Anthony LaMarca
Marc Langheinrich
Khai N. Truong (Eds.)

LNCS 4480

Pervasive Computing

5th International Conference, PERVASIVE 2007
Toronto, Canada, May 2007
Proceedings

Springer

Lecture Notes in Computer Science 4480

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

Anthony LaMarca   Marc Langheinrich
Khai N. Truong (Eds.)

Pervasive Computing

5th International Conference, PERVASIVE 2007
Toronto, Canada, May 13-16, 2007
Proceedings

Springer

Volume Editors

Anthony LaMarca
Intel Research Seattle
1100 NE 45th Street, Seattle, WA 98105 , USA
E-mail: anthony.lamarca@intel.com

Marc Langheinrich
ETH Zurich
Institute for Pervasive Computing
Clausiusstr. 59, IFW, 8092 Zurich, Switzerland
E-mail: langheinrich@inf.ethz.ch

Khai N. Truong
University of Toronto
Department of Computer Science
10 King’s College Road, Toronto ON M5S 3G4, Canada
E-mail: khai@cs.toronto.edu

Library of Congress Control Number: 2007924890

CR Subject Classification (1998): C.2.4, C.3, C.5.3, D.4, H.3-5, K.4, K.6.5, J.7

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

ISSN 0302-9743
ISBN-10 3-540-72036-7 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-72036-2 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12050604 06/3180 5 4 3 2 1 0

Preface

On behalf of the Organizing Committee for Pervasive 2007, welcome to the proceedings of the Fifth International Conference on Pervasive Computing. The year 2007 was a milestone for this young conference series, as it marked the first year that it was held outside of Europe. This "globalization" of the Pervasive conference is further evidence of its emergence as one of the most respected venues for publishing research on pervasive and ubiquitous computing. Each Pervasive conference represents a snapshot of our field, capturing the state of the art in technology, the research challenges we are addressing, as well as the people that form the Pervasive community. Each year's proceedings represent an opportunity to learn about new areas and identify upcoming research opportunities, and 2007 was no exception.

Pervasive 2007 attracted 132 high-quality submissions, from which the Technical Program Committee accepted 21 papers, resulting in a competitive 16% acceptance rate. In total, there were over 400 individual authors from 32 countries, coming from a wide array of disciplines and from both academic and industrial organizations. Papers were selected solely on the quality of their peer reviews using a double-blind review process. We were fortunate to draw on the combined experience of over 125 years of research in pervasive and ubiquitous computing in our 28 Program Committee members, coming from the most prestigious universities and research labs in Europe, North America, and Asia. Program Committee members were aided by no less than 177 external reviewers in this rigorous process, in which each Committee member wrote about ten reviews and each external reviewer between one and six. The total of 541 entered reviews resulted in an average of 4.2 reviews per paper, or almost 2,000 words of feedback (some 5 pages). To ensure that we had quality reviews as well as substantive deliberation on each paper, a final discussion phase generated 833 discussion items before the start of the Committee meeting. As a result, some 60 submissions were selected for discussion at the two-day, 14-hour Program Committee meeting, where a strict schedule and many additional reading assignments ensured that each candidate paper had been thoroughly evaluated by at least three Committee members present.

Pervasive has traditionally been a practitioner's conference, and the 2007 program certainly shows that tradition. With pervasive and ubiquitous computing gradually coming of age, research papers had to demonstrate not only novelty, but also applicability and feasibility. A major theme throughout the technical paper presentations was thus the rigorous evaluation and re-evaluation of novel and (seemingly) well-known principles in both home and work domains. Location continued to play a major role in the community, although refinement of existing principles and its implications for applications and services received increasing attention. Security and privacy issues, long since on the Pervasive agenda, were considered more and more with an eye towards usability and practicality, especially in the context of location-aware applications, where the increased availability of real-world data begins to allow realistic analyses. Last but not least, the prevalence of mobile phones and their use as an interface to a variety of applications and services emerged as an established new community research focus.

The research papers, although at the core of the Pervasive conference series, are hardly its only highlight. In addition to the technical sessions, Pervasive 2007 featured a keynote speech by Adam Greenfield—an internationally recognized writer and user experience consultant. As with all of its predecessors, the 2007 conference also contained a poster and demo reception, as well as a full day of workshops and doctoral colloquium. As a first for Pervasive, the final day of the conference was dedicated to tutorials by research community leaders. This series of short tutorials focused on the fundamentals of pervasive computing and were intended as an introduction for new researchers and practitioners in the field.

Several organizations provided financial and logistical assistance in putting Pervasive 2007 together, and we would like to acknowledge their support. We thank the University of Toronto for hosting the conference and managing the local logistics. We would also like to thank the staff and students at the Georgia Tech College of Computing for hosting the Program Committee meeting and ensuring an ample supply of food, coffee, and post-its throughout the two days. We very much appreciate the support of our Gold Sponsors Google, Intel and Nokia, Bronze Sponsor SMART Technologies Inc., along with the generous donations from FXPAL, MERL, and Microsoft Research—all of this was essential to helping us provide an excellent conference experience for all attendees. Lastly, we would like to thank, most of all, the authors who submitted their work to Pervasive, and the Program Committee members and our external reviewers who spent many hours reviewing submissions, shepherding papers, and providing the feedback needed to create this final program.

May 2007

Anthony LaMarca
Marc Langheinrich
Khai Truong

Organization

Conference Committee

General Chair	Khai Truong, University of Toronto, Canada
Program Co-chairs	Anthony LaMarca, Intel Research Seattle, USA Marc Langheinrich, ETH Zurich, Switzerland
Demonstrations	Darren Leigh, MERL, USA Gerald Morrison, Smart Technologies, Canada
Doctoral Colloquium	Boriana Koleva, University of Nottingham, UK Albrecht Schmidt, Fraunhofer IAIS and b-it University of Bonn, Germany
Late Breaking Results	Regan Mandryk, Dalhousie University, Canada Desney Tan, Microsoft Research, USA
Tutorials	John Krumm, Microsoft Research, USA Tom Pfeifer, Waterford Institute of Technology, Ireland
Videos	Hao Chu, National Taiwan University, Taiwan James Scott, Microsoft Research Cambridge, UK
Workshops	Anind Dey, Carnegie Mellon University, USA Gillian Hayes, Georgia Institute of Technology, USA
Publications	Lorcan Coyle, University College Dublin, Ireland Steve Neely, University College Dublin, Ireland
Publicity	Matthias Kranz, Technische Universität Braunschweig, Germany Alex Varshavsky, University of Toronto, Canada
Student Volunteers	Julie Kientz, Georgia Institute of Technology, USA Iqbal Mohomed, University of Toronto, Canada
Business Manager	Sandy Romas, University of Toronto, Canada
Events Coordinator	Cody Copeman, University of Toronto, Canada

Program Committee

Gregory Abowd, Georgia Institute of Technology, USA
Louise Barkhuus, University of Glasgow, UK
Michael Beigl, TU Braunschweig, Germany
A.J. Brush, Microsoft Research, USA
Nigel Davies, Lancaster University, UK
Maria Ebling, IBM T. J. Watson Research Center, USA
Alois Ferscha, Linz University, Austria
Christian Floerkemeier, ETH Zurich, Switzerland
Adrian Friday, Lancaster University, UK
Hans Gellersen, Lancaster University, UK
Mike Hazas, Lancaster University, UK
Ken Hinckley, Microsoft Research, USA
Hideki Koike, University of Electro-Communications, Japan
John Krumm, Microsoft Research, USA
Anthony LaMarca, Intel Research Seattle, USA
Marc Langheinrich, ETH Zurich, Switzerland
Pedro Marron, University of Stuttgart, Germany
Kenji Mase, Nagoya University, Japan
Yasuto Nakanishi, Keio University, Japan
Donald J. Patterson, University of California at Irvine, USA
Jun Rekimoto, Sony CSL, Japan
Tom Rodden, University of Nottingham, UK
Bernt Schiele, TU Darmstadt, Germany
Chris Schmandt, Massachusetts Institute of Technolgy, USA
Albrecht Schmidt, Fraunhofer IAIS and b-it University of Bonn, Germany
James Scott, Microsoft Research Cambridge, UK
Desney Tan, Microsoft Research, USA
Andrea Vitaletti, Sapienza - Università di Roma, Italy
Harald Vogt, SAP Research, Germany
Jacob O. Wobbrock, University of Washington, USA

Reviewers

Robert Adelmann, ETH Zurich, Switzerland
Luca Allulli, Sapienza - Università di Roma, Italy
Ken Anderson, Intel Research, USA
Takuya Araki, NEC Corporation, Japan
Yuji Ayatsuka, Sony Computer Science Laboratories, Japan
Maribeth Back, FXPAL, USA
Rafael Ballagas, RWTH Aachen University, Germany
Jakob Bardram, The IT University of Copenhagen, Denmark
John Barton, IBM Almaden Research Center, USA
Patrick Baudisch, Microsoft Research, USA
Martin Bauer, NEC Europe Ltd., Germany

Luca Becchetti, Sapienza - Università di Roma, Italy
Marc Bechler, BMW Group, Germany
Ashweeni Beeharee, University College London, UK
Marek Bell, University of Glasgow, UK
Zinaida Benenson, University of Mannheim, Germany
Alastair Beresford, University of Cambridge, UK
Gaetano Borriello, University of Washington, USA
Philippe Cattin, ETH Zurich, Switzerland
Keith Cheverst, Lancaster University, UK
Tanzeem Choudhury, Intel Research Seattle, USA
Mark Corner, University of Massachusetts, USA
Dan Cosley, Cornell University, USA
Andy Crabtree, University of Nottingham, UK
Lorrie Cranor, Carnegie Mellon University, USA
James Crowley, INRIA Rhone Alpes, France
Francisco Curbera, IBM T. J. Watson Research Center, USA
John Davis, IBM T.J. Watson Research Center, USA
Massimiliano de Leoni, Sapienza - Università di Roma, Italy
Christian Decker, TecO, University of Karlsruhe, Germany
David DeRoure, University of Southampton, UK
Anind Dey, Carnegie Mellon University, USA
Mark Dunlop, Strathclyde University, UK
Christos Efstratiou, Lancaster University, UK
Elena Fasolo, University of Padova, Italy
Kenneth Fishkin, Google, USA
Kentaro Fukuchi, University of Electro-Communications, Japan
Masaaki Fukumoto, NTT DoCoMo Multimedia Laboratories, Japan
Krzysztof Gajos, University of Washington, USA
Matthias Gauger, University of Stuttgart, Germany
Manuel Gonzalo, University of Stuttgart, Germany
Saul Greenberg, University of Calgary, Canada
Rebecca E. Grinter, Georgia Institute of Technology, USA
William Griswold, UC San Diego, USA
François Guimbretière, University of Maryland, USA
Stefan Gusenbauer, University of Linz, Austria
Jörg Hähner, Leibniz University Hannover, Germany
Jonna Hakkila, Nokia Group, Finland
Alastair Hampshire, University of Nottingham, UK
Matthias Handy, University of Rostock, Germany
Robert Harle, University of Cambridge, UK
Bjoern Hartmann, Stanford University, USA
Gillian Hayes, Georgia Institute of Technology, USA
Abdelsalam Helal, University of Florida, USA
Jeffrey Hightower, Intel Research Seattle, USA

Jaap-Henk Hoepman, TNO ICT and Radboud University Nijmegen, The Netherlands
Jesse Hoey, University of Dundee, UK
Paul Holleis, University of Munich, Germany
Lars Erik Holmquist, Swedish Institute of Computer Science, Sweden
Henry Holtzman, MIT Media Lab, USA
Jason Hong, Carnegie Mellon University, USA
Eric Horvitz, Microsoft Research, USA
Rachid Hourizi, University of Bath, UK
Elaine Huang, Georgia Institute of Technology, USA
Karin Hummel, University of Vienna, Austria
Amy Hurst, Carnegie Mellon University, USA
Heinrich Hussmann, University of Munich, Germany
Takeo Igarashi, The University of Tokyo / JST PRESTO, Japan
Sozo Inoue, Kyushu University, Japan
Stephen Intille, Massachusetts Institute of Technology, USA
Masaki Ito, Keio University, Japan
Seigo Ito, Nagoya University, Japan
Subu Iyer, Hewlett Packard, USA
Shahram Izadi, Microsoft Research, UK
Melody Moore Jackson, Georgia Institute of Technology, USA
Seiie Jang, Nagoya University, Japan
Rui Jose, University of Minho, Portugal
Hideaki Kanai, JAIST, Japan
Jong Hee Kang, University of Washington, USA
Günter Karjoth, IBM Zurich Research Lab, Switzerland
Amy Karlson, University of Maryland, USA
Nobuo Kawaguchi, Nagoya University, Japan
Nobuo Kawaguti, Nagoya University, Japan
Yoshihiro Kawahara, The University of Tokyo, Japan
Joseph 'Jofish' Kaye, Cornell University, USA
Minkyong Kim, IBM Watson, USA
Naohiko Kohtake, JAXA/Keio University, Japan
Panu Korpipää, Finwe Ltd., Finland
Albert Krohn, Particle Computer GmbH, Germany
Kai Kunze, University of Passau, Germany
Axel Küpper, University of Munich, Germany
Andreas Lachenmann, University of Stuttgart, Germany
Sven Lahde, Technical University of Braunschweig, Germany
Matthias Lampe, ETH Zurich, Switzerland
Johnny Lee, Carnegie Mellon University, USA
Stefano Leonardi, Sapienza - Università di Roma, Italy
Pamela Ludford, University of Minnesota, USA
Paul Lukowicz, University of Passau, Germany
Kent Lyons, Gerogia Institute of Technology, USA

Carsten Magerkurth, Fraunhofer IPSI, Germany
Alberto Marchetti Spaccamela, Sapienza - Università di Roma, Italy
Natalia Marmasse, Google, Haifa, Israel
Toshiyuki Masui, AIST, Japan
Rene Mayrhofer, Lancaster University, UK
Robert McGrath, University of Illinois, Urbana-Champaign, USA
Massimo Mecella, Sapienza - Università di Roma, Italy
Daniel Minder, University of Stuttgart, Germany
Luciana Moreira Sa de Souza, SAP Research, Germany
Ricardo Morla, Universitiy of California, Irvine and INESC Porto, Portugal
Paul Moskowitz, IBM T. J. Watson Research Center, USA
Hanspeter Mössenböck, University of Linz, Austria
Henk Muller, University of Bristol, UK
Yoshiyuki Nakamura, AIST, Japan
Carman Neustaedter, University of Calgary, Canada
David H. Nguyen, UC Irvine, USA
Jeffrey Nichols, IBM Almaden Research Center, USA
Alex Olwal, KTH (Royal Institute of Technology), Sweden
Leif Oppermann, University of Nottingham, UK
Patrik Osbakk, Jayway, Sweden
Benedikt Ostermaier, ETH Zurich, Switzerland
Sharon Oviatt, OHSU, USA
Shwetak Patel, Georgia Institute of Technology, USA
Matthai Philipose, Intel Research, USA
Steven Pickles, University of Florida, USA
Maria Pimentel, University of Sao Paulo, Brazil
Claudio Pinhanez, IBM T.J. Watson Research Center, USA
Ivan Poupyrev, Sony CSL, Japan
Stefano Puglia, WLAB Srl, Italy
Ali Rahimi, Massachusetts Institute of Technology, USA
Madhu Reddy, Penn State University, USA
Andrew Rice, University of Cambridge, UK
Till Riedel, TecO, University of Karlsruhe, Germany
Andreas Riener, University of Linz, Austria
Matthias Ringwald, ETH Zurich, Switzerland
Philip Robinson, SAP AG, Germany
Christof Roduner, ETH Zurich, Switzerland
Christian Rohner, Uppsala University, Sweden
Michael Rohs, Deutsche Telekom Laboratories, Germany
Mario Romero, Georgia Institute of Technology, USA
Umar Saif, LUMS, Pakistan
Daniel Salber, The Venice Project, The Netherlands
Antti Salovaara, Helsinki Institute for Information Technology, Finland
T. Scott Saponas, University of Washington, USA
Fabiano Sarracco, Sapienza - Università di Roma, Italy

Ichiro Satoh, National Institute of Informatics, Japan
Olga Saukh, University of Stuttgart, Germany
Robert Sauter, University of Stuttgart, Germany
Jochen Seitz, Technische Universität Ilmenau, Germany
Aline Senart, Trinity College Dublin, Ireland
Phoebe Sengers, Cornell University, USA
Steven Shafer, Microsoft Research, USA
Frank Siegemund, European Microsoft Innovation Center, Germany
Itiro Siio, Ochanomizu University, Japan
Timothy Sohn, University of California, San Diego, USA
Mirjana Spasojevic, Nokia Research, USA
Michael Steiner, IBM T.J. Watson Research Center, USA
Thomas Strang, DLR Oberpfaffenhofen, Germany
Martin Strohbach, NEC Europe Ltd., Germany
Yasuyuki Sumi, Kyoto University, Japan
Shigeru Tajima, Sony CSL, Japan
Hiroyuki Tarumi, Kagawa University, Japan
Tsutomu Terada, Osaka University, Japan
Masahiko Tsukamoto, Kobe University, Japan
Kent Unruh, University of Washington, USA
Kristof Van Laerhoven, Darmstadt University of Technology, Germany
Alex Varshavsky, University of Toronto, Canada
Sunil Vemuri, Massachusetts Institute of Technology, USA
Roel Vertegaal, Queen's University, Canada
Nicolas Villar, Lancaster University, UK
Daniel Vogel, University of Toronto, Canada
Amy Voida, Georgia Institute of Technology, USA
Steven Wall, University of Glasgow, UK
Leonardo Weiss F. Chaves, SAP Research, Germany
Andy Wilson, Microsoft Research, USA
Christopher Wren, Mitsubishi Electric Research Laboratories, USA
Zhiwen Yu, Nagoya University, Japan
Tobias Zimmer, Particle Computer GmbH, Germany

Gold Sponsors

Google
Intel
Nokia

Bronze Sponsors

SMART Technologies Inc.

Donors

FX Palo Alto Laboratory (FXPAL)
Microsoft Research
Mitsubishi Electric Research Laboratories (MERL)

Table of Contents

Reaching Out

Augmenting Looking, Pointing and Reaching Gestures to Enhance the Searching and Browsing of Physical Objects 1
David Merrill and Pattie Maes

Reach Out and Touch: Using NFC and 2D Barcodes for Service Discovery and Interaction with Mobile Devices 19
Eamonn O'Neill, Peter Thompson, Stavros Garzonis, and Andrew Warr

Combining Web, Mobile Phones and Public Displays in Large-Scale: Manhattan Story Mashup 37
Ville H. Tuulos, Jürgen Scheible, and Heli Nyholm

Context and Its Application

PersonisAD: Distributed, Active, Scrutable Model Framework for Context-Aware Services 55
Mark Assad, David J. Carmichael, Judy Kay, and Bob Kummerfeld

An Exploration into Activity-Informed Physical Advertising Using PEST 73
Matthias C. Sala, Kurt Partridge, Linda Jacobson, and James "Bo" Begole

Evaluating a Wearable Display Jersey for Augmenting Team Sports Awareness 91
Mitchell Page and Andrew Vande Moere

Using Location, Bearing and Motion Data to Filter Video and System Logs 109
Alistair Morrison, Paul Tennent, John Williamson, and Matthew Chalmers

Security and Privacy

Inference Attacks on Location Tracks 127
John Krumm

Shake Well Before Use: Authentication Based on Accelerometer Data 144
Rene Mayrhofer and Hans Gellersen

Virtual Walls: Protecting Digital Privacy in Pervasive Environments 162
Apu Kapadia, Tristan Henderson, Jeffrey J. Fielding, and David Kotz

Trust, Privacy and Relationships in 'Pervasive Education': Families' Views on Homework and Technologies 180
Katie Fraser, Tom Rodden, and Claire O'Malley

Understanding Use

Operating Appliances with Mobile Phones – Strengths and Limits of a Universal Interaction Device 198
Christof Roduner, Marc Langheinrich, Christian Floerkemeier, and Beat Schwarzentrub

Portable, But Not Mobile: A Study of Wireless Laptops in the Home 216
Allison Woodruff, Ken Anderson, Scott D. Mainwaring, and Ryan Aipperspach

Managing Communication Availability and Interruptions: A Study of Mobile Communication in an Oncology Department 234
Jeremiah Scholl, Per Hasvold, Eva Henriksen, and Gunnar Ellingsen

Sensing

Using Ground Reaction Forces from Gait Analysis: Body Mass as a Weak Biometric 251
Jam Jenkins and Carla Ellis

Weight-Sensitive Foam to Monitor Product Availability on Retail Shelves 268
Christian Metzger, Jan Meyer, Elgar Fleisch, and Gerhard Tröster

Assessing and Optimizing the Range of UHF RFID to Enable Real-World Pervasive Computing Applications 280
Steve Hodges, Alan Thorne, Hugo Mallinson, and Christian Floerkemeier

Movement-Based Group Awareness with Wireless Sensor Networks 298
Raluca Marin-Perianu, Mihai Marin-Perianu, Paul Havinga, and Hans Scholten

Finding and Positioning

Zone-Based RSS Reporting for Location Fingerprinting 316
Mikkel Baun Kjærgaard, Georg Treu, and Claudia Linnhoff-Popien

TrackSense: Infrastructure Free Precise Indoor Positioning Using Projected Patterns 334
Moritz Köhler, Shwetak N. Patel, Jay W. Summet, Erich P. Stuntebeck, and Gregory D. Abowd

Objects Calling Home: Locating Objects Using Mobile Phones 351
Christian Frank, Philipp Bolliger, Christof Roduner, and Wolfgang Kellerer

Author Index 369

Augmenting Looking, Pointing and Reaching Gestures to Enhance the Searching and Browsing of Physical Objects

David Merrill and Pattie Maes

MIT Media Lab, Cambridge MA 02139, USA
{`dmerrill, pattie`}`@media.mit.edu`

Abstract. In this paper we present a framework for attaching information to physical objects in a way that can be interactively browsed and searched in a hands-free, multi-modal, and personalized manner that leverages users' natural looking, pointing and reaching behaviors. The system uses small infrared transponders on objects in the environment and worn by the user to achieve dense, on-object visual feedback usually possible only in augmented reality systems, while improving on interaction style and requirements for wearable gear. We discuss two applications that have been implemented, a tutorial about the parts of an automobile engine and a personalized supermarket assistant. The paper continues with a user study investigating browsing and searching behaviors in the supermarket scenario, and concludes with a discussion of findings and future work.

1 Introduction

Today there is a wealth of online information and services available that reference the objects in our physical environment. Sitting behind our computers, we can consult websites to find the fair price of a used car, get book reviews, learn how to clear a jammed printer, create and explore a representation of our social network, and interact in countless other ways with information that references our physical experience in the world. However, when we are focused on an object in the physical world, away from our computers, it is often the case that we cannot easily access information and services related to the object. Ishii and Ullmer pointed out our modern-day 'dual-citizenship' in the digital and physical worlds [1], and given this reality, our experience would be enriched by a more seamless coupling between the two.

While some examples of widespread mobile access to contextually-relevant information are beginning to emerge [2], these approaches still fall short in terms of natural interaction and usability. But what characteristics should a mobile information-access device have? Jun Rekimoto [3] suggests two guidelines for designers to consider when creating unobtrusive wearable technology. The first, related to usability, is that any wearable device should support *hands-free operation*, meaning that the user should be able to operate the device without holding

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 1–18, 2007.

it in their hand, or at least the device should enable "quick changes between normal and operational mode." The second guideline is that these devices should be "*socially acceptable*, natural and (conceptually) as un-noticeable as possible for use in various social settings."

Our approach seeks to couple digital information to physical objects in a way that can be seamlessly browsed using natural actions while a user is mobile. We make this digital information accessible in ways that can respond to the focus of a person's attention, and our system remains hands-free and socially acceptable. Our design process was informed by our interest in direct interactions with physical objects, without requiring cumbersome gear.

This paper presents a novel system featuring small, powered 'transponders' and simple mobile devices that can communicate with these transponders via short-range, directional infrared (IR) light. The transponders can be fixed to objects or parts of objects in the physical environment, and they provide a range of in-situ visual feedback to support searching and browsing activities. In order to preserve social acceptability, one mobile device is built around a commercial hands-free Bluetooth audio earpiece, and the other is a finger-worn ring. When a person is wearing the augmented earpiece, the system can gauge which transponders are in their field of view, and can offer related information. Wearing the ring, a person can use the natural gesture of pointing or reaching towards an object to indicate interest, and the system can respond with related information. We present two example applications we have created to demonstrate these novel interactions: a tutorial about the parts of an automobile engine, and a supermarket assistant that helps a person find appropriate foods on the shelves. These systems explore the use of both auditory and on-object visual feedback in order to connect the user to relevant and personalized information about the physical objects around them.

A primary contribution of this work is to share our approach to designing attention-focused systems that leverage people's natural gestures to interact with on-object, personalized visual feedback. Secondly, we discuss the results from our user study, a physical-world searching and browsing task using our system, and what implications these findings have for mobile, digitally-augmented search in the physical world. We hope that our presentation of this work be useful to other designers of systems that seek to augment physical objects with lightweight in-situ feedback.

2 Background and Related Work

This work draws inspiration from research in the areas of location and context-aware systems. The Active Badge from Olivetti Research Lab [4], and the Xerox PARCTab [5] were important early demonstrations of how a person's location could be coarsely sensed using infrared transponders and used as context to provide services in a workplace like automatic call forwarding and continually updated maps for finding people. Our work also uses infrared transponders, but we use them in a more fine-grained manner, to identify a person's attention towards a particular object or a given part of an object rather than their presence in a space.

Many systems that connect digital information to physical objects make use of a handheld display device. Want, et al. connected electronic services and actions to physical objects using embedded RFID tags, enabling users to initiate responses on a tablet computer upon sensing the tags [6]. In HP's Cooltown project [7], users carried a PDA that used infrared communication to access information, services and maps associated with nearby real-world entities like meeting rooms and office printers. Likewise, the AURA platform [8] uses a PDA with a barcode scanner to retrieve and share product-related information. iCam [9] and Patel and Abowd's 2-Way Laser Assisted Selection Scheme [10] use a laser pointer in conjunction with a hand-held device to enable remote selection of tagged objects in a physical space, and exploration of their associated digital information (the laser pointer is modulated and used for object identification in the latter, while it is only used for rangefinding in iCam). Progress on optical glyph recognition on mobile phones is permitting their use as displays for digital annotations [11]. Finally, tour guide systems [12][13][14] also address a similar problem space, and have made extensive use of hand-held computers and PDAs.

The aforementioned systems that make use of a hand-held display device do not satisfy Rekimoto's hands-free guideline, and moreover they force the user to focus visual attention on a relatively small screen, rather than on the physical objects of interest. We feel that these intermediary handheld devices interfere with a user's experience of interacting directly with physical objects, so we have situated visual feedback on the objects themselves, and have used the auditory channel to deliver other related information. Our work shares some features with the FindIT Flashlight system [15], which also triggers visual on-object feedback using infrared communication. However our system is designed for different usage scenarios; whereas the hand-held FindIT Flashlight is more power-optimized than our system and targeted at allowing a single user to quickly locate specific objects, we have designed more flexibility in visual feedback into our hands-free system (different colors and blinking patterns) to augment casual browsing, as well as feedback personalization for individual users, visual feedback sync with synthesized speech, and navigation assistance.

Augmented-reality (AR) [16] can also present auditory output and overlay information on top of a view of the physical world. However, these systems tend to encumber the user with additional mechanisms, ranging from a hand-held PDA [17] [18] to a full heads-up-display [19]. Additionally, the ability to deploy AR systems [20] is often limited by the 'registration problem' in which the computer-generated graphics must be properly aligned with the user's view of their surroundings. Although augmented reality systems can excel at providing rich, contextually-relevant information, this extra gear is cumbersome, expensive and not generally socially acceptable for the everyday person.

Other AR-like approaches that avoid intermediary hand-held devices feature specially designed spaces or surfaces that are carefully instrumented with sensing apparatus that allow a person to interact directly with physical objects. The

Perceptive Workbench, and [21] Sensetable [22] are examples of this class of system. While these systems obviate the need for cumbersome wearable or handheld gear, they tend to be expensive to build and are not suitable for a mobile user. Furthermore, they have trouble scaling easily to large physical spaces or numbers of items. Our approach is designed to address the registration problem while still achieving an inexpensive, mobile, and scaleable solution at the cost of some richness in visual feedback possibilities.

Finally, we note the development of various systems that sense and leverage a user's visual attention towards people objects in their environment [23]. These systems usually utilize eye-tracking or head-tracking, and they tend to require a camera mounted on the user's head or desk. In contrast to our work, systems utilizing a desk-mounted camera [24] [25] are not suitable for mobile use, and head-mounted eye-tracking cameras are bulky and expensive. ViewPointer [26] improves on bulk and expense by using a small, cheap head-mounted camera trained on the user's eye to sense gaze towards specific infrared beacons in the environment, but the speed of data communication is limited by the camera's frame rate. EscGlasses and EyePliances [27] are two other camera-based systems that sense a user's attention towards individual objects in an environment. These systems succeed at detecting a user's attention towards a single object at a time. However on-object visual feedback and faster data-transmission rate makes our system better suited to support casual searching and browsing of many items simultaneously.

3 Two Example Applications

Prototyping new interaction scenarios is essential to understand how they will work in practice, and how to make them better. We built two distinct applications that utilize the same underlying infrastucture: Engine-Info, and My-ShoppingGuide.

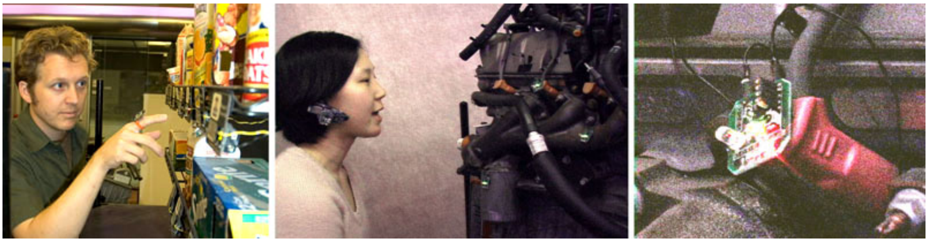


Fig. 1. A user pointing with the finger-worn ring (left), another user wearing the Bluetooth earpiece and interacting with the Engine-Info application (middle), and a close-up of a spark plug-mounted transponder blinking to visually point out the physical object being spoken by the system (right)

3.1 Engine-Info Application

In this application, situated[1] visual feedback and speech dialogue are used teach a user about the names, functions, and relationships between the components of an automobile engine. An engine was chosen because it is an object that is fairly complicated and densely populated with interesting external features.

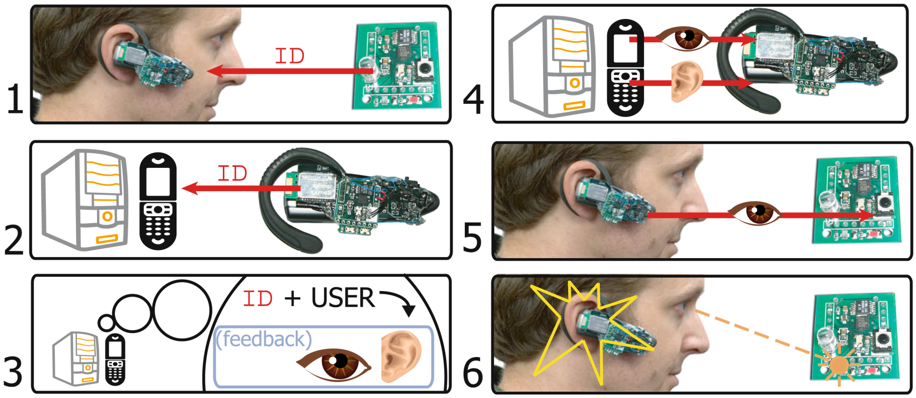


Fig. 2. Overview of message and feedback flow in the application. Steps are as follows: (1) On-object transponder broadcasts its ID over infrared which is received by the headset (2) Headset transmits transponders ID to server on a mobile phone or PC via Bluetooth serial channel (3) Server compares ID to user profile, generates auditory and/or visual feedback response (4) Server transmits visual feedback instruction to headset over Bluetooth serial channel, and plays synthesized speech to user over headsets built-in Bluetooth audio (5) Headset transmits visual feedback instruction to transponder over Bluetooth (6) User hears synthesized speech and sees coordinated visual feedback on transponder.

To interact, the user wears a modified Bluetooth audio earpiece (figure 3) and looks at a part of the engine that they are interested in. The earpiece has an infrared transmitter and receiver oriented to communicate in the direction of the user's gaze. Transponders on the engine's features glow green when they see the user's earpiece, giving the user a visual indication of their presence and position. An earpiece-mounted transponder relays the numeric identifiers of the detected on-object transponders back to an application server, providing the server with a continually updated report of which transponders are in the user's field of view (see figure 2 for details). Each transponder on the engine corresponds to an information node in a hierarchical knowledge-base about the engine. Given the user's attention, their profile, and their interaction history, the system conducts a speech dialogue with the user to give descriptions of the parts of the engine,

---

[1] We use *situated*, or *in-situ* feedback to refer to feedback that is located on the actual object in the environment. This is in contrast to systems that present feedback on the screen of a PDA or other mobile device.

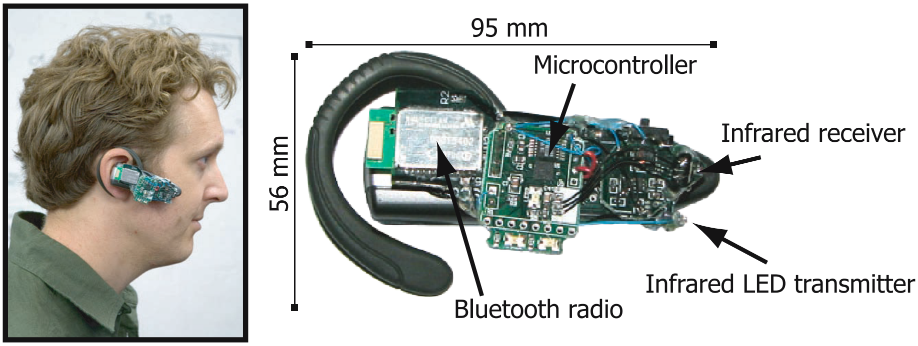


Fig. 3. The modified Bluetooth earpiece. The attached transponder is powered from the earpiece's battery through a step-up voltage regulator.

starting from one of the items in their field of view. When a particular item is mentioned verbally, a light-emitting diode (LED) on the given transponder flashes to draw the user's attention. The system gives navigation assistance to help the user locate each successive item, and confirms with the user that they have found each new item. For more details, please see section 4.2.

3.2 My-ShoppingGuide

This application uses situated visual feedback to help a user find appropriate foods on augmented supermarket shelves. Each shelf label has an embedded transponder that provides visual feedback, informed by the match between a user profile and the contents of the food item. My-ShoppingGuide works with either the Bluetooth earpiece or a finger-worn ring (figure 4). When the earpiece or ring receives an identifier broadcast from a shelf-mounted transponder, it consults the application server to determine the match between that product and the currently loaded personal profile. If there is a conflict between the product and the profile, the red light on the product-mounted transponder is activated. If the profile indicates a particular interest in the given product, the yellow light is triggered with a blinking sequence. The effect for the user is that when they orient their face (wearing the earpiece) or reach/point their hand (wearing the ring) towards a shelf, product labels lights up with personalized visual feedback.

4 Implementation

4.1 Hardware

Transponder. The transponder is an inexpensive hardware platform with limited processing, communication, data storage and feedback capabilities. Based around a Silicon Laboratories C8051F331 microcontroller [28], the transponder has an infrared (IR) light-emitting diode (LED), an IR receiver with integrated demodulator, three visible-light LED's, and a 1-megabit flash memory

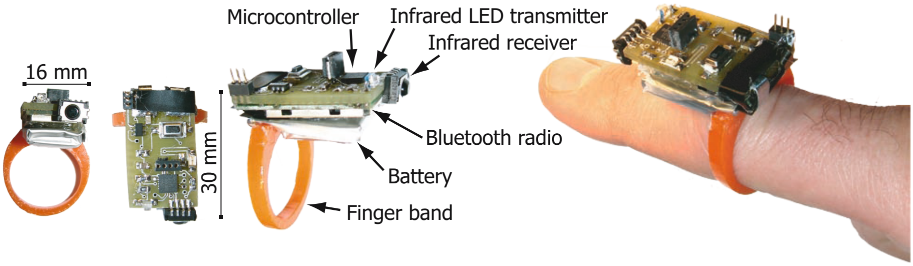


Fig. 4. The finger-worn ring

chip. A variation of the Sony-IR protocol is used for communication between transponders, and the IR light is modulated at 38kHz, for communication at 1200 bits/sec. Each transponder periodically broadcasts a message over IR containing its unique numeric identifier. The period between broadcasts is different for each transponder, in order to avoid repeated collisions from nearby transponders. Communication errors are detected by a single-byte checksum, and incoming packets that fail to match the checksum are discarded. Typical round-trip time from a transponder broadcast to visual feedback is 250 milliseconds.

For the Engine-Info application, twenty transponders were cut to half-size and affixed to all major external features of an automobile engine. For MyShoppingGuide, transponders were embedded into supermarket shelf labels. Each transponder was attached behind a label such that the infrared and visible LED's showed through a square hole cut into the label. In both applications power was routed to the transponders from a concealed DC power supply. Each transponder cost about $25USD (qty. 200), including parts and assembly.

Bluetooth Earpiece and Finger-worn Ring. An off-the-shelf Bluetooth audio earpiece was fitted with a modified transponder (figure 3), making it a wearable attention sensor and outgoing communication channel from the application server to the environmentally-placed transponders. The modified transponder was programmed to re-broadcast visual signaling instructions received from the server using a serial-over-Bluetooth link via an attached Bluetooth radio. Its infrared transmitter and receiver were re-oriented so that they would point in the direction of the earpiece wearer's gaze, and it draws power from the earpiece's battery. In a widespread adoption scenario, this extra hardware could be easily integrated into existing bluetooth earpieces at minimal extra cost.

For the supermarket application, a finger-worn ring (figure 4) was built with the same data communication capabilities as the modified earpiece. A small (20mm x 15mm x 5mm) rechargeable Lithium Polymer battery powers the ring.

4.2 Software

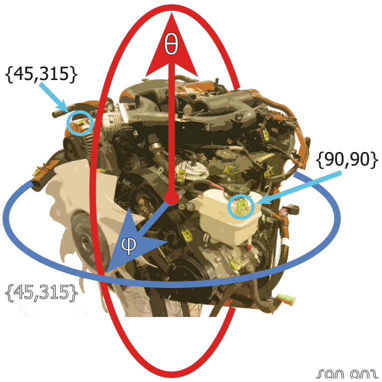
An application server was written in Java to be a general purpose, user-profile-aware state machine supporting physical-world interactions with augmented objects. The software was written on a PC running Windows XP, but could be moved to a mobile phone with modifications to the speech recognition and synthesis code. The primary services that the application server provides are mapping numeric identifiers to stored information and triggering feedback.

Information and Mapping. The mapping from a transponder's numeric identifier to related information and media is specified in an XML file loaded by the application server. Personalization is achieved easily by loading a user-specific file. The use of XML permits (but does not require) a hierarchical knowledge representation, as well as arbitrary links between any two nodes. In the Engine-Info application, grouped nodes are interpreted to be part of the same system (i.e. electrical, cooling). Dependencies, or suggested 'pre-requisite' relationships between nodes are specified as a field in the node's tag.

Input and Output. The application server is capable of limited spoken dialogue with the user, with the Bluetooth earpiece providing audio input and output. Speech input is supported using a small-domain speech recognizer implemented with the Sphinx software [29], and output is via the FreeTTS [30] speech synthesizer. Visual communication is implemented by triggering LED activity on transponders that are in the user's field of view. The spoken and visual outputs can be coordinated by triggering visual signaling on transponders during speech synthesis, a technique that conceptually connects the auditory information to tagged physical items.

Next-Node Selection and Navigation Assistance. In an open-ended exploration like Engine-Info, a method was needed for the system to decide which item to present next. We developed the 'attentional fair game' (AFG) metric, which provides some measure of the focus of a person's visual attention while remaining unobtrusive. The system tracks the set of transponders that have been spotted by the earpiece or ring during the past 5 seconds. When making a decision about which information to suggest next to the user, this set of most-likely-visible transponders is consulted, and they are given more weight in the selection heuristic. In this way the system has some approximation of the information that a human conversational partner would have about nearby objects, namely, which objects (or parts of objects) are in the user's field of view and are thus 'fair game' to be brought into the conversation and visually pointed out [31]. To choose which node to suggest next in Engine-Info, the system randomly picks another node in the same system (i.e. electrical, cooling) that has not been introduced yet, giving higher priority to nodes in the AFG. If the current group has been fully described, the system will randomly pick a node in the AFG from another group. If all nodes in the AFG have been described, the system will randomly pick any of the not-yet-described nodes on the object.

The spatial relationships between on-object transponders can also be encoded into the XML file, allowing the application server to give the user directions from one transponder to the next (figure 5). The system can tell the user which direction to look in order to find the next transponder, using a previously discussed transponder or one from the AFG as a starting reference point. See figure 5



Example Directions:

"From the spark plug, look up and to the left to find the engine oil"

"From the engine oil, look to the right to find the dipstick"

"From the exhaust manifold, look around to the other side to find the electrical connector"

Fig. 5. The spherical coordinate system's origin is at the object's center. θ represents a transponder's degrees from vertical, while φ is degrees from a given horizontal direction. Example coordinates of two transponders are indicated.

for some example directions. This coordinated visual feedback interleaved with the speech grounds the connection between the verbal dialogue and the physical world [32].

4.3 Personalization

The My-ShoppingGuide application loads a user profile from an XML file, which determines the visual feedback that each transponder should display (blinking yellow, green, red). A feedback mapping can be made explicitly by assigning the feedback category directly, or implicitly by categorizing a given ingredient, which may also be present in other food items. The profile supports, but does not *require* a semantic explanation for a categorization, and thus reasons for feedback could include considerations such as liking or disliking, conflict or accord with allergies, health considerations or the user's approval or disapproval of the manufacturer's environmental record.

5 Design Discussion

In designing the interaction, we were interested in leveraging existing behaviors for browsing and seeking information about physical objects. The existing behaviors that were augmented were: looking at, pointing at, and reaching for an object. These gestures were supported and augmented by the system's visual affordances, spoken referent/visual feedback connection, navigational guidance and personalization.

The Engine-Info interaction approximates some elements of having a co-located domain expert by sensing a user's looking behavior and engaging in a spoken dialogue. Borrowing interaction features from a typical person-person conversation, the system has a rudimentary understanding of which objects are in the user's field of view, and it utilizes this information to pick convenient topics of instruction. The system responds to verbal cues from the user and provides spoken information about objects. Like a human conversational partner, it can draw their attention to a given feature at the moment it is mentioned, to simulate a pointing gesture[2]. It tracks which topics have already been discussed, and although we have not yet explored these avenues, it can be personalized to level of expertise and language.

Our initial design for My-ShoppingGuide also intended to sense a user's visual attention, and to leverage their ability to efficiently browse many items simultaneously. Wearing the augmented Bluetooth headset, a user can stand back and look at an entire shelf of products. The labels in their field of view light up in a way that is informed by the match between each product and the currently-loaded personal profile. Though spatially distributed, this feedback can be quickly visually scanned. However, we learned from early feedback that when several people viewed the same shelves at the same time there could be some ambiguity about the intended feedback recipient. This ambiguity presents a challenge to *any* system in which personalized cues appear publicly in the environment. Thus the finger-worn ring was built to allow users to point at or reach towards a smaller range of objects that they are interested in. In pilot tests with early users, it became clear that this pointing/reaching ability serves two distinct purposes. Primarily, it allows more precise aiming of the infrared communication, avoiding activation of unintended feedback on nearby objects that are not of interest to the user. Second, pointing and reaching are bodily signals that make other nearby people aware of which object they are interacting with. In addition to pointing or reaching directly at a single object they are interested in, the ring also allows for a sweeping gesture in which the user can wave their hand across a shelf with a number of transponders, activating each in turn. This sweeping gesture preserves some of the convenient visual-scanning interaction that the headset allows, while making the interaction suitable for multi-person environments. Looking forward to even denser transponder deployments, it would be useful future work to quantify precisely how quickly visual feedback must appear in order to support sweeping gestures.

6 User Study

A small user study was run in order to investigate how our system would affect physical-world browsing and searching. 18 participants between the ages of 26 and 31 were recruited by email from the MIT Media Laboratory student population (7 female, 11 male). Each participant completed four browsing/searching tasks utilizing the My-ShoppingGuide augmented shelves. The study was between-participants, with six participants in each experimental

---

[2] Note that the system's ability to 'point' to an object or feature by triggering visual feedback is separate from our augmenting the user's actual pointing gesture.

Table 1. User study task instructions, same for all conditions

Task #	Printed instructions for each task
1	Shopping List: A food containing strawberries. An item containing sweet cream buttermilk. An item that is 'calcium enriched'. A breakfast cereal with no fat content. An item containing maltodextrin
2	Please find a beverage (a drink-able liquid) that does NOT contain any citrus or citric acid. (citrus/citric acid is found in oranges, lemons, grapefruits, etc.)
3	You are allergic to oatmeal, but enjoy cookies. Please find 2 types of cookies that do not contain oatmeal
4	Your body needs vitamins to grow and be healthy! But only vitamins B2 and B12, as it turns out. Find an item that contains the following vitamins (B2, B12) But that does NOT contain vitamin B1.

condition. The experimental conditions were (a) using the finger-worn ring, (b) using the Bluetooth earpiece, and (c) using neither device. Evaluation of the user experience with the Engine-Info application is left as future work.

The study consisted of four tasks, with a short questionnaire between each task. In the first task, participants were asked to collect five items with given properties from a printed shopping list (for example, 'a food containing strawberries' - see Table 1 for details). In each of the second, third and fourth tasks, participants were asked to collect small numbers of items from a certain category (i.e. beverages, cookies) while avoiding items that contained given 'allergens'. All participants were told that they were free to use whatever strategy they liked in order to find the given items, including the visual feedback (if applicable given their condition), the packaging of the products, and their existing knowledge. The correctness of items selected by participants was not enforced.

Task-related visual feedback was active in the conditions with the ring and earpiece. In the shopping list task, the feedback consisted of a flashing yellow LED on the label-embedded transponder for items on the printed list. In the allergy tasks, the feedback consisted of a glowing red LED on the label-embedded transponder of products containing allergens to avoid. This visual feedback appeared when the ring or earpiece was pointed in the direction of the given item label, and it was active throughout the entire task regardless of the participant's progress in collecting items. The non-feedback condition was the control group, and these participants received no visual feedback. Rather, they used only the packaging of the items and their existing knowledge to make their selections.

Completion time for each of the four tasks was measured by the experimenter with a stopwatch. After the final task, participants were asked to answer a few additional questions about their experience, and were given the opportunity to ask the experimenter about the study.

A presentation of experimental results and discussion follows.

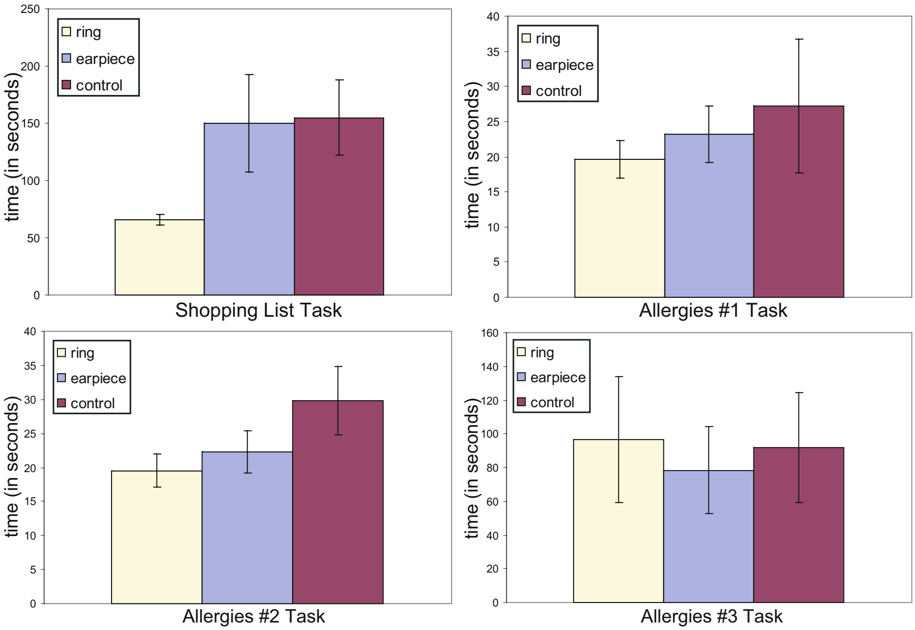


Fig. 6. Completion-time for the four tasks. Error bars show the standard error, computed as $\text{SD}/\sqrt{N}$.

6.1 Experimental Results

Average completion time (in seconds) on the shopping list task was significantly lower in the finger-worn ring group (M = 66) than for the control group (M = 155; t(10) = 2.67,p<.05,two-tailed). This result reflects the efficiency of the strategy that many participants in the ringer-worn ring condition reported using: sweeping their hand back and forth and trusting in the visual feedback to point out the desired items, rather than reading the ingredient lists or carefully examining other product packaging. The difference in average completion time between the earpiece group and control group was not significantly different, but there was a weakly significant difference between the finger-worn ring group (M = 66) and the earpiece group (M = 150; $t(10) = 1.95$,p=.08,two-tailed; ANOVA was inconclusive). We suspect that this difference was a result of the finger-worn ring setup permitting the wearable infrared transponder to be aimed more accurately, which made the appearance of visual feedback more reliable (some subjects in the earpiece condition allowed the earpiece to rotate downward at times, impairing the infrared communication between the wearable and the transponders in the environment).

Results in the allergies 1 and allergies 2 tasks suggest quicker completion times in the finger-worn ring condition as compared to the earpiece and control group, though the differences were not significant. In those same tasks, average completion times in the earpiece group came out slightly quicker than the control group.

The Allergies 3 task seemed to be the most difficult task for some participants, and differences in mean completion time between groups were not significantly different. In Allergies 3, participants were asked to find an item with vitamins B2 and B12, but *without* vitamin B1. Completion time on this task varied widely compared to the other allergies tasks, and participants reported that their search efficacy was largely determined by their pre-existing knowledge about these vitamins rather than the visual feedback or other factors. Participants who had familiarity with these vitamins tended to select the correct item very quickly, while other spent several minutes scrutinizing labels.

In the feedback conditions, participants reported using different strategies for the tasks, showing variation from low to high usage of the visual feedback. Some participants reported feeling that they had relied *too much* on the technology, and feared that they fell into what one participant referred to as ‘ring dependence’. One participant said:

> *“Its easy to rely on the lights without even thinking about what you are doing.”*

Some participants followed this concern with a remark about the potential danger posed by error in a deployment where the system would be recommending food products, especially if a shopper places a high degree of trust in the system to protect them from choosing foods that that contain allergens.

Not all participants felt that they over-trusted the technology though. A number of participants reported using a two-stage approach in the allergy-avoidance tasks. These participants would first browse the shelves in a traditional manner to find a potential candidate item, then they would use the visual feedback to confirm its acceptability. Some even followed this use of feedback by reading the ingredients list carefully, to make absolutely sure that the item was acceptable. Here are some reactions that participants reported:

> *“My common sense helps me narrow my search while the lights confirm my choices.”*
>
> *“I saw the cookie packages right away, but waited for the light to confirm that they were OK.”*

Finally, another participant referred to the visual feedback as a “secondary resource”, and said the following:

> *“I did not use the LEDs unless I was stumped.”*

Some participants noted that they appreciated being able to ignore the feedback if they wanted. Others suggested that they would like more ways to refine their search. The ability to change the search criteria while keeping the device pointed at a single item so as to ‘test’ it for multiple features or ingredients was requested, as was a feature to set up compound search criteria so that only items passing a multi-criterion profile match would be pointed out.

6.2 Discussion

> *"The ring really made things a lot easier - I could just scan it over the shelf and immediately identify if an item was on my list."*

The above sentiment about the shopping list task was expressed frequently among participants in the ring condition. It seems natural that a tool that visually points out exactly the items that a person is looking for will speed up a searching task. Interestingly though, as was mentioned in the results section, reliance on this easily available feedback varied from person to person. These individual differences suggest that for a physical world browsing or searching activity where people already have deep life experience patterns, such as finding food in a supermarket, there are likely to be a number of different strategies for using an attention-sensitive interactive augmentation like we have built. Furthermore, such a system should be available for use at the moment a person wants help or clarification, but also easy to ignore when they want to browse unassisted.

> *"Shopping in grocery stores is on some level about discovery - discovering new items, things you want to try etc. Its good though that it doesn't in any way impede normal browsing/shopping, you can totally ignore it if you want to."*

In observing participants, one common feature of their experience became obvious: reading the fine print on packaging takes a long time, and can be error-prone. Whenever a participant needed to read through the ingredients of a product, their completion time was dramatically slowed, and several made mistakes stemming from their unfamiliarity with reading and nutrition information and ingredients on food packaging. This lack of experience points to the benefit that a supplementary system like ours could bring the browsing experience. In order to reduce the need to read the fine print, some reported using the technology as a filter, narrowing the set of possibilities to investigate.

> *"I knew it would be a pain to read through the whole ingredients list. I found one that didn't have B1 using the light feedback, then checked the ingredients list to make sure it contained B2 and B12."*

> *"When I did look at the ingredients it was pretty much the same way I would do it at the supermarket. The difference was in helping me choose which items to bother picking up. It felt helpful more as a filter than as a selector."*

We find a recurring theme in these responses to the supermarket tasks, that for many participants, their usage of the system took place in concert with their existing searching and browsing strategies, augmenting rather than replacing them. Most quickly incorporated the visual feedback into the task, allowing it to either guide them directly, or to cut down the range of options that they needed to consider while finding the desired items. However, they did not abandon their pre-existing knowledge and habits.

We feel that this usage pattern applies to other physical world information scenarios where in-situ information about objects can be offered to people. As we saw in the My-ShoppingGuide recipe-finding task, on-object visual feedback can seamlessly augment a person's experience interacting with things in their environment, making the interaction more efficient. However this feedback is probably not an appropriate *replacement* for the pre-existing information. It can be used as a supplementary source of information during a browsing or searching task, but it is equally important that it can be ignored. We hope to see more systems in the future designed with these considerations in mind, enabling the seamless use of digital augmentation alongside existing information, and leveraging natural gestures that are cues of attention, to support searching and browsing in the physical world.

7 Conclusions

This paper has presented a general system for augmenting physical objects with relevant information and offering that information in a personalized way that responds to the user's attention through their natural looking, pointing and reaching gestures. We have introduced Engine-Info and My-ShoppingGuide, two applications that provide the background for the subsequent discussion of our interaction design. A user study of the My-ShoppingGuide scenario showed our finger-worn ring augmentation to speed up a shopping list searching task, and provided qualitative insights about how people search and browse physical items with mobile technology providing on-object feedback.

Subjects in our study reported a range of utilization behavior regarding in-situ feedback when browsing or searching for physical-objects. Variation in usage ranged from nearly complete dependence, to using the feedback only as a last resort, suggesting that while any system built to enhance pre-existing physical world search activities may improve task efficiency and be received enthusiastically by users, it must also support varying levels of utilization and be easily ignorable at any time. Furthermore, designers of such systems should be aware that some users will rely heavily on the system, so they must take care to ensure that no serious harm befalls this group in case of system error.

This investigation provides an example implementation for enabling seamless access to digital information associated with physical objects. Our work impacts many application areas that can be improved by enabling physical objects to teach and inform people by leveraging their gestures of attention and direct interaction. These areas include education, logistics, maintenance, repair and shopping, and our work illustrates a path towards providing seamless access to useful personalized information away from the desktop computer.

8 Future Work

In the future it would be interesting to investigate a larger and more realistic supermarket deployment of the My-ShoppingGuide system. Such an investigation could incorporate multimodal feedback, and could collect specific reactions about the social acceptability of the wearable devices. Our ideas for the use of auditory feedback include direction-giving within a store, and verbal explanation of given visual feedback. Care would be required to avoid overwhelming the user with excessive auditory information.

Also, Engine-Info was designed to teach a learner about the parts of an engine but it currently has access only to the ongoing state of a single user interaction. We are interested to incorporate a more sophisticated model of the learner into this interaction, as well as to build support for multi-person interaction. A followup study would help us to better understand the pedagogical potential of this application.

Finally, in this paper we have focused on our prototyping of the interactive user experience, and have not addressed the complexity of real-world deployment of a system like ours. The problems of transponder attachment and lifespan, of recording the connection between the transponder's numeric identifier and the physical object, and of easily creating the corresponding digital annotations are all interesting and challenging. There is much work still to be done towards solutions to these problems for our system and for others like it.

Acknowledgements

Thanks to Professor Martin Culpepper for his loan of the engine that we used for the Engine-Info application. Additional thanks to Mat Laibowitz for his assistance in creating the circuit board and related firmware for the transponders and earpiece, and to Dan Schultz for his work on the application server. Webb Chappell took the photo on the left-hand side of figure 3, also used in figure 2, and Zahir Dossa did a fine job of testing infrared communication range and hand-wiring all of the transponders for Engine-Info. We appreciate the contributions of advice and editing assistance from Sajid Sadi, Enrico Costanza, Cory Kidd, and our colleagues in the Ambient Intelligence group. This work was supported by the Things That Think Consortium and the sponsors of the MIT Media Lab.

References

1. Ishii, H., Ullmer, B.: Tangible bits: Towards seamless interfaces between people, bits, and atoms. Proceedings of CHI'97 (1997) 234–241
2. Google: Google maps for mobile. http://google.com/gmm/ (2006)
3. Rekimoto, J.: Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In: 5th IEEE International Symposium on Wearable Computers. (2001)
4. Want, R., Hopper, A., Falcao, V., Gibbons, J.: The active badge location system. ACM Transactions on Information Systems **10**(1) (1992)

5. Want, R., Schilit, B.N., Adams, N.I., Gold, R., Petersen, K., Goldberg, D., Ellis, J.R., Weiser, M.: An overview of the PARCTAB ubiquitous computing experiment. IEEE Personal Communications **2**(6) (1995) 28–33
6. Want, R., Fishkin, K.P., Gujar, A., Harrison, B.L.: Bridging physical and virtual worlds with electronic tags. In: Proceedings of CHI '99, New York, NY, USA, ACM Press (1999) 370–377
7. Kindberg, T., Barton, J., Morgan, J., Becker, G., Caswell, D., Debaty, P., Gopal, G., Frid, M., Krishnan, V., Morris, H., Schettino, J., Serra, B., Spasojevic, M.: People, places, things: Web presence for the real world. In: IEEE Workshop Mobile Computing Systems and Applications (WMCSA 2000), Los Alamitos, Calif., IEEE CS Press (2000) 19–28
8. Smith, M., Davenport, D., Hwa, H.: Aura: A mobile platform for object and location annotation. In: The Fifth International Conference on Ubiquitous Computing, Seattle, WA (2003)
9. Patel, S., Rekimoto, J., Abowd, G.: icam: Precise at-a-distance interaction in the physical environment. Proc. of Pervasive Computing (2006)
10. Patel, S., Abowd, G.: A 2-way laser-assisted selection scheme for handhelds in a physical environment. Proceedings of UbiComp (2003) 200–207
11. Rohs, M., Gfeller, B.: Using camera-equipped mobile phones for interacting with real-world objects. Advances in Pervasive Computing (2004) 265–271
12. Abowd, G.D., Atkeson, C.G., Hong, J., Long, S., Kooper, R., Pinkerton, M.: Cyberguide: A mobile context-aware tour guide. ACM Wireless Networks **3** (1997) 421–433
13. Cheverst, K., Davies, N., Mitchell, K., Friday, A., Efstratiou, C.: Developing a context-aware electronic tourist guide: some issues and experiences. In: CHI. (2000) 17–24
14. Oppermann, R., Specht, M.: A context-sensitive nomadic exhibition guide. In Thomas, P., Gellersen, H., eds.: Handheld and Ubiquitous Computing: Second International Symposium. Volume 1927/2000., Bristol, UK (2000) p 127
15. Ma, H., Paradiso, J.: The findit flashlight: Responsive tagging based on optically triggered microprocessor wakeup. Proceedings of UbiComp (2002) 160–167
16. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Computer Graphics and Applications **21**(6) (2001) 34–47
17. Rekimoto, J., Ayatsuka, Y.: Cybercode:designing augmented reality environments with visual tags. Proc. Designing Augmented Reality Environments (2000) 1–10
18. Rekimoto, J., Nagao, K.: The world through the computer: computer augmented interaction with real world environments. Proceedings of the 8th annual ACM symposium on User interface and software technology. ACM Press, Pittsburgh, Pennsylvania, United States (1995) 215639.
19. Hoellerer, T., Feiner, S., Terauchi, T., Rashid, G., Hallaway, D.: Exploring mars: developing indoor and outdoor user interfaces to a mobile augmented reality system. Computers & Graphics **23**(6) (1999) 779–785
20. Feiner, S.K.: Augmented reality: A new way of seeing. Scientific American **April'02** (2002)
21. Leibe, B., Starner, T., Ribarsky, W., Wartell, Z., Krum, D., Singletary, B., Hodges, L.F.: The perceptive workbench: Towards spontaneous and natural interaction in semi-immersive virtual environments. In: Proceedings of the IEEE Conference on Virtual Reality. (2000) 13–20

22. Patten, J., Ishii, H., Hines, J., Pangaro, G.: Sensetable: A wireless object tracking platform for tangible user interfaces. In: Proceedings of CHI'01, Seattle, WA, ACM Press (2001) 253–260
23. Shell, J., Selker, T., Vertegaal, R.: Interacting with groups of computers. Communications of the ACM **46**(3) (2003) 40–46
24. Stiefelhagen, R., Yang, J., Waibel, A.: Estimating focus of attention based on gaze and sound. Workshop on Perceptive User Interfaces (PUI01) (2001)
25. Oh, A., Fox, H., Van Kleek, M., Adler, A., Gajos, K., Morency, L., Darrell, T.: Evaluating look-to-talk: a gaze-aware interface in a collaborative environment. Conference on Human Factors in Computing Systems (2002) 650–651
26. Smith, J., Vertegaal, R., Sohn, C.: Viewpointer: lightweight calibration-free eye tracking for ubiquitous handsfree deixis. UIST (2005) 53–61
27. Shell, J.S., Vertegaal, R., Cheng, D., Skaburskis, A.W., Sohn, C., Stewart, A.J., Aoudeh, O., Dickie, C.: Ecsglasses and eyepliances: using attention to open sociable windows of interaction. Proceedings of the 2004 symposium on Eye tracking research & applications (2004) 93–100
28. Labs, S.: Silicon labs. http://www.silabs.com/ (2006)
29. Sourceforge: Cmu sphinx. http://sourceforge.net/projects/cmusphinx (2006)
30. Sourceforge: Freetts 1.2 - a speech synthesizer written entirely in the java$^{\text{TM}}$programming language. http://freetts.sourceforge.net/docs/ (2006)
31. Barbara, J.G., Candace, L.S.: Attention, intentions, and the structure of discourse. Comput. Linguist. **12**(3) (1986) 175–204 12458 0891-2017 MIT Press.
32. Roy, D.: Grounding words in perception and action: Insights from computational models. Trends in Cognitive Science **9**(8) (2005) 389–396

Reach Out and Touch: Using NFC and 2D Barcodes for Service Discovery and Interaction with Mobile Devices

Eamonn O'Neill[1], Peter Thompson[2], Stavros Garzonis[1], and Andrew Warr[1]

[1] Department of Computer Science, University of Bath, UK, BA2 7AY
`{eamonn, s.garzonis, cspaw}@cs.bath.ac.uk`
`http://www.cityware.org.uk`
[2] Vodafone Group R&D, Newbury, UK, RG14 2FN
`peter.thompson@vodafone.com`

Abstract. We investigate the use of 2 tagging technologies: Near Field Communication (NFC) and 2-dimensional barcodes. Our investigation combined a field trial and interview based study with an experimental evaluation. The field trial focused on users' experience of the usability of NFC for a range of trial services, users' perceptions of NFC use in their daily life context, and users' suggestions for potential applications of NFC. The tags were embedded in a variety of postcards, table-top signs and posters. The experimental evaluation compared the ease of use of NFC and 2D barcodes, operationalised in terms of time taken to read a specified sequence of tags on posters. We found that for untrained users the 2D barcodes were quicker to use than the NFC tags but that training significantly improved users' performance with the NFC tags while having no effect on their performance with the barcodes.

1 Introduction

The use of ambient tags read by mobile devices presents new possibilities for service discovery and user interaction. Service discovery is an increasingly important issue as we move towards realising pervasive systems. How does the user identify and access a particular service from the plethora of services that may potentially be available around him at any moment? Having discovered a service, a second important issue is how the user interacts with it. Tags in the environment may be used to provide fast, zero-configuration service discovery and support interaction between mobile and fixed devices in pervasive systems [8]. Tags may be used to store small amounts of data on the tag itself. However, more typically, the tag may store a link to more sophisticated information and services. The link may connect the tag-reading device to a URL, email address or phone number, or may provide a unique identifier as a reference to information or services. Two very different tagging technologies, 2D barcodes and NFC, are each beginning to see widespread interest but as yet there are few reported studies of either in use [10].

2D barcodes can be easily produced and printed. They can store more information than the currently ubiquitous 1D barcodes but, unlike 1D barcodes, have not yet been standardised. Recent software applications enable any device with a camera, such as a mobile phone, to read 2D barcodes with no additional hardware required. A 2D

A. LaMarca et al. (Eds.): PERVASIVE 2007, LNCS 4480, pp. 19 – 36, 2007.

barcode may be printed on a surface or on an adhesive tag which may in turn be attached to a surface. Alternatively, a 2D barcode may be displayed on a device's screen while another device is used to read the tag. A 2D barcode provides one-way communication. The barcode image encodes data. The mobile phone's camera scans the image and the reader application decodes the data from the image. Typical uses include stock control but also e-ticketing (e.g. airline boarding passes) and triggering mobile data services from a location or object [e.g. 4]. [12] used a form of 2D barcode on posters and active displays to establish wireless connections between mobile phones and site-specific services. This work integrated previous work [11] that used the barcodes for device discovery to bypass the Bluetooth scanning protocol, significantly reducing discovery time.

NFC is a short range wireless technology, similar to other RFID technologies but with additional features (see www.nfc-forum.org). NFC devices operate in reader/writer mode, peer-to-peer mode, or card emulation mode, based on the ISO/IEC 18092 NFC IP-1 and ISO/IEC 14443 standards. In reader/writer mode, an NFC device can read or write NFC tags similar to the uses of standard RFID tags. In peer-to-peer mode, two NFC devices can directly exchange data, simply by touching the devices together to initiate a wireless link. NFC can be used to initiate a channel between devices and then hand over to complementary wireless technologies such as Bluetooth and WiFi. In card emulation mode, an NFC device appears to an NFC reader as an NFC tag with secure data storage capacity, allowing the device to be used for applications such as contactless payment and eticketing.

The market for RFID systems is currently driven by traditional applications such as access control, car engine immobilization, automated toll collection, and stock control and transportation. NFC has the potential to drive new applications and markets [13]. NFC devices are currently in use in the form of simple payment cards like the Oyster card used by Transport for London in the UK. In the US, NFC use has so far been confined to the development of NFC-enabled credit cards. The major mobile phone manufacturers are committing to NFC. Nokia got to market first with an NFC shell for its 3220 phone, which we used in the studies reported here. Although the use of NFC-equipped mobile phones is currently still in its infancy, a study by ABI Research predicts that 50% of all mobile phones will support NFC by 2009, with sales of NFC phones exceeding 500 million by 2010[1]. The study also predicts that by 2007, NFC deployments will be common in many kinds of consumer electronics, from computers to cameras, printers, set-top boxes etc.

NFC tags and 2D barcodes require slightly different use. NFC requires the user to bring her device very close to the tag. 2D barcodes require the user to point her camera at the tag (and with some readers, take a photo). NFC offers secure data storage features that 2D barcodes cannot offer. In general, however, NFC and 2D barcodes support many of the same services. Over the next few years, it seems likely that NFC and 2D barcodes will offer alternative means of accessing some of these services, such as eticketing (where the data may be stored securely in an NFC device or

---

[1] ABI Research (2005) Near Field Communications (NFC): simplifying and expanding contactless commerce, connectivity and content. Report RR-NFC. Available at http://www.abiresearch.com/products/market_research/Near-Field_Communications_(NFC). Last accessed 13 October 2006.

displayed as a barcode on the user device's display, in either case being read by an external reader) and service discovery (where NFC tags or 2D barcodes are printed on or attached to objects in the environment).

In the work reported here, we investigated the use of tagging to support service discovery and interaction with mobile phones using NFC and 2D barcodes. In the first study, we conducted a field trial using NFC in which most of the services offered, other than writing personalized tags directly from the phone, could equally have been supported by 2D barcodes. In the second study, we directly compared the use of NFC and 2D barcodes in a lab-based experimental evaluation. Our field trial provided empirical data on the popularity of various tag based services, user feedback on the use of NFC phones and tags, and suggestions for NFC applications. The field trial did not attempt to distinguish between NFC and 2D barcodes as appropriate tagging technologies. This was investigated in our second study, in which we conducted an experimental comparison of NFC and 2D barcode use.

2 NFC Field Trial

The main objective of the field trial in our first study was to gather user feedback on the use of NFC to discover and interact with mobile services. In particular, we were interested in (i) users' experience of the usability of NFC for a range of trial services, (ii) users' perceptions of NFC use in their daily life context, and (iii) users' suggestions for potential applications of NFC.

2.1 Participants

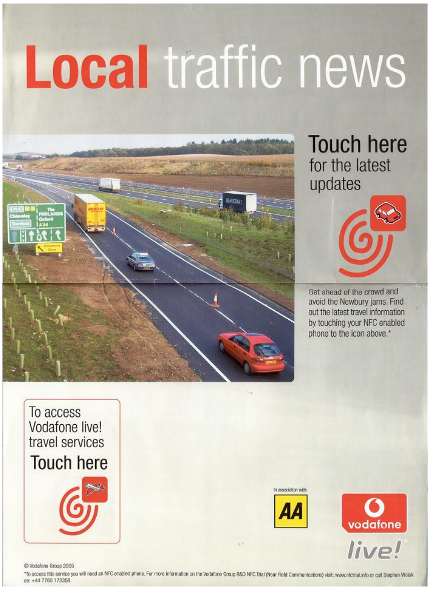
A total of 60 Vodafone employees from various departments volunteered for the trial. Ages ranged from 22 to 51 years of age, with a median of 32 and a mean of 31. Several volunteers were unavailable at short notice so only 55 participants completed the opening questionnaire, 36 male and 19 female. Throughout their regular workdays the participants are exposed in varying degrees to the development, use or marketing of mobile phone technologies. They are relatively knowledgeable technically and are experienced in using new features on mobile phones.

2.2 Equipment

A total of 1226 NFC tags were used in the field trial. Prior to the study, table-top signs and A3-size wall posters were distributed in 40 locations across six buildings, including business centres, meeting areas and kitchens. NFC tags were embedded in the posters and signs. The services accessed by reading a tag ranged across the services offered by mobile phone operators: voice, data and SMS. There were 90 table-top signs, each containing tags with links via URLs to Vodafone Live! sport, music and games sites. 16 of the posters had a variety of tags. For example, the "Local traffic news" poster (see Figure 1) had a tag that triggered an SMS request for information on local traffic conditions. The response was sent by SMS to the user's phone. A second tag on this poster linked to the Vodafone Live! traffic news site.

A "Music" poster contained a tag that linked to the Vodafone Live! music site, and a second tag that entered the user in a prize draw competition with a CD as the prize.

A "Sport" poster contained a tag that linked to the Vodafone Live! sports site, and a second tag that entered the user in a prize draw competition with tickets to a football game as the prize. A further 20 posters contained tags that formed a "treasure hunt". Participants who managed to read all the tags embedded in the posters with a treasure hunt icon were entered in a prize draw for a miniature Ferrari.



Get ahead of the crowd and avoid the Newbury jams. Find out the latest travel information by touching your NFC enabled phone to the icon above.*

© Vodafone Group 2005

*To access this service you will need an NFC enabled phone. For more information on the Vodafone Group R&D NFC Trial (Near Field Communications) visit: www.nfctrial.info or call Stephen Wolak on: +44 7760 170358.

Fig. 1. A poster with embedded NFC tags

A trial pack was also given to each participant at the start of the trial. The pack included a Nokia 3220 phone with NFC shell, a set of postcards each with an embedded tag, and an introductory letter with 2 embedded tags. One linked to a WAP-based site offering information about the trial and NFC. The other triggered a voice call from the user's phone to a "trial helpline" supported by the trial administrators. One postcard tag downloaded free wallpaper for the user's phone. The wallpaper offered changed daily. Another postcard tag downloaded a free ringtone for the user's phone. Similarly, the ringtone offered changed daily. Three "Keep in touch" postcards had "blank" tags. Participants were encouraged to write their own content (e.g. text messages, URLs or phone numbers) to these tags and leave the postcards around for others to find and read. Finally, a "directory assistance" postcard triggered a voice call from the user's phone to the main Vodafone directory assistance service.

2.3 Procedure

The trial lasted 9 days, beginning on a Thursday and running through to the following Friday. On the first day, the participants were given an introductory briefing and a trial pack. 55 of the participants completed a pre-trial questionnaire. This covered the participant's usage of mobile phones, including voice, text and data services and phonecam use; their usage of the Internet, including email and web use; and their wishes for new mobile applications and features.

The services were activated from the first day of the trial (Thursday) with the exception of the "treasure hunt" which began on the fifth day (Monday). The trial administrators and 17 "wizards" (staff who were very familiar with the technologies) were available to provide help and advice to the participants throughout the study.

The data services were delivered via an NFC server implemented for the trial. However, this server was not involved in interactions with the travel information SMS tags, since the SMS transmissions bypassed the NFC server, or the interactions that accessed a voice service (i.e. reading the "trial helpline" and "directory assistance" tags), in which a voice call was placed directly to the service provider.

At the end of the trial, 47 of the participants completed a post-trial questionnaire. This covered their experiences during the field trial, and their comments and suggestions for future applications and uses of NFC based services. The questionnaire was followed up with an interview.

2.4 NFC Field Trial Results

2.4.1 Tag Usage

The NFC server log provided a quantitative measure of a subset of the participants' interactions with the NFC system, i.e. excluding voice calls and SMS messages. A total of 2005 tag interactions by participants were logged by the server during the first 8 days of the trial. (Tag events were not logged for the final day of the trial.) Table 1 shows the number of *different* tags accessed by a participant, i.e. not counting repeated accesses of a particular tag by the same participant.

Table 1. Distribution of data service interactions with different NFC tags

Distribution of tags	Number of participants
0	04
1-10	35
11-20	16
21-30	02
31-40	03

Participants read a high number of tags in the first 2 days after receiving their NFC phones and trial packs (Thursday and Friday). 64 tag events were recorded during the weekend despite a network problem that became apparent on Friday. The trial administrators sent an email early on the Monday morning, informing the participants that the problem had been fixed. Over half the total tag interactions were then logged during the Monday and Tuesday. Tag events decreased over the next two days (Figure 2).

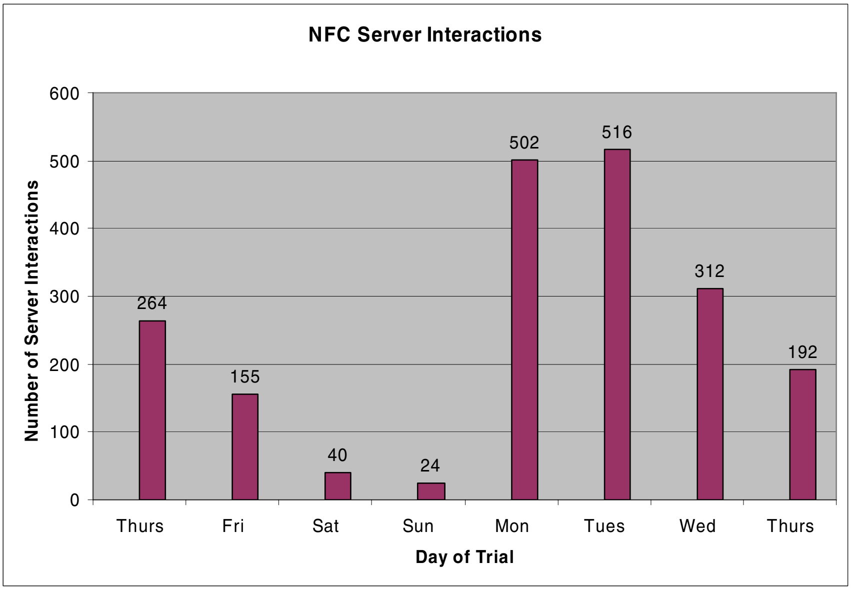


Fig. 2. NFC server events per day

Table 2 compares the use of the various tags for the days when the treasure hunt was active. Content downloads (free daily wallpapers and ringtones) were the most popular tags, followed by tags that accessed the Vodafone Live! information sites.

Table 2. Tag usage for days when the treasure hunt was active

	Min/day (Mon-Thu)	Max/day (Mon-Thu)	Total (Mon-Thu)
Content downloads	58	168	464
Vodafone Live! information	34	166	404
Competitions	22	68	174
Treasure hunt	64	194	398
Trial Information WAP site	2	36	82

2.4.2 User Experience with NFC Trial Services

The participants' extensive experience with mobile phones helped make them comfortable using the NFC phones. Occasional problems with the software were overcome by restarting the phone, participants developed workarounds for unimplemented menu options, and NFC tag reading was approached with more patience than might be the case with less experienced users.

Although 96% of participants regarded the NFC symbols indicating the locations of the tags on the posters, signs and postcards as easy to find and identify, and 90% claimed that the location of the NFC reader in the phone was obvious, only 40% felt that the use of the reader was clear from the start. Most reported that they had to experiment with reading tags, as it was not clear if one had to hover, slide, wave or press the reader on to the tag.

After an initial period of experimenting with the NFC interaction process, participants developed their own ways of reading a tag. Different users reported different techniques, including touching the tag with the top of the phone, sliding the phone across the tag, and holding the phone 3 cm from the tag. Participants became convinced that their device responded best to a particular angle, motion or distance from the tag. After the initial experimentation, most participants found the process of reading tags straightforward and intuitive, although 65% of participants reported occasional problems reading tags, 26% reported frequent problems and only 9% reported no problems.

25% of participants regarded the speed of accessing services as good, 37% as adequate and 37% as too slow. These should be interpreted as the perceptions of very experienced mobile phone users, accustomed to delays in downloading data services on pre-3G networks.

Some participants found that the screensaver and keyboard lock features on the phone interfered with the use of NFC. One could read a tag with the screensaver or the keyboard lock enabled, but had to unlock the keyboard before the corresponding service could be accessed or viewed on the screen. Given that reading an NFC tag is intended to require an explicit user action, participants felt that reading a tag should automatically turn off the screensaver and keyboard lock.

Half of the participants wrote to their “Keep in touch” tags, mainly storing contact details or URLs on the tag. The remaining participants did not write to the tags. The follow up interviews revealed that some of these participants attempted to but were unable to complete the tag writing successfully.

Privacy and security concerns arose during the trial and were discussed during the post-trial interviews. While some participants did not perceive any security problems with the technology, others were more concerned. Many of the latter were wary of the “always on” status of the NFC reader in the phone. At times the NFC phone would unintentionally interact with nearby tags, e.g. when a tagged postcard was next to the phone in the participant’s bag.

Participants were also concerned that tags could advertise one service, for example through a poster, while actually connecting to another, perhaps at a premium rate. This could be perpetrated deliberately by the producer of the poster. Even more insidiously, a poster offering a legitimate service could be subverted by having its tag rewritten. Participants were also concerned that rewriting tags could be used to create a nuisance for third parties via the reader’s phone. By writing a message to a tag that included the third party’s number, one could encourage people reading the tag to phone the third party, with potentially unwelcome results. Worse still, one could write the tag to trigger a call directly from the reader’s phone to the third party. Perhaps more subtly, one participant suggested, “I could write someone else’s name to a tag and link it to an inappropriate website or include an impolite message and leave it for his manager [to read the tag]”.

2.4.3 NFC in the Users’ Daily Life Context

The potential of a tag-based system to identify precisely an individual’s location and activities is sometimes viewed as a threat to people’s privacy [1; 6]. Our participants regarded this issue as relatively unimportant. Participants felt that the devices they currently use – swipe cards, computer log-ins and so on – already record their locations and activities sufficiently closely that reading some NFC tags would not introduce a significant extra opportunity for monitoring their behaviour. To paraphrase Scott McNealy, these users have lost their privacy and got over it.

Participants were more concerned with how the use of NFC readers in public spaces made them appear to other people around them. Many participants noted that they felt awkward at first using NFC due to the very explicit public act of reaching out and touching a tag embedded in a poster or sign. However, despite describing it as "odd" to use at first, many participants lost their reservations about using NFC over the course of the trial.

While the process became more familiar to participants, their interactions with the tags remained objects of curiosity to many of their colleagues. (This became clear when the trial administrators began to receive requests to be included in the trial from people who had noticed the participants reading tags.) The very requirement to make NFC tags both noticeable and accessible makes the reader-tag interactions public.

When asked if they would use NFC in public places, 13 participants responded that the nature of the place was a determining factor. In high traffic public places where people often seek information (e.g. an airport or train station concourse), they would use NFC. In low traffic areas they would be less likely to use NFC. Even in high traffic places, participants suggested that the interaction time with tags – and potential for consequent interaction with other people in the vicinity – should be minimised. The location of some tags during the trial was regarded as intrusive. In cases of repeated tag reading errors, the trial participant would be blocking the access to a photocopier, for example, thereby interfering with other people's work. And of course if many people wanted to interact with a particular tag, long interaction times would cause another form of blocking, particularly in high traffic places.

In addition to the problem of blocking others' movements and activities, participants voiced another concern with the use of NFC in public places. When an NFC user buys a ticket for a concert, for example, by touching a public poster, a bystander observing the interaction could in principle acquire information about the user's shopping and leisure preferences. However, it seems unlikely that a would-be profiler following NFC users around would be able to gather sufficient reliable simply from observation.

Many participants were concerned that reaching out to read a tag in public was openly advertising one's phone, potentially catching the attention of thieves. There was a perception amongst participants that busy places are potentially high crime areas, and using NFC in such places was regarded as potentially dangerous. In response to this perceived threat, participants suggested that tags might be placed where users could read them close to their body, without reaching out.

2.4.4 Suggested NFC Applications and Services

In addition to providing feedback on the trial services, participants suggested potential applications for NFC. In addition to existing NFC services, such as mobile electronic payment, their suggestions included applications in:

- access control;
- using mobile services; and
- data storage, retrieval and transfer.

Participants commented that they currently gain access to their offices, meeting rooms etc by the use of passive RFID swipe cards. NFC was seen as a valuable alternative, the notion of carrying one device fewer appealing to participants. Embedding access control into the mobile phone was regarded as adding a level of security since our participants believed that people neither leave their phone at home very often nor lend their phone to other people. The little empirical evidence that has been published tends to support this assumption [7]. The risk of losing the phone, and the consequent ability for others to access one's office, was initially viewed by some participants as dangerous. However, they noted that losing a swipe card has the same inherent problem. With NFC, there is the potential to add security features such as deactivation of the sim, access rights, credit card etc with a single action.

A typical NFC service proposed by the participants was ordering a taxi by reading a dedicated tag in a restaurant or shop. Similarly, using tags at bus stops to invoke an SMS message giving the arrival time of the next bus was proposed. At Vodafone HQ, a similar service already exists for the shuttle bus to nearby Newbury, but it is invoked by an SMS sent by the user. Many participants commented that they do not use this service because they do not know the number to which to send their SMS. A shuttle-tag in the HQ lobby could be used to inform users of the time of the next shuttle's departure, simply by touching the tag.

The most commonly suggested application of NFC was contact information management, with the NFC phone being used to store and transfer business card details. In one proposal, an NFC tag could be embedded in a business card, with the recipient of the card reading the tag to transfer the data to her phone's contacts list. Another proposal was the direct interaction of two NFC phones. By touching the phones to each other, the users' contact details could be transferred between the phones.

Another proposal for managing and using data was to "bookmark" physical objects with tags. Rather than trying to remember the details of a friend's CD, one could simply read its tag and store it in a "Favourites" folder on the phone. This information could then be used to link to, for example, online retailers or review sites.

Overall, our field trial gave us some interesting empirical data on the popularity of various tag based services amongst our participants, provided user feedback on the use of NFC phones and tags, and provided suggestions for NFC applications. Almost all of these services and applications could be supported by either NFC tags or 2D barcodes, or indeed a combination of both. (The exceptions include electronic payment where the NFC technology is augmented with secure data storage, and the facility to write directly to tags from the phone.) The field trial did not distinguish between NFC and 2D barcodes as appropriate tagging technologies. Our second study filled this gap with an experimental comparison of NFC and 2D barcode use, in which we employed 2D barcodes with the same physical dimensions as the NFC tags.

3 Experimental Comparison of NFC and 2D Barcodes

Our experimental comparison of NFC and 2D barcodes reflected the use of posters with embedded tags in our field trial. However, in order to control confounding variables such as poster content, the posters in the experiment were as simple as possible, containing just the number of the poster and the embedded tags (see Figure 4). We were interested in the ease of use of the tags themselves, comparing NFC against 2D barcodes, rather than people’s experience with the poster content and linked services that were the focus of our field trial.

Therefore, we provided the same, very simple “service” from every tag. Successfully reading a tag provided the user with an instruction on the phone to move on to the next tag. The main difficulty with using the tags in our field study, and in the pilot for our experimental study, was repeated unsuccessful attempts to read a tag. Hence, time taken to read the complete set of tags in the correct sequence was a useful and reliable quantitative comparison of the 2 forms of tagging.

In both the field study and the experimental pilot, we experienced the problem of users not knowing the precise location of the NFC reader in the phone. With multiple adjacent NFC tags, this resulted in users repeatedly scanning the wrong tag because they aimed the wrong part of the phone at the target tag, causing the reader to read an adjacent tag. With multiple adjacent 2D barcodes, the reader had difficulty in interpreting a single tag when parts of more than one barcode appeared in the camera’s viewfinder, so distance to the tag and image composition became important.

To investigate these effects in our main experiment, we arranged multiple tags in different patterns on the posters and varied which of the group of tags the user was required to read.

In our experimental pilot, we found that the fastest 2D barcode reading was achieved with the phone approximately 100mm from the tag, in contrast to the optimal distance of about 20mm found with the NFC tags. The angle of the phone also had a marked effect on 2D barcode reading. The phone had to be reasonably parallel to the plane of the barcode for successful reading (although this problem could be mitigated in software by the reader). To investigate this effect in our main experiment, we arranged the posters at different heights to change the angle of incidence to the tags.

Overall, our pilot study suggested that using 2D barcodes was at first easier than using NFC, with fewer of the initial difficulties and need for experimentation with the NFC reader also found in our field study. However, the difference in usability between the 2 technologies lessened quite quickly as users became more experienced with the NFC reader and settled on a particular way to handle and aim the reader. This reflected our findings from the field study.

3.1 Experimental Design

Our experiment had a mixed factorial design. There were two factors, each of which had two levels. The independent variables were: (i) whether 2D barcodes or NFC tags were used; and (ii) whether the participants were given training or not. The dependent variable was the time it took a participant to scan all the indicated tags in the correct sequence. It was predicted that for untrained users the 2D barcodes would be quicker to use than the NFC tags (H1). It was also predicted that training would significantly improve users' performance with the NFC tags (H2).

3.2 Participants

Twenty-eight participants took part in the experiment, 20 male and 8 female. The participants varied in age from 19 to 30 years of age, with a mean age of 21 years (SD 2.8). The participants were all university students. The participants were recruited via mailing lists and word of mouth, and volunteered to participate in the experiment.

3.3 Equipment

The NFC tags and the 2D barcodes were located on the bottom of an A3 poster. 12 posters were attached to a wall in a usability lab (Figure 3). Each poster had a large number printed in the centre ranging from 1 to 12. Each poster had either 1 or 3 2D barcodes in the bottom left corner and the same number and layout of NFC tags in the bottom right corner.

The posters were arranged in numerical order in four columns and three rows. The numbering of the posters went from 1 in the top-left corner through to 12 in the bottom-right corner, in an across-then-down order. Posters in a row were placed so that the target tags were horizontally aligned with each other, to present consistent target heights of 210, 150 and 90 cm. 150cm represented a comfortable reaching out height for a standing person. Based on our pilot study findings, 210cm and 90cm were used

Fig. 3. Reading an NFC tag on a poster

to test usability for standing users when presenting the phone close to and parallel to the plane of the tag was less easy than at shoulder height. 90cm also represented a standard height for wheelchair-accessible facilities such as ATMs. 210cm also reflected experiences of regulatory requirements on the placement of NFC tags in a wide area installation [4].

Across the four columns a variety of tag arrangements was used. As noted above, pilot testing revealed that different arrangements of tags could cause erroneous interactions. The arrangements of tags in our main experiment obliged the user to select the appropriate tag and not merely wave the tag reader around until a tag was read. The target tag in each group of 3 tags was indicated by an adjacent arrow. The groupings of the tags are illustrated in Figure 4, with the target tag highlighted in each group.

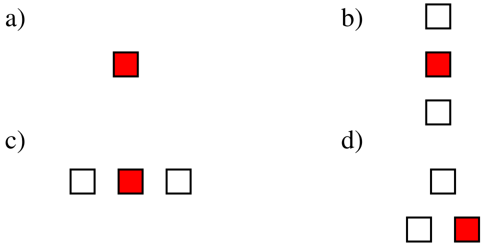


Fig. 4. The arrangements of NFC tags and 2D barcodes on the posters

The 2D barcodes were read using a Nokia 6600 mobile phone running the Glass 2D barcode reader application (http://www.activeprint.org/). The NFC tags were read using a Nokia 3220 with an integrated NFC reader. Ideally we would have liked to use the same phone, however hardware incompatibilities prevented this. The differences between the phones were negligible for our purposes.

When running the Glass application, the view from the camera is shown on the phone's display. When the camera is pointed at a 2D barcode, the application automatically decodes the information held in the barcode and displays it on the screen of the phone as a scrolling message. In our experiment, the message told the user which poster to go to next.

To read the NFC tags, we used Nokia's Service Discovery application. When this is running, it displays the message "Please touch tag" on the phone's display. When the reader is placed close to an NFC tag, the application reads the information held on the tag. This is then displayed as a message on the phone's screen. As in the 2D barcode condition, the message told the user which poster to go to next. Because the Service Discovery application is designed to connect to a URL, the user is also given options to "Connect" or to "Cancel" each time a tag is read. For the experiment, we limited the user option to just "Cancel". Pressing the "Cancel" button returned the phone to "Please touch tag", ready to read the next tag.

Using either of the technologies, when a tag was read that was not the current target tag, a message appeared on the phone's display informing the user that an incorrect tag had been read.

During the experiment, 3 video cameras recorded the trials. 2 cameras built in to the usability lab captured a wide and a high view respectively of the user interacting with the tags, while a tripod mounted camera captured the view from behind the user facing the posters on the wall. An observer also took notes during the trials.

3.4 Procedure

Participants were run individually. At the start of each trial, the participant was given a phone and instructions on how to complete the activity. The instructions were read from a script by one of the experimenters. Fourteen of the 28 participants were given training. Training included describing to the participant how the technology worked. The participant was introduced to the phone and told how the tags were to be read, including details of where the readers were in the phones, as well as optimum distances and angles for tag reading. The experimenter demonstrated how to read a tag using the reader on the phone. The trained participants were also given practice runs with a practice poster.

The 14 untrained participants were told only that the reader was on the back of the phone (i.e. the camera or the built-in NFC reader). They were instructed to read the target tag as quickly as possible and move on to the next tag when instructed by a message on the phone's display.

Each participant started at poster 5 and obtained the number of the next poster by scanning the appropriate tag. For example, reading the target tag on poster 5 displayed the message "Go to poster 2". When the target tag was successfully read, the participant read aloud the number of the poster they were to go to next and moved on. The target tags instructed each participant to read the tags in the order 5 > 2 > 11 > 7 > 4 > 12 >10 > 6 > 9 > 1 > 3 > 8. The target tag on poster 8 instructed the participant that she was finished. An error was recorded whenever the participant successfully read a tag that was not the current target tag. In this case, the participant continued attempting to read the target tag until it was successfully read.

After a participant had traversed all 12 posters, she completed a questionnaire. Taining in the other tagging technology was then given (dependent on the trained/untrained condition) and the trial was repeated with the other tagging technology. The order in which the tagging technologies were used was counter-balanced. After each participant completed her trials with both tagging technologies, she was given a third questionnaire to compare her experiences of the two technologies.

3.5 Experimental Results

The dependent variable was the cumulative time taken from the first tag being read on poster 5 to the target tag on the final poster being read. Figure 5 shows the total times (in seconds) for NFC and 2D barcode use by the untrained participants.

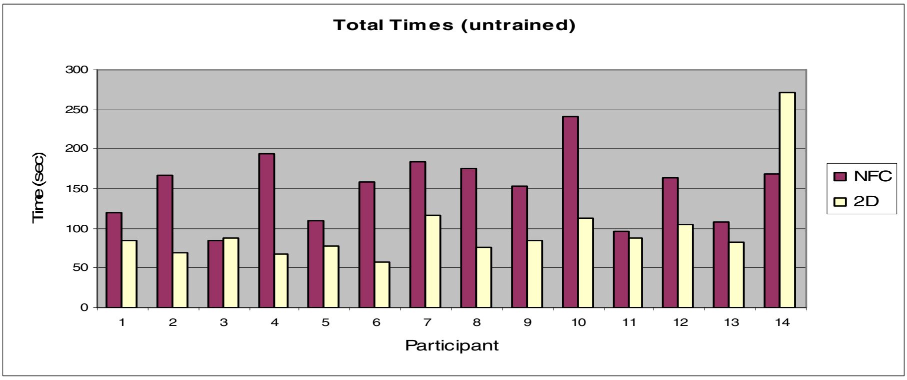


Fig. 5. Times (in seconds) for untrained participants

Figure 6 shows the total times (in seconds) for NFC and 2D barcode use by the trained participants.

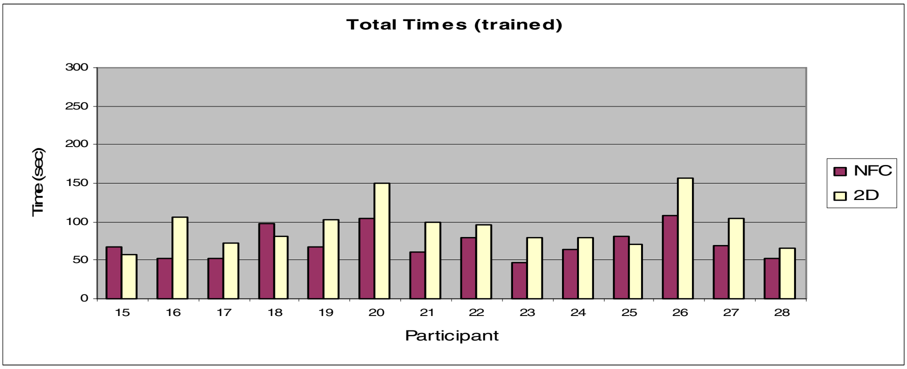


Fig. 6. Times (in seconds) for trained participants

Table 3 shows the mean (and SD) time to read all 12 tags using each technology, with and without training.

Table 3. Mean (and SD) time to read all 12 tags for both the 2D barcodes and NFC tags, with and without training

	NFC	2D barcodes
Untrained	151.50 (43.33)	98.43 (52.42)
Trained	71.71 (20.01)	94.57 (29.18)

A mixed 2-way ANOVA for the data in Table 3, with tagging technology and training as the independent variables, found a significant effect ($F_{3,24} = 16.67$, $p = 0.000154$). In order to determine the direction of the differences, post hoc Tukey tests were performed.

A Tukey test for the main effect of tagging technology found a significant difference between NFC and 2D barcodes for untrained users ($T = 5.18$, $Q = 4.7$, $T > Q$, $p < 0.01$), supporting hypothesis 1, while there was no significant difference between NFC and 2D barcodes for trained users ($T = 2.23$, $Q = 3.79$, $T < Q$, $p > 0.05$).

When using NFC, there was a significant difference between the training and no training conditions ($T = 7.79$, $Q = 4.7$, $T > Q$, $p < 0.01$), supporting hypothesis 2, while there was no significant difference between the training and no training conditions when using 2D barcodes, ($T = 0.38$, $Q = 3.79$, $T < Q$, $p > 0.05$).

So, both experimental hypotheses were supported. Without training, using the NFC tags took significantly longer than using the 2D barcodes (H1). With training, users' performance with the NFC tags significantly improved (H2), to the point at which there was no difference between using 2D barcodes and NFC. Training had no effect on users' performance with 2D barcodes.

The observations of the trials, the video recordings and the post-trial questionnaires provided supporting qualitative data. On a Likert scale (1: very difficult, 5: very easy), the untrained participants found it easier to use 2D barcodes than NFC. This finding was consistent across the immediate post-trial questionnaires (NFC: 3.8, 2D barcode: 4.6) and the third questionnaire (NFC: 3, 2D barcode: 4.1). When asked which phone they would pick if they had to repeat the task, almost all the untrained participants stated that they would prefer to use the 2D barcode technology. Not knowing which part of the phone to touch on the tag was the main frustration reported with the NFC phone.

In contrast, the trained participants found it easier to complete the task using NFC rather than the 2D barcodes. Again, this was consistent across the immediate post-trial questionnaires (NFC: 4.1, 2D barcode: 3.9) and the third questionnaire (NFC: 4.1, 2D barcode: 3.9). When asked which phone they would pick if they had to repeat the task, almost all the trained participants stated that they would prefer to use the NFC technology. The main complaint about the 2D barcode phone was the requirement to align the phone with the 2D barcode, in contrast to just touching the tag with the NFC phone.

When asked in the third questionnaire "Which technology did you find easier to use?", only 3 of the 14 untrained participants said they found NFC easier, while 11 of the 14 trained participants said they found NFC easier.

Video analysis showed that while many untrained users struggled to use the NFC phone, many developed more or less successful techniques in their attempts to get the device to work. Some users made quick stabbing motions at the tag, rather than holding the phone over the tag until it was read. Such interactions meant that many attempts were needed to read the tag and this contributed to increasing the overall time taken. Some participants carefully held the phone horizontally, apparently believing that this facilitated tag reading.

One participant found it extremely difficult to read any 2D barcodes on the bottom row of posters. The video record suggested that his problems were due to not holding the phone at an appropriate angle to compensate for aiming the phone's camera downwards (thereby moving it away from parallel to the plane of the tag).

Posters 3, 7 and 11 caused particular problems for the untrained participants using NFC, with between 20 and 30 errors (i.e. reading non-target tags) for each of these posters. The arrangement of tags on these posters was that shown in Figure 5 (b). This arrangement led to participants repeatedly reading the wrong tag. Training reduced the errors, contributing in turn to significantly reduced task completion times.

4 Conclusions and Future Directions

Our field trial and our experimental pilot study suggested that users were unsure of the location of the NFC reader in the mobile phone. In the field trial and the experimental evaluation, users developed their own techniques for successful tag reading, although these often involved manipulating or holding the phone in ways that were not actually necessary. Our experimental evaluation tested the hypothesis that an initial training period would overcome users' lack of familiarity with the NFC technology, leading to better performance. This was found to be the case. Initial training had no effect on users' performance with the 2D barcodes. It seems likely that this is partly due to users' previous familiarity with cameraphones, and partly due to the obvious location of the camera on the phone and the visual feedback on the display of the camera image. These factors are likely to have contributed to users' performance being at a level that was unaffected by the brief training offered. It is likely that longer term familiarity with NFC would produce much the same effect for that technology.

Participants in the field trial reported feeling awkward or embarrassed reaching out to use the tags in a public place. Similarly, in a user study of service discovery and interaction using mobile phones and RFID tags, [9] found that users were reluctant to use this from of interaction outside their own homes "because they were embarrassed about touching RFID tags in public places. For example, they felt that touching might seem too eye-catching on the street or at a shopping mall. So, although touching is itself a natural action, it isn't necessarily considered a socially acceptable way to use new technology" [p. 45]. However, even during the limited course of our field trial, participants – and their colleagues – became more accustomed to tag-reading in public. Not so long ago, voice calling and texting in many public places would have seemed unusual but they have rapidly become appropriated. There is still a frisson of the weird about seeing a user of a discreet hands-free headset wander along, talking animatedly apparently to thin air, but that too is rapidly becoming a commonplace. It seems unlikely that it will take long for public tag reading to be similarly appropriated.

[9] also found that users were unwilling to read tags in public places "because they considered the security risk too high" [p. 45]. Again, this was reflected in our participants' concerns. Mobile phone theft accounts for about half of all street crime in the UK (http://www.met.police.uk/mobilephone/) but there is evidence that high traffic places are often safer than low traffic places [3], suggesting that the actual threat may be perceived differently from the real threat. Reading a tag may be slightly more risky than calling or texting in a public place only in so far as one is stationary while reading the tag. Along with the issue of long interaction times blocking others' access both to the place and to the tag, this would suggest designing to minimize interaction

time with the tag, retaining most of the interaction on the phone via a wireless connection to services initially identified by the tag reading event.

The concern of some users about unintentional interactions with NFC tags suggests that users should have more control of the NFC functionality so that, at least, they can turn it on and off when desired. The issues of impersonating others and rewriting tags may suggest a requirement for lockable or “read-only” NFC tags. However, disabling write access to a tag would disable some of the intrinsic advantages of NFC. 2D barcodes, at least in their printed form, will require a different approach to security since they may be overlaid or replaced quite easily.

In proposing access control as an application of NFC, our participants voiced no concerns about its potential role in monitoring their movements. This is in contrast to the commonly reported fears of privacy being eroded by tagging systems [2]. Whether or not a tag based, or indeed swipe card based, access control system logs events is a design decision. Such decisions and users’ acceptance of such monitoring will depend at least in part on how much they value the services offered.

Our participants’ most commonly suggested application of exchanging contacts data by touching 2 phones together is the functionality that raises NFC beyond a simple tagging technology. In [5] we report an application that does this, comparing the contrasting influences on the user experience of using NFC and Bluetooth. The changing costs of tags and barcode production, the production of phones with appropriate readers and, ultimately, the development of useful, usable applications and services will influence the trajectory of tagging technologies, but it seems likely to be upward.

Acknowledgements. We thank Nokia, and Jason Cooper, Nilusha Piyasena, Trevor Richards, Mickael Vaillant, Craig Webb and Jan Kietzmann.

References

1. Garfinkel, S.L., Juels, A. and Pappu, R. (2005) RFID privacy: an overview of problems and proposed solutions. IEEE Security & Privacy 3 (3), 34-43.
2. Günther, O. and Spiekermann, S. (2005) RFID and the perception of control: the consumer’s view. Commun. ACM 48 (9), 73-76.
3. Hillier, B. and Shu, S. (2000) Crime and urban layout: the need for evidence. In: Ballityne, S., Pease, K. and McLaren, V., Secure foundations: key issues in crime prevention, crime reduction and community safety, London, IPPR.
4. Kindberg, T. and O’Hara, K. (2006) “You are here” – experiences with located systems. HP Labs Technical Report HPL-2006-120. http://www.hpl.hp.com/techreports/2006/HPL-2006-120.pdf.
5. Kostakos, V., O’Neill, E. and Shahi, A. (2006) Building common ground for face to face interactions by sharing mobile device context. LOCA 2006, LNCS 3987, 222-238.
6. Ohkubo, M., Suzuki, K., and Kinoshita, S. (2005) RFID privacy issues and technical challenges. Commun. ACM 48 (9), 66-71.
7. Patel, S.N., Kientz, J.A., Hayes, G.R., Bhat, S. and Abowd, G.D. (2006) Farther than you may think: an empirical investigation of the proximity of users to their mobile phones. Ubicomp 2006, Springer LNCS 4206, 123-140.

8. Pering, T., Ballagas, R., and Want, R. (2005) Spontaneous marriages of mobile devices and interactive spaces. Commun. ACM 48 (9), 53-59.
9. Riekki, J., Salminen, T. and Alakarppa, I. (2006) Requesting pervasive services by touching RFID tags. IEEE Pervasive Computing, 5 (1), 40-46.
10. Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A. and Chin, J. (2006) An experimental comparison of physical mobile interaction techniques: touching, pointing and scanning. Ubicomp 2006, Springer LNCS 4206, 87-104.
11. Scott, D., Sharp, R., Madhavapeddy, A., and Upton, E. (2005) . Using visual tags to bypass Bluetooth device discovery. *SIGMOBILE Mob. Comput. Commun. Rev.* 9, 1 (Jan. 2005), 41-53.
12. Toye, E., Sharp, R., Madhavapeddy, A. and Scott, D. (2005) Using smart phones to access site-specific services. IEEE Pervasive Computing 4 (2), 60-66.
13. Want, R. (2006) An introduction to RFID technology. IEEE Pervasive Computing Jan-Mar 2006, 25-33.

Combining Web, Mobile Phones and Public Displays in Large-Scale: Manhattan Story Mashup

Ville H. Tuulos[1], Jürgen Scheible[2], and Heli Nyholm[3]

[1] Helsinki Institute for Information Technology
`tuulos@cs.helsinki.fi`
[2] University of Art and Design Helsinki
`jscheib@uiah.fi`
[3] Nokia Research Center
`heli.nyholm@nokia.com`

Abstract. We present a large-scale pervasive game called Manhattan Story Mashup that combines the Web, camera phones, and a large public display. The game introduces a new form of interactive storytelling which lets an unlimited number of players author stories in the Web while a large number of players illustrate the stories with camera phones. This paper presents the first deployment of the game and a detailed analysis of its quantitative and qualitative results. We present details on the game implementation and game set up including practical lessons learnt about this large-scale experiment involving over 300 players in total. The analysis shows how the game succeeds in fostering players' creativity by exploiting ambiguity and how the players were engaged in a fast-paced competition which resulted in 115 stories and 3142 photos in 1.5 hours.

1 Introduction

Manhattan Story Mashup (MSM) combines the web, mobile phones and one of the world's largest public displays in Times Square to a large-scale pervasive game. The game was played by over 150 players in the web and 184 street players in Midtown Manhattan. The web players used the storytelling tool at the game web site to mash up stories, either by writing new sentences or by reusing already illustrated sentences. A noun from each new sentence was sent to a street player's camera phone for illustration. The street player had to shoot a photo which depicted the word within 90 seconds. The photo was then sent to two other street players who had to guess amongst four nouns, including the correct one, what the photo depicts. If the photo-noun pair was guessed correctly, the original sentence was illustrated with the new photo and the resulting illustrated sentence turned into an ingredient for new stories. The best stories were shown on the Reuters Sign in Times Square in real-time.

The game introduces a new form of interactive storytelling which lets distant people to collaborate in real-time. The web players get a real-time human-mediated sneak peek to the physical world which they may steer at a desired

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 37–54, 2007.

theme. The street players, who are taking the photos, may use their imagination at the fullest while trying to find the requested targets in a fast-paced competition. Since both the requested individual nouns and the returned photos are highly ambiguous in nature, the game feels somewhat mysterious, yet meaningful and exciting to all the players. On the web player's viewpoint, this may be seen as *collaborative leisure* [1] whereas the street player may see it as an urban game.

Manhattan Story Mashup brings together many concepts from earlier pervasive games. It links the physical and virtual worlds, as the games described in [2], and engages the players in a collaborative and competitive effort of storytelling [3]. It also entices players to share their experiences through a public display [4,5]. and provides an entertaining and motivating context to produce experimental data for further purposes [6].

MSM is a part of the SensorPlanet project at Nokia Research Center. In its origins, it was motivated by the need to understand issues related to mobile phone-centric sensing. Especially we wanted to get hands on vast amounts of real-world data, collected by actual mobile phones. Design, implementation and orchestration of the game provided valuable knowledge on experimentation with a pervasive application in the real world.

The game produced a rich set of data. All game events were collected in a database which allowed us to analyze the whole game event afterwards in detail. Immediately after the game, the street players were asked to fill in a questionnaire which contained questions about the game design and experience. Analysis of both the quantitative and qualitative data is presented in section 5. Some of the design decisions proved to be remarkably successful; they are listed in the conclusions. We also outline the system implementation and the set up process which present some practical lessons learnt.

We hope that the scale and the multi-faceted nature of Manhattan Story Mashup provides useful information for the design and orchestration of future experiments which utilize mobile phones, web and / or public displays. Central contribution of this paper is to strengthen some earlier findings, such as benefits of ambiguity, and to present a new field-tested game design which provides a working example of collaborative leisure and co-presence.

2 Prior Work

Positioning has been a defining feature in many earlier pervasive games. For instance, Pirates! [8] uses the physical world as a game board in which the game takes place. In this constrained game arena locations of the players may be determined using short range radio beacons. Yoshi [9] and Bill [10] exploit spotty coverage of WiFi networks in a clever way by taking advantage of this seeming limitation. In these games, the gameplay is designed around the concept of physical location. MSM was about to utilize locationing as well but only in a minor role. However we could not come up with a technically straightforward locationing method which would have fitted in the game without taking too much attention from the key concepts. Despite of this, we logged internally the GSM cell IDs of the players during the game.

MSM is first and foremost an urban photo hunt with a twist. A similar concept, involving both camera phones and public displays, was sketched by PhotoPhone Environment [11], although they did not present any implementation. MSM and Snagu[1] share the common concept of ”reverse-Flickr”: Given a keyword or tag, the player takes a photo resembling the word. Sharing the Square [1] is a system for sharing leisure between distant people through photographs, voice and location. The MSM concept is tangential to the Square’s idea of implementing co-presence through an interactive photographing process. Their treatise of shared photographing is relevant for MSM as well.

Manhattan Story Mashup builds on previous works of research which combine the physical and the virtual into a seamless game experience. Two large-scale games were produced in collaboration between the artist group Blast Theory and the Mixed Reality Laboratory at the University of Nottingham. In Can You See Me Now [12] online players are chased through a virtual model by street players who play in a real urban environment. Similarly, Uncle Roy All Around You [7] involves players both in the field and in a parallel virtual world. Street players were given a task to search for a character named Uncle Roy. Remote players, together with professional performers, guided the street players in their quest. Experiences from these games provide a solid background for designing and orchestrating collaboration in MSM.

Both the previous games are carefully orchestrated and controlled together with professional performers, which play a major role in these games. In contrast, our motivation was to gain understanding in spontaneous behavior of players, both in the Web and in the streets. Therefore we deliberately left room for emergent features and unexpected events to happen. Furthermore we designed MSM so that there would not be any inherent limitations in number of participating players. This sets MSM apart from many previous games; for example Can You See Me Now can be played only by fifteen online players at time. We hypothetize that this approach might give us realistic data on real-world sensing and user behavior, which may be generalizable to non-game related settings as well. The approach is akin to some previous games for collecting “serious” data, such as the ESP game for labelling images in the web [6].

As an example of interactive storytelling on mobile devices, Bamford et al. [3] reported an innovative mobile game in the form a multi-authored mobile book based on the 1920s surrealist technique of *Exquisite Corpse*. The book builds from a series of standard text message length contributions, each author being given only the previous message on which to base their own contribution. MSM goes a step further by including the images illustrated by some people and stories written or mashed up by other people, using recent techniques of the social web.

The usage of public displays haven been researched in various contexts. One of the main challenges associated with interactive public displays is how to entice people to interact with them [4]. Another challenge is how to share a single

---

[1] `http://snagu.com`

public display between multiple users. This is where personal mobile phones come handy in terms of dispersing access and control. For example, Scheible and Ojala [5] demonstrated with MobiLenin a solution for realizing multi-user interaction with a public display using personal mobile phones.

3 Game Design

The overall goal from the perspective of a street player in MSM is to collect points by shooting photos for illustrating stories written by people in the web. The player who collects the most points is the winner of the game. There are two tasks for the street player: One task is to shoot photos representing nouns that are extracted from story sentences created by people on the game's website in real-time. The mobile phone receives these nouns automatically and shows them as keywords in a list on the screen and the player just needs to select a keyword. The phone camera will then open automatically and start a countdown timer that grants between 60-90 seconds time to shoot the photo. The photo is then sent automatically to the game server. The other task is to match keywords to images taken by other players to validate the quality of the images going into the system. The screen will show an image and a list of four keywords including the correct one. The player needs to guess which of the keywords matches the image. The image is sent at the same time to another player for guessing in order to let the players compete. The first player to guess the word-photo pair correctly gets the points.

The player can gain different amounts of points: 1 point for shooting a photo, 6 points for guessing correctly, 9 points for a photo that is taken by the player himself and correctly guessed by some other player. The points are accumulated during the whole game. The players were able to see their own score and rank all the time on the lower part of the phone screen.

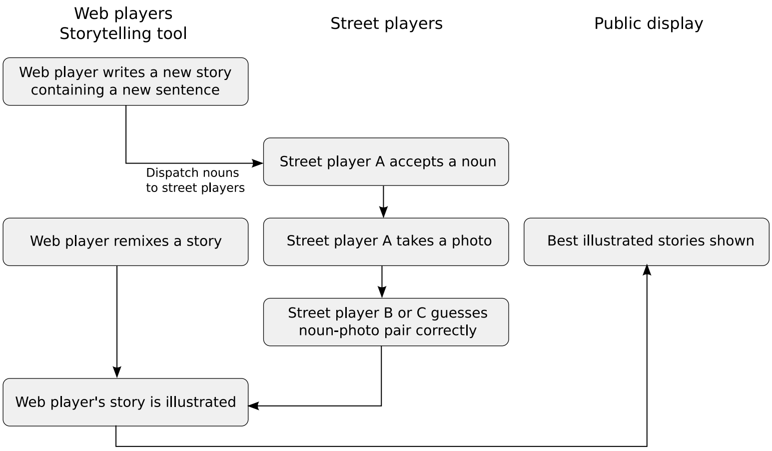


Fig. 1. Gameplay

The core of the game is to produce imaginative noun-photo pairs. The rationale for guessing is to make sure that the photos actually represent the desired target, or that the association between word and photo is conceivable by another human being, even though it may be highly ambiguous. The reason for having two players to guess the same photo is simply to increase likelihood of having a successful guess and to motivate each player to guess well – otherwise a competitor would gain more points.

The gameplay is illustrated in Fig. 1. If a storyteller in the web chooses to use only already illustrated sentences by other people, no street player actions are needed. If the storyteller contributes new sentences, a noun from each sentence must go through the illustration process. If the street players fail to accomplish any step, the story will not finish since at least one sentence will be left unillustrated.

In case that the web players are unable to keep the street players busy by writing enough stories, there is a backup mechanism which dispatches nouns to the street players from a predefined list. If such an automatically dispatched noun passes the illustration process, it may be used in a new sentence without additional delay.

3.1 Discussion

Based on their experiences with Uncle Roy All Around You, Benford et al present a design framework for mobile experiences [7]. In the following, we reflect main points of our design to this framework.

We wanted to avoid the situation in which the web players would be mere spectators, while the street players would be the *de facto* performers. Games like Uncle Roy and Can You See Me Now provide the remote player a virtual model of the physical city, thus mimicking the physical experience to some degree. In contrast to this approach, our design provides two different, yet equally important facets to the game, both respecting the natural context of action: The web player feels that she is participating in a Web 2.0-ish collaborative effort, so she may well regard herself as a major performer in the game. In parallel, the street player takes part in the hectic urban game which makes her another true performer. This two-sided approach is akin to so called *seamful design* [10] of ubiquitous systems. Instead of trying to hide differences between the virtual and physical worlds, we try to exploit the best features of both worlds.

However, both players are also spectators. Following the taxonomy presented in [13], we consider the street player to have a *suspenseful* view to the game, since she is unable to see the effects of her actions immediately: She takes a photo and hopes that someone guesses it correctly and it will get integrated to a story. In the web player's point of view, the game is *magical*, since she cannot identify the exact source of the photos: She sees newly illustrated sentences, written by other web players popping up every now and then and her own sentences becoming illustrated by some random street players.

The large public display supports the suspenseful and magical nature of the game. After some delay the street player may see her own images on the display as a part of a story. The web players could see their stories through the Reuters' webcam magically presented in a physical place in some distant location.

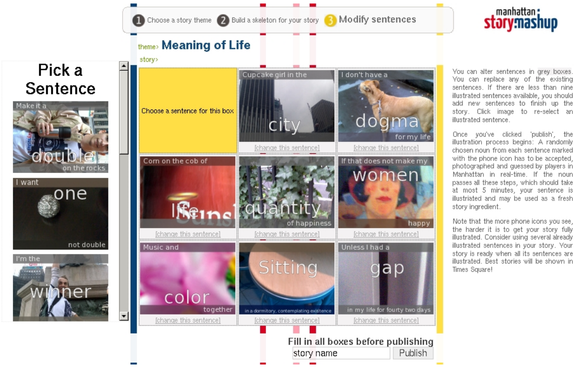
Another important feature in our design is deliberate ambiguity in the tasks. This approach is suggested for game design in [7] and [14]. Instead of forcing or encouraging the web players to use only unambiguous and concrete words, such as "house", "milk" or "sun", we picked a random noun from each sentence. This made the game more exciting and left plenty of room for players' creativity. Ambiguity is especially apparent in the guessing part of the game which requires a human player to interpret photos: Consider an image of a building. Is it a hospital or a dormitory? All these tasks boil down to being able to guess another player's intent, even though the message is mediated through an ambiguous channel, namely through a word or a photo [14]. It is even desirable that some ambiguity remains in the resulting stories since the results are often hilarious.

4 Implementation

4.1 Storytelling Tool

The Storytelling tool lets an individual web site visitor to leave her creative handprint to a large public sign, as well as interact with the real people in the streets of Manhattan. However presenting just a collection of individual contributions would be rather uninteresting. Also the required effort to make an individual contribution from the scratch might be prohibitively high for many. Instead, the storytelling tool lets the user pick a previously contributed story and use it as a basis for her own story. Alternatively one can use previous sentences to mash up or remix a personal story.

To enhance usability and approachability, the first steps with the tool were made as convenient as possible. No registration was needed: There was a link in the front page leading directly to the tool. There were no options from which to



You can alter sentences in grey boxes. You can replace any of the existing sentences. If there are less than nine illustrated sentences available, you should add new sentences to finish up the story. Click image to re-select an illustrated sentence.

Once you've clicked 'publish', the illustration process begins: A randomly chosen noun from each sentence marked with the phone icon has to be accepted, photographed and guessed by players in Manhattan in real-time. If the noun passes all these steps, which should take at most 5 minutes, your sentence is illustriated and may be used as a fresh story ingredient.

Note that the more phone icons you see, the harder it is to get your story fully illustrated. Consider using several already illustrated sentences in your story. Your story is ready when all its sentences are illustrated. Best stories will be shown in Times Square!

Fig. 2. Storytelling tool in the web

Fig. 3. Mobile Client: Keyword selection mode, shooting mode, player statistics and gallery views

choose. The user could start playing around with the tool immediately and she could contribute her own sentences with no restrictions.

A story consists of nine illustrated sentences as shown in Fig. 2. For each cell or slot the user may either pick an already illustrated sentence, written by someone else, or she may write a new sentence. Once the user has chosen appropriate content for each slot and the resulting story looks somewhat meaningful, she may publish the story.

It is practically impossible to detect dubious or nonsensical sentences automatically as the problem is highly semantic in nature. However, we performed some simple filtering using a blacklist of 333 common profanities. We also made some syntactical checks, ensuring that the sentences are long enough and do not contain any meaningless characters.

After publishing, each new sentence is tokenized. Each token is checked against WordNet [15] and nouns are collected. From each new sentence a random noun is chosen for illustration. The nouns, the number of which may vary from zero to nine, are dispatched to random players in Manhattan, yet taking care that no player has more than ten nouns at time. After this the story moves to the pool of incomplete stories, until all its sentences are illustrated. For an overall view to the gameplay, see Fig. 1.

Storytelling tool has an important feature that should keep the stories meaningful, even though users are free to write anything to sentences: When the user picks sentences for her story, she is presented a list of available illustrated sentences to choose from. This list is ordered by descending popularity: The more frequently a sentence gets picked, the higher ranking it will have in the list. This feature is similar to filtering mechanisms found in many social bookmarking sites, such as Digg[2] or Reddit[3]. The rationale is that even though someone may find entertaining to input nonsense to the system, almost no one regards nonsense written by an unknown person interesting. Thus nonsense gets disregarded by many and its ranking drops. This phenomenon is further amplified by the fact that most users consider only the top entries in the list, being too lazy to browse everything through, and thus increase the popularity of the already popular items.

---

[2] `http://digg.com`

[3] `http://reddit.com`

4.2 Mobile Client

We decided to use Nokia N80 mobile phones in the game, mainly due to their new S60 3rd edition software platform, WiFi support and high-quality 3 megapixel camera. The game client software was implemented in Python for S60 (PyS60) which is an open-source port of the Python programming language to the S60 platform [16]. Python was chosen due to its suitability to rapid prototyping and easy extendibility in native C++. As described below, source code availability was an important factor as well.

We implemented a set of custom UI widgets for the game instead of using the standard UI library. This gave us more flexibility in usability design and made the client look and feel more game-like. The standard PyS60 distribution includes support for taking photos, but does not provide the viewfinder. Since smooth camera usage is central to the game experience, we implemented viewfinder support as an extension. Thus the player did not have to use any other phone functionality or software besides our game client, which greatly enhanced game immersiveness.

The game client has three modes: Keyword selection mode, shooting mode and guessing mode. The game leads the player from one mode to another. In the keyword selection mode player chooses one of the available nouns as the next target for shooting. The keyword choice activates the shooting mode which opens the viewfinder and lets the player to find a suitable target. Once the photo is taken, it is sent automatically to the server and the player returns to the keyword selection mode.

The game client polls the server every five seconds to update the list of keywords and retrieves requests for guessing. Once a new request for guessing is received, the guessing mode is activated automatically unless the player is shooting photo. In the guessing mode the player is presented four alternative nouns, one of which is the correct one, together with a photo taken by another player. After choosing one of the alternatives, the player is taken back to the keyword selection mode.

In each of the modes, there is a visible countdown timer which forces the player to make quick decisions. In the selection mode, the timer expires keywords which have been shown for over 90 seconds. In the guessing and shooting modes, the player is taken back to the selection mode if she was unable to act in 60 seconds. In technical point of view, tight timeouts make sure that players get a constant stream of fresh tasks and passive players cannot stall the game dynamics. Rather surprisingly, since players see that the game keeps going without their explicit action, they feel motivated or even urged to act. This was a major factor in making the game highly engaging.

In addition to the three main modes, there is a screen showing current player statistics. A simple gallery is provided so that the player could see the most recent photos taken by other players. These features were added to increase feeling of competitiveness and collaborative effort. In practice however, the single line in the keyword selection mode showing the player's current score and ranking in real-time proved to be sufficient for this purpose.

An important detail in large-scale field experiments involving mobile phones is how to set up a large number of devices. Installing software to 200 phones manually is not impossible, but it is a remarkable feat. There are some standardized methods for large-scale software deployment for mobile handsets, such as Over-The-Air Programming, but often these methods are only available for operators and phone manufacturers.

Our approach was twofold: First, we were able to compile a customized version of Python for S60 since its source code is freely available. We modified the user interface of the interpreter to include functions for game client update and launch. Also the viewfinder extension was included in the new build. This reduced the number of packages which needed to be installed. Secondly, we were able to automatize the installation process to some degree by using freely available Obexftp[4] tool for Linux to transfer the installation packages to phones. We considered using several Bluetooth dongles to transfer packages to multiple phones simultaneously, but this proved to be somewhat unreliable in Linux. Instead, we used a USB-hub to connect five phones to a laptop at the same time, which greatly reduced time needed to transfer the files. After one has figured out the process, we can estimate that installing packages to some 200 phones would take 10-15 person hours. However other tasks, such as sorting out the SIM cards and recharging the phones, took two days from three persons.

4.3 Large Public Display

We rented the Reuters Sign[5] in Times Square for exclusive use during the game. The display system consists of 11 individual displays which may be used either separately or as a single large display. This display was chosen due to its prominent location and enormous visibility. Times Square is an iconic location in global scale, thus the possibility to create personal content to be shown there was attractive for people around the world. Considering the storytelling tool, it was crucial that the web users could relate to the place where the stories were to be shown, even though the location was distant.

Once all the nine sentences of a story had successfully gone through the illustration process, depicted in Fig. 1, the story became a candidate for presentation on the display. We had a human moderator filtering the candidate stories using a separate moderation interface in the web. The interface allowed the moderator to “bless” a story for presentation, or to blacklist individual sentences containing unwanted content.

Once a story was blessed, it was automatically sent to the display system by the game server. Depending on the display status, the new story was shown there after some 1-5 minute delay. All the nine illustrated sentences were shown at once on the display. One by one, each sentence was enlarged and shown on the large middle display for six seconds, as seen in Fig. 4.

The display added a unique twist and a big wow-effect to the game. It provided a feedback channel for the street players who were able to follow in real-time

---

[4] http://openobex.triq.net

[5] http://www.timessquare2.com

Fig. 4. The Reuters Sign in Times Square. Photos by Kitty Kat, Jürgen Scheible and _snipp.

how their photos were intervowen into various stories. This role, providing a shared view to the game, was probably the most important feature of the display. In addition, the unique opportunity of getting a personal fingerprint to Times Square motivated both the web and the street players to produce imaginative content.

5 Experience

The actual game event took place on September 23rd 2006 between noon and 1:30pm in Midtown Manhattan. The game was one of the featured games in Come Out and Play street games festival [6]. We had invited 140 university students to participate from New York University, Parsons The New School for Design and Brooklyn Polytechnic. In addition we had invited some 70 persons from various companies and institutions to join the game.

In total 184 players played the game. During the game the players took 3142 photos and made 4529 guesses, 2194 (48.4%) of which were correct. Technically, there should have been 6284 guesses in total, assuming that every photo was guessed by two separate players. However, if all the players were busy, i.e. taking a photo or already guessing, only one guesser would suffice. In the extreme case that no guessers were available, the image was accepted as such. Also if the guessing timeouted or the player closed the client without guessing, no guess was recorded.

Figure 5 visualizes the whole gameplay. Each row in the graph corresponds to a player. Rows are ordered by descending score, thus the topmost row corresponds to the winner of the game. X-axis is time, from the game start to the end, 4902 seconds or 81 minutes in total. Color segments indicate player action at every moment, which correspond to the three modes of the game client: White indicates keyword selection mode, green/light shooting and blue/dark guessing.

The most distinctive feature in the graph is the large gap in the middle. This corresponds to an unknown bug which appeared after the game had been running for some 30 minutes. Game server stopped dispatching nouns, apparently

---

[6] `http://comeoutandplay.org`

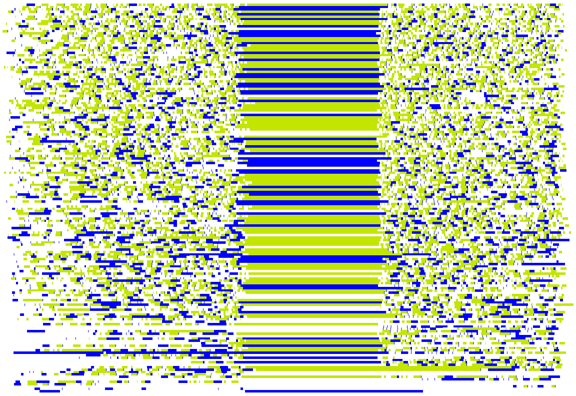


Fig. 5. Gameplay visualized: Rows correspond to players, x-axis is time. Green/light segments denote time spent taking photo and blue/dark segments denote guessing. Rows are ordered by descending score.

assuming that all players already have the maximum amount of pending keywords. It took a while to notice that the bug affected all the players. After this, the game server was restarted and the game history was re-evaluated, which normalized the situation and the game continued.

The following phenomena can be interpreted from the graph:

1. Gameplay was fair: All the players seem to have done comparable amount of shooting (green/light) and guessing (blue/dark).
2. Nouns were not too hard: There are only few long white gaps in player actions, thus players accepted proposed nouns quickly.
3. Players were motivated: Even the bottommost players played hard all the time.
4. Players did not get bored: Right-hand side of the graph looks the same as the left.
5. Scoring is not random: Segments are slightly longer in the bottom part of the graph, indicating that low scores correlate with slow actions.

Observation 1. proves that the game dynamics worked correctly: It was a deliberate decision that a player could not avoid guessing by staying in the shooting mode all the time. Observation 2. proves a hypothesis correct: Players were remarkably creative with the words and they were able to illustrate even the most abstract words, such as “synergy”, by utilizing the rich urban environment and spontaneous acting. Observation 3. was unexpected: We assumed that the players who notice their ranking to be rather low, say 83th, would soon lose interest in playing. However it seems that the players often got spurts of easy targets, leading to sudden increase in ranking which kept motivation high. This leads to the conclusion that showing the current ranking in the game UI is a good idea. Similarly observation 4. was unexpected: We decided to make the game short enough, so that people would not get bored during the game. Clearly the game was engaging enough to keep players fully focused for 90 minutes. Observation 5. was expected, but we assumed that the effect would be much stronger. The data shows that the players were surprisingly homogeneous, regardless of their final score.

We analyzed the game history also quantitatively, to confirm the observations from the graph and to find hidden patterns. Basic statistics over players are presented in Table 1. Player characteristics were distributed rather uniformly and no clusters of different types of players were found, which agrees with observation 3. Overall, it seems that that the players were a homogeneous group of well-motivated individuals which resulted to a uniform outcome. This provides us a solid baseline of the game design against which different, more heterogeneous groups of players may be evaluated in the future.

We were interested in finding out whether the winning players had some special characteristics compared to the others. We found no clear pattern related to number of correct guesses or guessing in general. The most remarkable correlation exists, not surprisingly, between the average shooting time and the final score. To conclude, it seems that the winners won by making good decisions quickly and by acting fast. This might be considered as a positive outcome for a new game design.

Table 1. Gameplay statistics over players

Variable	Min	Max	Median
Nouns chosen	2	64	35
Photos shot	2	58	32
Guesses	3	56	28
Correct guesses	1	29	14
Score	18	430	246
Guessing time (avr secs.)	6.5	39.3	15.0
Shooting time (avr secs.)	15.9	91.9	33.4
GSM cells visited	1	26	14

We also recorded player locations based on GSM Cell IDs. Even though the game took place in a restricted area, namely between the 59th and 43rd streets, around the 5th Avenue and Broadway, in total 197 unique cell IDs were recorded. There seems to exist a weak positive correlation between the number of cells visited and the final score.

5.1 Photos

The most intriguing tangible result of the game were the 3142 noun-photo pairs. The game was designed to stimulate creativity. Our research motivation for this was to get some preliminary ideas on what kind of “action possibilities” urban environments might provide and what kind of non-trivial features people are able to find in their surroundings.

We analyzed 523 photos manually, to gain understanding in the types of associations present in the photos. The largest single theme was 193 photos showing human beings, or players in our case. Sometimes acting was used to set out a role, facial expression or a character, as in the leftmost photo in Fig. 6. However, the photos did not always depict a human being as such, but another player acting out an abstract concept such as “evidence”, “link” or “equal” as seen in the middle photo. Sometimes the association was related to the textual form of the word, resembling a wordplay, as in the rightmost photo.

We were surprised to see how effortlessly the players were able to cross the boundaries of ordinary public behavior, e.g. by acting out publicly, once they started to look at the world through the game client, especially when playing in a team. This suggests that concerns related to blurring the line between the virtual and physical, which are discussed e.g. in [7], are valid even though the game by itself may be quite harmless and abstract, as in our case.

Fig. 6. Example photos and sentences

Another imaginative approach to deal with abstract concepts was to deliberately blur or shake the photo to hide irrelevant details. Of the analysed photos, 155 were shaky. Shaky photos were also used to depict colors or movement, as in “explosion” and “speed” or kicking something to show “temper”. In case of unambiguous but unavailable objects, such as “mustard”, “film” or “balloon”, players tried to blur another object with some resemblance to the target object to produce the desired effect. Naturally some photos were shaken or out of focus by accident. This happened often with close ups (“powder”, “ear”, “bruise”) if the player forgot to switch the macro mode on. Since the players were forbidden to use any visible trademarks or copyrighted items in the photos, some players tried to blur the offending target to circumvent the rule.

Since the nouns were recognized using WordNet, which also includes semantic relations between words, we were able to group nouns under some generic categories or themes automatically. To reduce the number of categories shown, Table 2 presents only those categories or hypernyms which were used to extract words for automatic dispatching before the game. The numbers in parentheses indicate the total amount of nouns illustrated in the corresponding category. Accuracy indicates the percentage of correctly guessed noun-photo pairs in the category.

It is worth noting that the lowest accuracy is well above the default, 0.44, which is the probability that at least one of the two guessers makes the correct choice assuming that the choices are random. Since the actual accuracies are above this, we can assume that the players were paying attention to the guessing part and probably tried to perform as well as possible.

Table 2 shows that explicit human-related subjects, such as facial expressions or characters, are easy to act out and photograph regardless of the surroundings. Likewise unambiguous concrete objects, such as bodies of water, shops, and beverages are easy to guess. In contrast, objects that are difficult to set out, such as "dormitory" or "breakfast", produce often ambiguous photos. However, the players seemed to enjoy ambiguity and in some cases they deliberately took ambiguous photos for the sake of fun and ingenuity. Thus although ambiguous words and photos made the game more difficult, they were likely to provoke engagement and exploration in the game [13].

Table 2. Guessing accuracies per WordNet categories. Total number of guesses in parentheses.

Hypernym	Accuracy	Hypernym	Accuracy	Hypernym	Accuracy
facial expression	0.94 (17)	car	0.74 (34)	currency	0.67 (3)
character	0.86 (7)	show	0.72 (25)	friend	0.66 (35)
writing implement	0.85 (13)	bread	0.71 (14)	edible fruit	0.64 (28)
body of water	0.84 (31)	wheeled vehicle	0.69 (36)	light	0.64 (67)
athlete	0.83 (24)	waste	0.69 (29)	road	0.62 (29)
shop	0.81 (26)	motor vehicle	0.68 (19)	toy	0.61 (18)
performer	0.77 (22)	tool	0.68 (50)	material	0.61 (79)
chromatic color	0.77 (61)	human	0.68 (499)	piece of work	0.59 (22)
beverage	0.75 (24)	garment	0.68 (87)	dish	0.56 (48)
figure	0.74 (72)	device	0.67 (181)	building	0.56 (70)

5.2 Web

Approximately 4000 unique IP addresses had visited the game web site at `storymashup.org` before the game launch. The storytelling tool was open only during the game for 90 minutes. During this time 165 unique IP addresses visited the game web site. 115 stories were published, which were mashed up using the 271 sentences written during the game.

Since the players in Manhattan accepted 5603 nouns in total, only 4% of the words originated from the new sentences. However, the storytelling tool offered a possibility to use already illustrated nouns in a sentence. This feature was used 129 times. We hand-picked 26 best stories which were shown in Times Square. The game design, as well as the implementation, would have been able to handle much larger number of storywriters in the Web.

A major technical research motivation for having the web site in the first place was to gain better understanding in different time-scales of the web and the physical world, and how it affects the system dynamics. In practical terms, minutes make a big difference while one is standing on a busy street in Manhattan, compared to web site which is still mostly conceived as a rather static entity. Especially it is not accustomary to have a web site open only for 90 minutes, even though such a happening makes sense in the physical world. Thus to achieve smooth real-time interaction between the web and the physical world, the system must carefully take into account inherent differences between the two time-scales.

Correspondingly, the difference in the user base between the web and mobile devices may be huge, yet it is not easily predictable – another lesson learned. Being able to accommodate highly volatile user bases is a challenge to the system dynamics. In our case, for example, the system should not have discarded stories so eagerly (see Fig. 1), having so few storytellers. However, the design was based on the assumption that there are far more web users than players in the field, and thus the stories should have been in abundance. In the future we are determined to increase the system flexibility to adapt to situations like this.

The web site included a questionnaire which was answered by eight web players. Apparently the players who answered the questionnaire had generally enjoyed the experience, so it is difficult to make any conclusive remarks. To exemplify nature of the results, we include here answers for three of the questions:

Any surprise elements? “yes, when it became possible to be funny and make connections ... very surreal”, “Yes, when I saw the image that was for the words I selected.”, “discovering own sentences in other stories was very surprising.”, “It was a surprise that ”the game” didn’t seem to work at all. Also the term ”game” is misleading. It is more comparable to a mobileentertainment solution. Not into a game, as you don’t play it anyway.” **What did you like most?** “crafting the story. and wondering what photos would come out. using my knowledge of the city to make puns”, “Seeing everyone else’s creative efforts.”, “How people used my story und how the pics where done to my keywords.”. **Do you want to play this game again?** “Now that I have the hang of it, I’m thinking of new ways to play it...”, “yes! all the time”, “Yes. I would like to master whta I dont understand”, “Yes, please!!!!! It could lead to addiction, cause it is so much fun to write stories to the topics you offered and to wait how they will develope.”, “No, as it is not a game. If it’d work, I would reconsider writing in there.”.

5.3 Questionnaire

Qualitative data reflecting the user experience was evaluated with a questionnaire, which was collected immediately after the experiment. The questionnaire contained 26 statements on which the users were asked to answer on a 5-point scale between 1 (disagree completely) and 5 (agree completely). In addition there were 23 open-ended questions. A few individual users and 2-3 groups of users were also video interviewed after the experiment. Observation was carried out during the experiment with a video camera.

Of 184 players who participated the game, 99 returned a filled in questionnaire (56 males and 43 females). The age distribution was as follows: 24 players of age

18-24, 64 of age 25-34, and 15 of age 35+. Each player got an invitation to the Story Mashup evening party upon returning the questionnaire. In the following, we adopt the following notation for brevity: “Statement” (X), where X denotes the average of responses of 99 players on the 1-5 point scale.

As a general observation it was very clear that people really enjoyed playing the game: “It was fun and engaging to play this game” (4.07). This underlines that creating an interactive, engaging experience is key for a successful user participation for any large scale research experiment.

Of all the questions, it was a surprise that exactly the two major tasks that street players had to perform, got the highest rating: “I enjoyed hunting for photos” (4.63) and “I enjoyed the guessing part of the game” (4.39). This can be seen as a strong success from the game design point of view. It shows that players made an intellectual and emotional investment which Ryan [17] claims to be a precondition of an interactive medium to open its world to the user. Clearly, the use of mobile device as an interaction device and as an image capturing device in the context of our game is strong: “When I was holding the phone, I felt confident hunting for images and doing the guessing part” (4.18).

It seems that hunting for points and competing with other players in game context indeed provides a motivated activity. We can see from the data that the game mechanisms played a major role in getting players engaged and motivated. The ratings are very hig: “I like the competitive style of play” (4,32) and “I was motivated by the scoring mechanism” (4,25).

People were very active and also various interactions between people took place, which can be easily seen in the photos as well (see Section 5.1). People enjoyed socializing and teamplay: “I did prefer to play this game alone rather than joining a team of other players” (1.43). Players also felt, with some variation, that they are part of a joint authorship: In this case it was by contributing images to the web stories and ultimately to the large public display: “While playing I felt I was part of a joint activity between players on the web and street players in Manhattan” (3.22), “I felt I belonged to a joint, collaborative action contributing to a common goal” (3.59) and “For me it was an important part of this game to see the illustrated stories at the public display at Times Square (3.28)”.

Also the usability goals were met, meaning that the system should be efficient to use, easy to learn and easy to remember: “The mobile application was easy to use” (3.19), “It was easy to understand the game concept” (4.21) and “At any given moment it was clear to me what I was supposed to do” (3.51).

6 Conclusions

The game was a greater success than we expected. The game design proved to be engaging which supports our core design decisions. We were delighted by the amount of imaginative photos taken during the game. On the other hand the game was almost too motivating, since the data turned out to be quite homogenous, although we expected that some of the players would play rather lazily.

We think that the following factors contributed the most to the success of MSM: **Lazy Shooting**: A user with a camera is a powerful way to collect interesting data from the physical world. In addition, in our case it was effortless: The camera activated and sent the image automatically. This encouraged players to shoot many photos. **Ambiguity**: Players felt that clever and imaginative thinking contributed to their ranking. Players spent most of their time doing something that is natural to human beings: Trying to infer other people's intentions or trying to infer how other people would infer my intentions. **Teamplay**: Players formed teams spontaneously and used them to act out difficult words. Teamplay also lowered barrier for crossing the boundaries of public behavior. **Speed**: The game felt immersive since the players did not have time to think about anything else. Since the players' ranking was updated all the time, even a short pause resulted in a noticeable drop in score. **Integrated Game-Flow**: A single simple game client was used to play the game. Even the camera was included in the client. This ensured that we were able to streamline the UI to the bare minimum. With rapid prototyping tools like Python for S60, one does not have to accept suboptimal interfaces.

On the other hand, we could not find satisfying answers to the following issues: **Time and Web**: Even though technically our web site worked as expected, we would have expected more web players participating in the game. We assume that a major reason for the smallish number of visitors was that the game interface was open only during the game, for some 1.5 hours. In the future, we have to pay more attention to matching the physical and the virtual time-scales. **Scalability**: Organizing a game involving almost 200 physical players was a major feat. The semi-manual approach for installing the phones does not scale easily to even larger settings. **Duration**: The game lasted for 1.5 hours which is enough for an intensive game but in order to get more data, we would like to keep the system running for a longer period of time. **Adaptability**: If the game involves web or it is otherwise open for anyone to participate, the rules should adapt to varying number of players and activity. This is especially important if the game runs for a longer period of time. Many of the previous issues relate to the natural friction between the web and the physical world. It is clear that further experiments are needed to find best practices and design frameworks for successful interaction between the two worlds.

References

1. Brown, B., Chalmers, M., Bell, M., MacColl, I., Hall, M., Rudman, P.: Sharing the square: collaborative leisure in the city streets. In: Proc. of the Ninth European conf. on Computer-Supported Cooperative Work, Springer (2005) 427–447
2. Benford, S., Magerkurth, C., Ljungstrand, P.: Bridging the physical and digital in pervasive gaming. Communications of the ACM **48**(3) (2005)
3. Bamford, W., Coulton, P., Edwards, R.: A massively multi-authored mobile surrealist book. In: Proc. of the SIGCHI conf. on Advances in computer entertainment technology, ACM Press (2006)
4. Brignull, H., Rogers, Y.: Enticing people to interact with large public displays in public spaces. In: Proc. of INTERACT-03. (2003) 17–24

5. Scheible, J., Ojala, T.: Mobilenin combining a multi-track music video, personal mobile phones and a public display into multi-user interactive entertainment. In: Proc. of the ACM conf. on Multimedia, ACM Press (2005) 199–208
6. von Ahn, L., Dabbish, L.: Labeling images with a computer game. In: Proc. of the SIGCHI conf. on Human factors in computing systems, ACM Press (2004) 319–326
7. Benford, S., Crabtree, A., Reeves, S., Sheridan, J., Dix, A., Flintham, M., Drozd, A.: Designing for the opportunities and risks of staging digital experiences in public settings. In: Proc. of the SIGCHI conf. on Human Factors in computing systems, ACM Press (2006) 427–436
8. Falk, J., Ljungstrand, P., Bjork, S., Hannson, R.: Pirates: Proximity-triggered interaction in a multi-player game. In: Extended Abstracts of Computer-Human Interaction (CHI). (2001) 119–120
9. Bell, M., Chalmers, M., Barkhuus, L., Hall, M., Sherwood, S., Tennent, P., Brown, B., Rowland, D., Benford, S.: Interweaving mobile games with everyday life. In: Proc. of the SIGCHI conf. on Human Factors in computing systems, ACM Press (2006) 417–426
10. Chalmers, M., Bell, M., Brown, B., Hall, M., Sherwood, S., Tennent, P.: Gaming on the edge: using seams in ubicomp games. In: Proc. of the SIGCHI conf. on Advances in computer entertainment technology, ACM Press (2005) 306–309
11. Thoresson, J.: Photophone entertainment. In: CHI'03: Extended abstracts on Human factors in computing systems, ACM Press (2003) 896–897
12. Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M., Row-Farr, J.: Can you see me now? ACM Trans. Comput.-Hum. Interact. **13**(1) (2006) 100–133
13. Reeves, S., Benford, S., O'Malley, C., Fraser, M.: Designing the spectator experience. In: Proc. of the SIGCHI conf. on Human factors in computing systems, ACM Press (2005) 741–750
14. Aoki, P.M., Woodruff, A.: Making space for stories: ambiguity in the design of personal communication systems. In: Proc. of the SIGCHI conf. on Human factors in computing systems, ACM Press (2005) 181–190
15. Fellbaum, C., ed. In: WordNet: An Electronic Lexical Database. MIT Press (1998)
16. Laurila, J., Tuulos, V., MacLaverty, R.: Scripting environment for pervasive application exploration on mobile phones. In: Adjunct Proc. of Pervasive 2006. (2006)
17. Ryan, M.: Narrative as Virtual Reality. Johns Hopkins University Press (2001)

PersonisAD: Distributed, Active, Scrutable Model Framework for Context-Aware Services

Mark Assad, David J. Carmichael, Judy Kay, and Bob Kummerfeld*

The University of Sydney, Sydney, Australia
{massad,dcarmich,judy,bob}@it.usyd.edu.au

Abstract. PersonisAD, is a framework for building context-aware, ubiquitous applications: its defining foundation is a consistent mechanism for scrutable modelling of people, sensors, devices and places. This paper describes the PersonisAD features for supporting distributed models with active elements which can trigger when relevant events occur. This framework makes it possible to quickly create new context-aware applications. We demonstrate the power of the framework by describing how it has been used to create two context aware applications: MusicMix which plays music based on the preferences of the people in the room; MyPlace, which informs people of relevant details of the current environment. Major contributions of this work are: the PersonisAD framework which provides a powerful and consistent means to respond to significant changes in the models of people, sensors, devices and places; support for distributed models and associated resource discovery; two applications that illustrate the power of PersonisAD.

1 Introduction

An essential element of the ubiquitous computing vision is that relevant contextual information is used to drive the behaviour of systems. This should also take account of the fact that different people have different needs and a single person has different needs over time, based on a vast array of their own attributes such as their changing knowledge and current goals. People's needs are also driven by the many and varying elements of their current environment, such as the people present or nearby and what is happening. So, for example, if Alice enters an unfamiliar building to meet Bob, she needs to know about the relevant people and services in that place. Bob is a relevant person since Alice's goal is to meet Bob. Alice's friend Carol is a relevant person if Carol is nearby. A context-aware system might model the fact that Alice is unfamiliar with this building, that she is meeting Bob and that she knows Carol. It may use this and other information from sensors to tell Alice how to make her way to Bob and it may point out Carol is nearby.

There is considerable ubiquitous computing research directed at determining a person's location: indeed, it is arguably the most explored aspect of context-aware applications. There have also been many applications that use location

---

* This research was part funded by the Smart Internet Technology CRC Australia.

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 55–72, 2007.

information to customise their actions in some way, for example [1–3]. By contrast, there has been far less work that also takes account of factors such as user attributes and preferences.

To move towards the broad vision of timely and useful information availability and management, we need to maintain *models* of the significant elements of the ubiquitous computing environment. First, and foremost, we need to *model people.* A starting point is to model a person's location. However, a rich ubiquitous computing future also requires models that capture other relevant information about people, such as their knowledge, preferences and goals. A ubiquitous computing application needs to make use of these models of people. It also needs to make use of models of *places, sensors, services and devices* in the environment. All these models can be implicit, perhaps within the application. However, there are real advantages in making them explicit, with existence outside any one application: we can then use consistent mechanisms to reason about any of these entities and to transmit information between them.

Since the model of a person is inherently personal information, an important requirement on the modelling is to ensure that people can maintain control of both what is modelled about them and how it is used. We describe models as *scrutable*, if they are designed so that a person who chooses to investigate them can determine just what is modelled. This is a foundation for ensuring the people will be able to *control* their personal information and its use in a ubiquitous computing environment. We envisage that people like Alice in our scenario should be able to ask questions like: How did the system determine my location? Who else can see this information about me? What does the system believe that I know about this building? What evidence sources can contribute to the model of my location?

PersonisAD is framework which takes account of these issues. It grows from our earlier work on scrutable modelling of people [4] and exploration of interfaces that can support scrutability of user models [5] even when they are large [6]. We recognised the close relationship between the people in a ubiquitous environment and the devices, sensors and places [4,7]. In this paper, we describe two important advances in this work: we have enhanced Personis [4,7] to support modelling that is both *active* and *distributed*.

In Section 2, we describe the PersonisAD architectural framework, explaining the notions of *active* and *distributed* models. Section 3 reports our evaluation of its power and flexibility in terms of applications built on top of it and analysis of their implementation. We then describe related work, and conclude by reviewing the contributions as well as future work.

2 PersonisAD Overview

We describe the broad architecture of an application based upon the PersonisAD framework in terms of Figures 1 and 2. Then we work through a detailed example of an application called MusicMix to illustrate how the framework serves as a foundation for building context-aware applications.

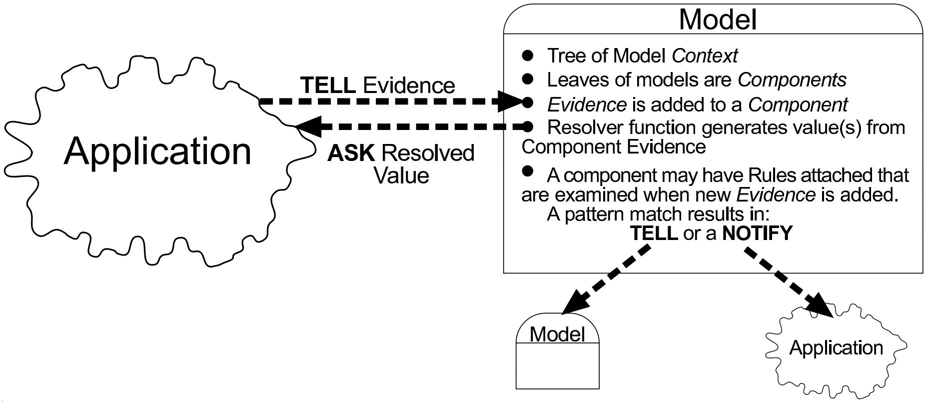


Fig. 1. Interaction between an application and a model

First consider the model representation and reasoning, as outlined at the right of Figure 1. The first core notion of the model is that it is organised as a tree of *model contexts* which contain the *component*s of the model. This hierarchical structure serves to define namespaces and to collect the information and associated reasoning elements.

The figure shows the general mechanism of PersonisAD is to collect *evidence* for the value of the context attributes (*component*s) and when a value for the *component* is requested we examine the available *evidence* using a *resolver* function to generate a value.

We call this mechanism *accretion/resolution* and it has many advantages. The value for the *component* is only calculated when it is required rather than continuously as in many systems. For an attribute such as location, where many data sources (eg GSM, bluetooth, WiFi, GPS) may be consulted to calculate a location value, it is a waste of resources to calculate a location whenever a new data point is received. Only when the location value is required should the value be calculated.

Another major advantage is that different situations may require different answers for the same attribute. For example, the answer to the question "Where is Bob?" may need to be a latitude/longitude for one application but another application may need a detailed symbolic value such as room/building/address while another might just be given the resolved value, "at work". This can be catered for by using an appropriate resolver function.

A key problem of context-aware computing is to provide the context information to an application at the time it becomes relevant. As shown in Figure 1, our system provides for the timely notification of applications when an appropriate context is present. It does this by associating triggers or *rules* with components. We call components that have rules attached, *active components* and hence *active models.* A rule is evaluated when new *evidence* is received for the *component*. If the rule succeeds, it can result in *evidence* being added to any *component* or an application being notified. Chains of these rules can be used to generate a sophisticated response to changes in context. The rules take the form:

`value ∼ pattern : action`

where `value` is a resolved value of a *component*, `pattern` is a regular expression. As illustrated at the lower right of Figure 1, the trigger language has two possible `actions`: a `tell` which adds *evidence* to a *component* in any model; or a `notify` which can send information to an application.

A major part of our approach is the desire to allow users access to the models of them and other entities in order to explain the actions of the system. We use the term *scrutable* to indicate that the model is understandable and that the system can generate explanations. We have designed the whole PersonisAD system with scrutability as a key concern and one of the design goals is to decouple the elements and to achieve simplicity since both of these are critical for providing explanations that the user can scrutinise to understand and control their personalisation in applications.

The implementation of PersonisAD [4][1], provides a small number of simple operations for maintaining the models. The main operations are:

access: Each entity and its associated model has a globally unique ID. The *access* operation is used to locate the model server and connect to it.
tell: A piece of evidence (value, source, type) is added to a given model component within a given model context.
ask: A value for a model context/component is returned after resolution by a nominated resolver function.

As shown in the left part of Figure 1, the normal communication between applications and the models is via the `ask` to request the resolved value of a *component* in the model and the `tell` to send new *evidence* to a *component*.

The next important part of the PersonisAD framework is the support for distributed models as shown in Figure 2. All interaction with models is based on the same basic operations just described in Figure 1. To support distributed models, these operations (and others for model management) are implemented using a simple remote procedure call mechanism based on JavaScript Object Notation (JSON) with HTTP as the transport protocol. This is shown as the broken lines from the application to the models in the figure.

The solid line in the figure indicates how PersonisAD locates the server for a model using the multicast and wide-area Bonjour [8] (Zeroconf) service discovery system is used. This allows efficient discovery of model servers both on the local area network and across the Internet. When models are created they are registered locally with the multicast DNS server, and remotely on the `personis.org` domain name server. Addresses of servers are not directly tied to the name of the model, allowing models to be moved from one physical computer to another without the need to update any dependent models. Even the simple example

---

[1] Personis Pty Ltd.

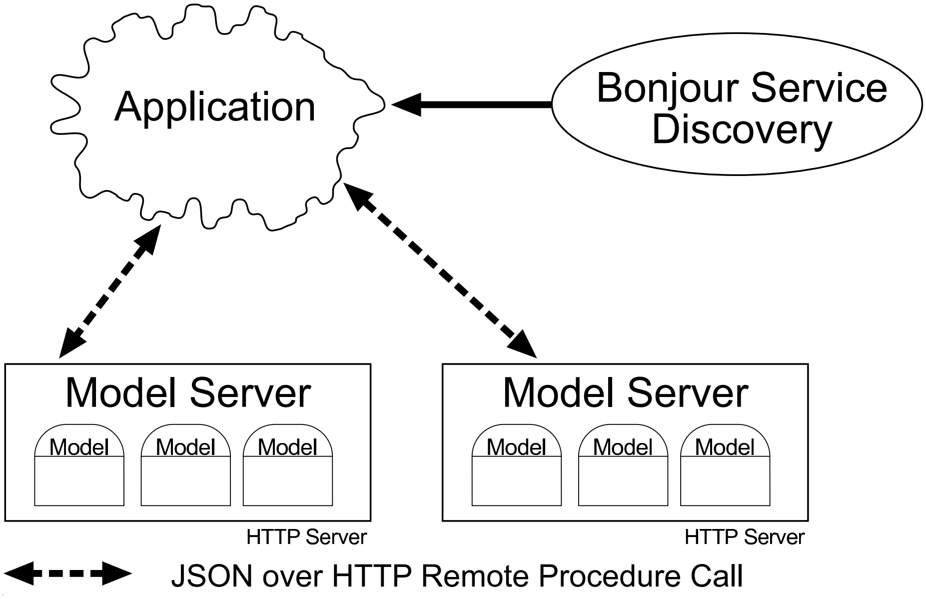


Fig. 2. Interaction between an application and model servers

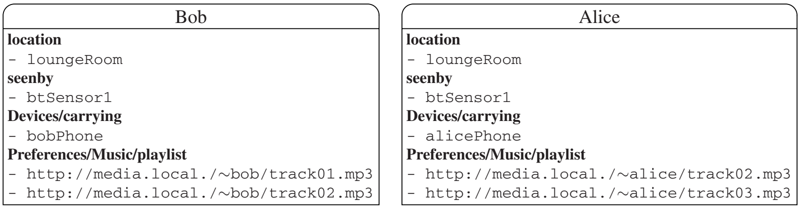
systems described in this paper involve model servers on several machines. We routinely move models between machines for convenience with no change required to the applications.

2.1 The MusicMix Application

We now illustrate our PersonisAD approach using one application built upon it: MusicMix is a demonstrator which plays background music, selecting the playlist according to the music preferences of people who are detected in the listening area. So, for example, if Bob is alone in his living room, the MusicMix in that room selects music he likes. When a friend, Alice, drops in, MusicMix plays music from the combined set of music that they like. We use this example to illustrate how we model people's location and make use of models of their music preferences to drive MusicMix.

MusicMix needs models of the people, the space and the devices involved in playing the music and determining the current location of the people. The music player needs to be notified when the context changes, that is when a person enters or leaves the area or an existing person changes their music preferences.

Figure 3 gives a simple view of parts of user models for Bob and Alice at the time when they are both in the lounge room. The first *component* is **location**. In the figure its *resolved value* is `loungeRoom` for both users.



Bob

location
- loungeRoom

seenby
- btSensor1

Devices/carrying
- bobPhone

Preferences/Music/playlist
- http://media.local./~bob/track01.mp3
- http://media.local./~bob/track02.mp3

Alice

location
- loungeRoom

seenby
- btSensor1

Devices/carrying
- alicePhone

Preferences/Music/playlist
- http://media.local./~alice/track02.mp3
- http://media.local./~alice/track03.mp3

(a) Bob's user model (b) Alice's user model

Fig. 3. Examples of simplified user models for Bob and Alice

The next *component* in Figure 3 is **seenby**. Its resolved value is the identifier for the sensor that last detected the user. We will return to this when we explain how a user's location is determined in PersonisAD. In this case, this is `btsensor1`, a Bluetooth sensor for both users.

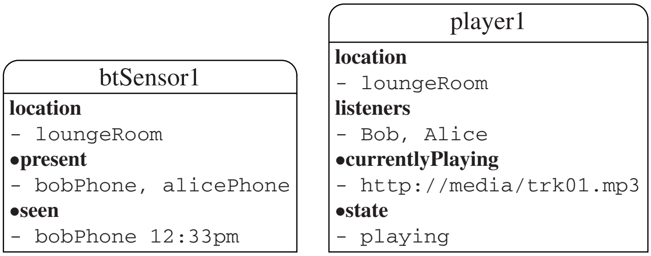
Following this, the figure shows the resolved value for the **Devices** context's *component*, called **carrying** which models the devices that the user has. This has *evidence* for each device the user is carrying. Its resolved value is a list of the identifiers for the models associated with such devices. In Figure 3, both Bob and Alice are carrying their respective Bluetooth phones. To describe PersonisAD, we take the simple case where the only means to locate these users is when their Bluetooth phones are detected. Our deployed system has other location sensors.

The remaining *component* in Figure 3 is for the *model context* for **Preferences** and within it, the subcontext for **Music** preferences and within that the **playlist** *component* whose resolved value is the list of the person's preferred music. In the figure, this has just two tracks, where one is common to both users.

2.2 Modelling Sensors, Devices, Places

The same accretion/resolution representation is used to model sensors, devices and places. Figure 4 shows examples of each of these. The Bluetooth sensor model in Figure 4(a) has a *component* called **location** which gives its location, in this case the **loungeRoom**. It also models the devices it has **seen**; each time it detects a device, a piece of *evidence* is added to this *component*. The intuitive interpretation of this evidence is that recent evidence indicates which devices were in the place where that sensor is located.

The device model in Figure 4(b) is for the music player. It has a *component* for *location* and the current value for this is `loungeRoom`. The second *component* for the player device is **listeners** and its value is the list of people modelled as listening to it. In the figure, this is `Bob` and `Alice`. The **currentlyPlaying** *component* has the url of the music that is currently being played and, finally, the *component* called **state** indicates that it is playing.



(a) Bluetooth sensor (b) Audio player

Fig. 4. Examples of models for sensors and devices and places. The bullets indicate *active components*.

2.3 Determining Location

We illustrate the PersonisAD approach in determining Bob's location and also in determining who is present in a location for the MusicMix system. We gather evidence from many sensors to infer location but in this example we show our use of Bluetooth capable mobile phones. Figure 5 shows the Bluetooth Sensor, and models for the sensor, user, user's phone and the lounge room.

There are three main stages in the process illustrated in the diagram. Step 1 occurs when Bob's phone is detected by a Bluetooth sensor. This causes the addition of a piece of *evidence* to the models for Bob's phone and for the sensor. This triggers Step 2, which causes the addition of *evidence* indicating that the phone is in the lounge room and that there has been a change in the devices present at this sensor. Step 3 adds *evidence* to the lounge room model indicating that Bob is present and to Bob's model indicating he is in the lounge room. We now go through this in more detail to illustrate how the PersonisAD approach operates.

Bob's phone, with model `bobPhone`, is detected by the sensor with model `btSensor1`. The sensor program places *evidence* into the **seen** *component* of the sensor model (`btSensor1` in Fig. 5) and the **seenby** *component* of the phone model (`bobPhone` in Fig. 5) using two **tell** operations. This is Step 1 in the figure.

Both of these components have a rule attached to them, indicated by bullets next to them.

When a new piece of *evidence* is delivered to the **seenby** *component* the attached rule is interpreted. The resolved value of the **seenby** *component* is matched with a regular expression that always returns True (since in this case we want to update the location when a new piece of *evidence* arrives). The action in the rule adds a piece of *evidence* containing the location of the sensor to the **location** *component* of the device that had been sensed, in this case the phone of the user.

The action of this rule is labelled 2b in the figure. The solid line represents the **tell** occurring, and the dotted line shows where another model has been accessed during evaluation of the rule. For example, in this case, the value of the sensor's `location` *component* is needed to create the evidence indicating that the phone is in the location with model `loungeRoom`. Expressions within a rule can be nested: in this example one **ask** is evaluated to form the model id for another **ask**.

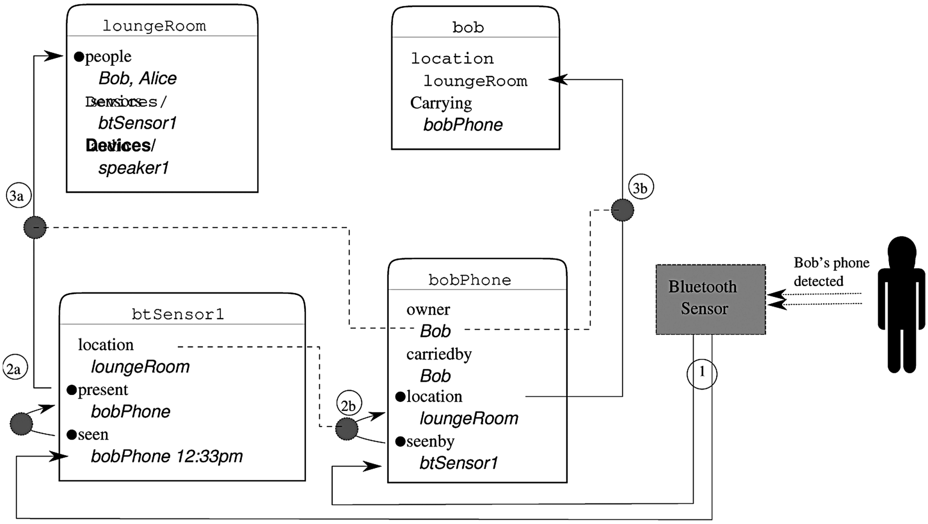


Fig. 5. Updating a user's location and the people present in a place: Bluetooth sensor and models for bob, bobPhone, btSensor1 and the room. Solid lines represent **tell**s, with the dashed lines showing the additional models accessed while evaluating the rules.

Similarly, there is a rule attached to the **location** *component* of the phone and this has the effect of **tell**ing the location of the phone to the model of the person carrying the phone. This is labelled 3b in the figure.

There is also a rule on the **seen** *component* in `btSensor1` model. Labelled 2a, this causes the recent **seen** items to be examined to create a list of **present** devices. It only triggers when the resolved value of this *component* changes. So, although evidence is continuously added to this sensor *component*, the rule only adds evidence to the `present` *component* when a new device is detected.

This in turn causes the rule attached to the **present** *component* of the model `btSensor1` to be evaluated. This step is 3a in the figure, and it **tell**s the `people` *component* of the `loungeRoom` model who is now present based on the devices detected. A similar process has to operate as phones leave an area.

The rules and actions on components of models for sensors, devices, places and people provide location information for applications. The system is not limited to location: it can notify applications when the context changes in any way. For example, if a user indicates a change their music preferences a simple active *component* rule can notify the music player.

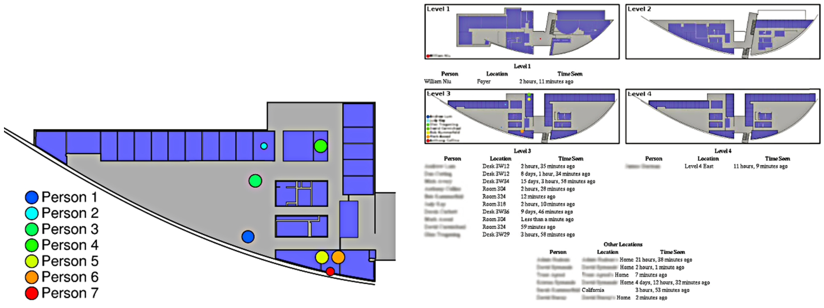
3 Evaluation

Since the goal of PersonisAD is to provide a framework for creating new ubiquitous applications, we have been evaluating it by using it to create the applications that we describe in this section. We first describe the actual applications. Then we provide some analysis of the implementation using PersonisAD. This gives an indication of the power and flexibility offered by the PersonisAD framework in terms of the amount of effort required to build a new application. In addition to this, the use of a consistent framework means that we can reuse active models: we illustrate this in the reuse of the first application, which models location, in the other two applications.

3.1 Location

This basic demonstrator application can help people finding others within our building. Using the location information, built from evidence collected from the sensors, a live image is generated and displayed on a webpage. Figure 6(a) shows a detailed display for just one wing of the building at the left. On the right, in Figure 6(b), is the full display for four floors. Each person's location is indicated by a coloured dot on the map at the place they were last seen. The key for these is shown at left of Figure 6(b) (anonymised for this paper). The age of this information is indicated by size of the dot on the map. So a person who has not been detected for some time has a smaller dot on the map. Each person's model is **ask**ed for that person's current location with a resolver that returns both the location and the timestamp indicating when that was last updated. This is used to provide the input to the map display. People have to subscribe to this service for their sensor evidence to be used. Users can control which resolver is used and so control the information available to this application.

The raw sensor data comes from two types of sensors; Bluetooth sensors detect Bluetooth devices within range, System activity sensors detect mouse and keyboard activity on a terminal. The Bluetooth activity sensors **tell** four kinds of evidence messages: on initial detection of a device a "Found" message is sent;



Level 1

Person	Location	Time Seen
William Niu	Foyer	2 hours, 11 minutes ago

Level 3

Person	Location	Time Seen
[redacted]	Desk 3W12	2 hours, 35 minutes ago
[redacted]	Desk 3W12	8 days, 1 hour, 34 minutes ago
[redacted]	Desk 3W34	15 days, 3 hours, 58 minutes ago
[redacted]	Room 304	2 hours, 28 minutes ago
[redacted]	Room 324	12 minutes ago
[redacted]	Room 318	2 hours, 10 minutes ago
[redacted]	Desk 3W36	9 days, 46 minutes ago
[redacted]	Room 304	Less than a minute ago
[redacted]	Room 324	59 minutes ago
[redacted]	Desk 3W29	3 hours, 58 minutes ago

Level 4

Person	Location	Time Seen
[redacted]	Level 4 East	11 hours, 9 minutes ago

Other Locations

Person	Location	Time Seen
[redacted]	[redacted] Home	21 hours, 38 minutes ago
[redacted]	[redacted] Home	2 hours, 1 minute ago
[redacted]	[redacted] Home	7 minutes ago
[redacted]	[redacted] Home	4 days, 12 hours, 32 minutes ago
[redacted]	California	3 hours, 53 minutes ago
[redacted]	[redacted] Home	2 minutes ago

(a) Detailed view of a single wing (b) Screenshot of the web interface

Fig. 6. An example of a map that plots the location data collected by the ActiveModels on a floorplan of our building. The radius of the circles are proportional to currency of the data, with smaller circles indicating older data.

every 5 minutes a “Present” is sent for each device in range; and a “Lost” message is sent when a device moves out of range. When no devices are in range a “Heartbeat” message is sent every 5 minutes to inform the system that the sensor is still alive. System activity sensors work in a similar way when detecting activity of users.

3.2 MusicMix

As described in Section 2.1, MusicMix is a demonstrator application that plays background music, generating the playlists based on who is currently in the general area. The application uses the location information provided by the pre-existing models, and combines this with user preferences expressed in terms of the individual’s playlists. There is an additional model for the player, representing the physical speaker in the room where the music is to be played.

The player model stores the list of the people who are nearby, the current track being played, and the status of that track. This is the core information needed to support this application. A rule is used to update the player model with the list of listeners. Whenever the list of people changes, the playlist manager is automatically made aware via a **notify** rule. It is then able to generate a new playlist by **ask**ing the user models for their preferences. A track is selected and **told** to the Speaker model. The Speaker Manager application controls the playing of music tracks: it is made aware of the current track by its own **notify** rule. When the track finishes playing, the model is updated again, causing the playlist manager to select a new track and the process repeats. It should be noted that there is no direct communication between these two applications. All their actions are a direct result of updates to the models: this ensures complete decoupling of application components. (For more technical details please see [9].)

3.3 MyPlace

MyPlace presents people with personalised information about a place. It builds models of the people so that it can reason about their knowledge of a place, their learning needs about it, their preferences for information delivery as well as their location. So, for example, when Alice first visits Bob at his workplace, MyPlace needs to give her relevant and timely information about Bob’s location. If we suppose that her friend Carol is also nearby, MyPlace should tell her so. Since this paper’s focus is the PersonisAD support for building applications, we give just a brief outline of MyPlace.

For example, the screenshots in Figure 7 illustrate the different information that MyPlace would present Alice and Bob. At the top of the screen is an information bar. She has selected her status as “In transit”, where Bob has selected “Available”. This provides *evidence* to the user models. Next to the user’s selected status is the location as modelled by the system: Alice is in the “Foyer” and Bob is in “324”, which is his office. The “details” link provides an explanation of how the system models location. The bulk of the display shows items modelled as relevant to the user. These are grouped into expandable headings of Devices at this location, Nearby Devices, Nearby Places, Services/Events, and People. If a category contains no items, the heading is not shown.

As a visitor to the building, Alice's information needs and therefore user model are very simple. The stereotype for a visitor states that they are interested in the devices of the Foyer Noticeboard (for general information) and the Information Kiosk (for contacting their host on arrival). Thus when Alice is located in the Foyer these items are shown in the "Devices at this location" category. Bob's user model indicates he wants to know about any sensors which may detect him. Accordingly, the "Devices at this location" category contains the Bluetooth sensor in his office, the system activity sensor in his office, details of the airconditioning systems control panel. Alice is shown the location of Bob's office in the "Nearby Places" category. For Bob, the "Nearby Places" category is empty and not shown, as he has worked at this location for some time, and thus already knows all the nearby places modelled to be of interest to him. In the "People" category, Alice is shown the location and status of Bob, as he is her host, and Carol, because the system models her as knowing Carol. Bob's "People" category contains details of Alice as he has a meeting scheduled with her, and David and Fred, two of his colleagues with whom he is working closely on a project. At the bottom of both Alice's and Bob's displays is a "show all items" link which allows them to see all the items which the system has chosen not to show them.

Status: In transit, Location: Foyer [details] User:**alice**[logout]

Devices at this location

- **noticeboard-foyer** Electronic personalised noticeboard.
- **information-kiosk** Information kiosk for contacting staff.

Nearby Places

- **324** Bob Smith's office

Services/Events (4)

People

- **bob** Available, 324
- **carol** Busy, 302

[Show all items]

Status: Available, Location: 324 [details] User:**bob**[logout]

Devices at this location

- **bspy-324** Bluetooth sensor.
- **aircon-324** Air conditioning control panel.
- **syssensor-324** System activity sensor.

Nearby Devices (4)

Services/Events (5)

People

- **alice** In transit, Foyer
- **david** In meeting, 302
- **fred** Coffee, 125

[Show all items]

(a) Alice (b) Bob

Fig. 7. Two user views in the MyPlace system, left (a) shows a view for Alice, a Visitor, and right (b) shows a view for Bob, her host

3.4 Analysis of Applications Built with PersonisAD

We now analyse the development of applications that have been built with PersonisAD. By looking at the amount of evidence we have collected, the models that have been created, the rules required and the code that has been written to build them, we show the power, flexibility and utility of the architecture.

Evidence. We have been collecting data for location modelling for 17 months. There are a total of 3,706,237 items of raw sensor data from Bluetooth and system activity sensors. Users must register themselves and their Bluetooth devices before their models are built. There are 21 registered users, 23 registered mobile Bluetooth devices and 22 system activity sensors. There are 12 fixed Bluetooth sensors placed in key locations around the building. Some users have only one system sensor associated with them, while others each have two Bluetooth devices (a phone and a laptop) and system activity sensors running on multiple machines (work desktop, home desktop, laptop, and another communal machine).

Table 1 shows a summary of one month's data for four users with diverse evidence sources. The `Active` column shows the number of data items indicating activity on that system. The `Heartbeat` column has the number of items showing that no activity has been detected on a mobile system. These messages have been omitted for system activity sensors whose location is fixed. The columns `R1,R2 ... R12` are Bluetooth sensors around the office building.

Table 1. A summary of one month of location data for four users with diverse evidence sources

User	Device	Active	Heartbeat	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
Person 1	phone			79	196	33	491	913	64	2	395	256	15		763
(academic)	laptop	615	117												
	Desktop	832													
Person 2	phone				14		766	55	11		1446	74		85	10
(postgraduate)	laptop	857	439												
Person 3	laptop						449		22						
(postgraduate)	phone			2	94	52	1296	1759	471	231	1032	235	71	9	209
	desktop	1257													
Person 4	mobile			10	214	18	541	677	63	55	2071	282	2	111	220

Location Models. In order to provide location data from the raw sensor data, models are required to represent the sensors, users, devices, and locations. Table 2 gives the number of models of each type in the system. In addition the number of components per model is shown. This indicates the amount of data required per model. Finally the number of subscriptions per model shows the complexity of the interactions between models and components. The low number of components per model shows the lack of complexity required to model these items using PersonisAD.

Table 2. Number of models, components and subscriptions used in modelling location

Type	Models	Components per model	Subscriptions per model
People	21	5	-
Places	66	4	1
Devices	23	5	3
Bluetooth Sensors	12	4	2
System Sensors	22	5	2

Application Development. This requires writing a number of simple rules that link the models together and to external applications. The MusicMix application required 4 rules to operate. This section describes one of those rules in detail to illustrate the character of the work involved in building systems based on PersonisAD.

```
player1:currentlyPlaying
   <./currentlyPlaying> ~ ".*" :
       NOTIFY 'http://player:2001/playSong?song=' <./currentlyPlaying>
```

Fig. 8. The code for the **currentlyPlaying** *component* of the `player1` model. When the **currentlyPlaying** *component* is changed, the notify rule alerts the Speaker Manager via the HTTP request, which begins playing the new track.

The rule shown in Figure 8 notifies the Speaker Manager of the address of the current music track to play. The `player1:currentlyPlaying` line shows that we are adding this subscription to the **currentlyPlaying** *component* of the `player1` model. This means that the rule will be examined whenever new evidence is told to this component. To decide if the rule should be fired the `<./currentlyPlaying>` value will be examined. The match performed is a regular expression, as indicated by the ∼ symbol. The expression matched is `".*"`, which matches all possible values. In this case the rule will always fire, performing the action `NOTIFY 'http://player:2001/playSong?song=' <./currentlyPlaying>`. This action makes a HTTP GET request with the given URL and appends the current value of **./currentlyPlaying**. The `./` indicates that the value comes from the current model. By inserting a model name here, a value could come from any local or remote model.

The simple rule language allows great flexibility in responding to new evidence in the models. Any change on any model can trigger a rule, and the rule can be triggered conditionally depending on the matching operator. When a rule is triggered, any information that can be obtained from the models can be passed to an application by varying the **notify** URL.

Code Effort. By keeping the management of the data within the models, and by describing the relationship between the models with the rules, the specific application code that needs to be written is very short. The MusicMix application required two small Python scripts to run: the Playlist Manager and the Speaker Manager. The Playlist Manager was the longer of the two being 95 lines, the Speaker Manager only 79 lines. Adding location support to the player models required adding a single rule. This means that simple applications can be built simply without limiting the complexity of larger systems.

The code for updating the location models based on sensor data is also quite simple. It consists of a small number of special *resolver* functions, and a short python script to do **tell**s whenever sensor data *evidence* is received. The architecture allows us to make the location modelling more intelligent by changing the *resolver* functions.

4 Related Work

A major contribution of our work is a generalised framework to simplify the creation of ubiquitous computing applications. There are many existing systems that already do this, but our PersonisAD provides a unique approach, by storing data within consistent models associated with people, places, devices and sensors, allowing applications to access these as required. Unlike other systems where applications are developed by combining application level building blocks, we focus on the storage of information about the entities within the environment, and the links between these entities. We review a selection of other systems to illustrate how they differ from PersonisAD; an exhaustive review is beyond the scope of this paper.

The work by Schilit et. al. [10] describes a system that reacts and changes to an individual's changing context. They have developed context triggered services by following a sequence of IF-THEN rules to decide how applications should adapt. By contrast, the main focus of our work is the modelling of the environment: the functionality of our rules is encapsulated by attaching **notify** rules to *components* which respond to changing context.

The Context Toolkit [11] allows applications to be built using a variety of input and output widgets, which encapsulate common programming tasks. An application controls how the input events map to the output actions. Metaglue [12] provides infrastructure support to embed controlling applications, or agents, within a wider scale pervasive environment. By contrast, PersonisAD has no concept of a controlling application. The information is stored within the models embedded in the environment and individual applications query this information to provide services to the user.

The Event Heap [13] is a tuple-based communication message architecture which has been extended to a distributed environment [14]. Applications filter the aggregated data based on subscription-like data requests, as opposed to PersonisAD where this filtering is a combination of the application and the models.

Another method of providing general ubiquitous computing services is Activity Zones [15], where location is mapped into a particular geographic zone. Certain rule-based actions are performed based upon the zone they are in. It is similar to our system in that actions are performed by following rules as evidence arrives, but PersonisAD is able to respond to more than location information.

Liquid [16] provides a framework for persistent queries where results are drawn continuously from both local and remote sources as events take place. This differs from PersonisAD in the way that queries are dynamically processed when a client application makes a request: the PersonisAD rule set results in the models being constantly kept up to date as new information is provided.

An important feature of the PersonisAD approach is that it builds upon the homogenous modelling of all the entities relevant to supporting ubiquitous applications [7]. Moreover, the underlying accretion/resolution representation supports flexible interpretation of a range of heterogenous sources of evidence about the user [17]. This distinguishes the work from the many systems [18] that model location using a single class of evidence, such as ultrasonic beacons [19,20], GSM Mobile phone networks [21], WiFi beacons [22,23] and Bluetooth sensors [23,24]. Each of these produce location evidence at different resolutions and reliability.

Our work has unusual origins for ubiquitous computing research in that PersonisAD grows from work on user modelling and personalisation [4,7]. This has influenced the philosophical underpinnings of the design of the accretion / resolution approach, which models people by collecting *evidence* about aspects of them, in the components of the model. The sources and nature of the evidence are important and play a role in the interpretation of the evidence by **resolvers**. Importantly, since user models are essentially repositories of personal information, the accretion/resolution approach supports scrutability, a notion that has also been explored by Cheverst et al [25]. The importance of user control, and the closely associated issues of privacy, have been recognised in ubiquitous computing from its earliest days as well as more recently [26]. One part of this is providing people with the right amount of information about the invisible sensors and services in the environment. Another important part is to provide the user with support to drill down into that information, to scrutinise the details of just what information or knowledge is held in the ubiquitous computing environment and how it is used.

5 Conclusions and Contributions

We have described PersonisAD, a framework for distributed, active, scrutable models as a foundation for building context-aware applications. We have explained how it has enabled us to quickly create a set of such applications. We have illustrated how the elements of the accretion/resolution approach, extended with active components provides a powerful, flexible framework for supporting ubiquitous computing applications. The foundation of the approach is the accretion of *evidence* from arbitrary sources into the *components* of the models. These are structured into an hierarchy of *model-contexts.* The access to information about entities in PersonisAD is controlled at the level of the model-context, the evidence source and type and by the resolver functions which are selectively available to different applications. This enables different levels of granularity of values to be returned for any component.

The current system operates with models distributed across different machines, but any one entity's model is on a single machine. So, for example one user's model is entirely on one machine. Future work will move towards distribution of partial models, especially for user models [27] so that subsets of the user models and some contexts within the model are located on different machines. Another future enhancement will support disconnected operation so that information flows, such as described in Section 2, can be supported even when parts of the system are temporarily disconnected. We are also conducting experiments to assess several of the dimensions of scalability, similar to testing the basic accretion/resolution implementation in Personis [7]. A different order of future work involves the interface design challenge of supporting scrutability in a ubiquitous environment. This can build from work on scrutably adaptive hypertext [28] as well as Cheverst et al [25].

The main contribution of PersonisAD is as a framework for context aware computing, supporting personalised ubiquitous services, integrating the collection of diverse types of evidence about the ubiquitous computing environment, interpreting it flexibly. In this paper, we have focussed on the distributed and active nature of the models.

The addition of triggers to the models is an important innovation. In contrast to our earlier work with passive models, where the application was responsible for all the control, PersonisAD makes it easier to build context aware applications. Essentially, this is because it decouples the systems more effectively. Of course, at one level, the notion of triggers is not new, since this class of mechanism has been part of many systems, such as the many rule-based systems, from the very early ubiquitous computing frameworks [10] and including general operating systems and Active Databases Systems [29]. The important contribution of our work is that the user model representation and reasoning, with its design for scrutability, has been generalised to modelling the other elements, sensors, devices, services and places.

It is this consistent and generalised representation of the context data that has made it feasible to use a simple mechanism to achieve distributed models. This is an important contribution as it means that the models can be kept in the places that make sense for various pragmatic reasons such as user control. The PersonisAD architecture is based upon a small set of operations for applications to interact with the models: *access*, *ask* and *tell*. The service discovery facilitates distributing the models across various machines across a network. The application writer can simply use these, in conjunction with the active models, to build applications. The service discovery for the models of a PersonisAD-based system ensures that different models can be dispersed across arbitrary machines in a network, without burdening the application builder. We have demonstrated the use of the PersonisAD framework by building the applications described in this paper. The analysis of this process indicates that the PersonisAD framework makes for low implementation cost with carefully decoupled elements. This is due to the consistent, simple, flexible and powerful representation of models for each of the elements in a context aware environment. The representation was initially motivated by the goal of building user models that can support scrutability of the modelling and personalisation processes. The accretion/resolution representation also supports flexible control of access to the models and the interpretation of uncertain and noisy evidence. PersonisAD uses this same representation for all the elements in the context-aware environment. This, combined with its carefully designed, simple but powerful rule language that is specific to the models, and transparent support for distribution of the models, provides a new way to think about and build context-aware applications.

References

1. Lamming, M., Flynn, M.: Forget-me-not: intimate computing in support of human memory. In: FRIEND21 Symposium on Next Generation Human Interfaces. (1994)
2. Dey, A.K., Abowd, G.D.: Cybreminder: A context-aware system for supporting reminders. In: Intl. Symposium on Handdeld and Ubiquitous Computing. (2000) 172–186
3. Beigl, M.: MemoClip: A Location-Based Remembrance Appliance. Personal Technologies **4**(4) (2000) 230–233
4. Kay, J., Kummerfeld, B., Lauder, P.: Personis: A server for user models. In: AH. (2002) 203–212
5. Kay, J.: A scrutable user modelling shell for user-adapted interaction. PhD thesis, Basser Department of Computer Science, University of Sydney (1999)
6. Kay, J., Lum, A., Uther, J.: How can users edit and control their models in ubiquitous computing environments? In Cheverst, K., Carolis, N.d., Kruger, A., eds.: Workshop on User Modeling in Ubiquitous Computing, 9th International Conference on User Modeling, Johnstown, USA (2003)
7. Carmichael, D., Kay, J., Kummerfeld, R.: Consistent modeling of users, devices and environments in a ubiquitous computing environment. User Modeling and User-Adapted Interaction **15** (2005) 197–234
8. Cheshire S., K.M.: DNS-based service discovery (2005)
9. Assad, M., Carmichael, D.J., Kay, J., Kummerfeld, B.: Active models for context-aware services. Technical Report TR 594, The School of IT, The University of Sydney, Australia (2006)
10. Schilit, B., Adams, N., Want, R.: Context-aware computing applications. In: IEEE Workshop on Mobile Computing Systems and Applications, Santa Cruz, CA, US (1994)
11. Dey, A.: Providing Architectural Support for Building Context-Aware Applications. PhD thesis (2000)
12. Coen, M., Phillips, B., Warshawsky, N., Weisman, L., Peters, S., Finin, P.: Meeting the computational needs of intelligent environments: The metaglue system. In: Proceedings of MANSE'99, Dublin, Ireland (1999)
13. Johanson, B., Fox, A.: The event heap: A coordination infrastructure for interactive workspaces. In: WMCSA, IEEE Computer Society (2002) 83–93
14. Storz, O., Friday, A., Davies, N.: Supporting ordering and consistency in a distributed event heap for ubiquitous computing. In: Second Workshop on System Support for Ubiquitous Computing Workshop (Ubisys 2004) in association with Sixth International Conference on Ubiquitous Computing (online proceedings), Nottingham, England (2004)
15. Koile, K., Tollmar, K., Demirdjian, D., Shrobe, H.E., Darrell, T.: Activity zones for context-aware computing. [30] 90–106
16. Heer, J., Newberger, A., Beckmann, C., Hong, J.I.: liquid: Context-Aware Distributed Queries. [30] 140–148
17. Kay, J.: The um toolkit for cooperative user modelling. User Modeling and User-Adapted Interaction **4** (1995) 149–196.
18. Hightower, J., Borriello, G.: A survey and taxonomy of location systems for ubiquitous computing. Extended paper from Computer **34**(8) (2001) 57–66
19. Harter, A., Hopper, A., Steggles, P., Ward, A., Webster, P.: The anatomy of a context-aware application. In: Mobile Computing and Networking. (1999) 59–68

20. Minami, M., Fukuju, Y., Hirasawa, K., Yokoyama, S., Mizumachi, M., Morikawa, H., Aoyama, T.: Dolphin: A practical approach for implementing a fully distributed indoor ultrasonic positioning system. In Davies, N., Mynatt, E.D., Siio, I., eds.: Ubicomp. Volume 3205 of Lecture Notes in Computer Science., Springer (2004) 347–365
21. Otsason, V., Varshavsky, A., LaMarca, A., de Lara, E.: Accurate GSM Indoor Localization. [31] 141–158
22. LaMarca, A., Hightower, J., Smith, I.E., Consolvo, S.: Self-mapping in 802.11 location systems. [31] 87–104
23. Hightower, J., Consolvo, S., LaMarca, A., Smith, I.E., Hughes, J.: Learning and recognizing the places we go. [31] 159–176
24. Madhavapeddy, A., Tse, A.: A study of bluetooth propagation using accurate indoor location mapping. [31] 105–122
25. Cheverst, K., Byun, H.E., Fitton, D., Sas, C., Kray, C., Villar, N.: Exploring issues of user model transparency and proactive behaviour in an office environment control system. User Modeling and User-Adapted Interaction **15**(3-4) (2005) 235–273
26. Rehman, K., Stajano, F., Coulouris, G.: Visually interactive location-aware computing. [31] 177–194
27. Brar, A., Kay, D.J.: Privacy and security in ubiquitous personalized applications. Technical report, The School of IT, The University of Sydney, Australia (2004)
28. Czarkowski, M., Kay, J.: Giving learners a real sense of control over adaptivity, even if they are not quite ready for it yet. In: Advances In Web-Based Education. Information Science Publishing (2006) 93–126
29. Paton, N.W., Díaz, O.: Active database systems. ACM Computing Surveys **31**(1) (1999) 63–103
30. Dey, A.K., Schmidt, A., McCarthy, J.F., eds.: UbiComp 2003: Ubiquitous Computing, 5th International Conference, Seattle, WA, USA, October 12-15, 2003, Proceedings. In Dey, A.K., Schmidt, A., McCarthy, J.F., eds.: Ubicomp. Volume 2864 of Lecture Notes in Computer Science., Springer (2003)
31. Beigl, M., Intille, S.S., Rekimoto, J., Tokuda, H., eds.: UbiComp 2005: Ubiquitous Computing, 7th International Conference, UbiComp 2005, Tokyo, Japan, September 11-14, 2005, Proceedings. In Beigl, M., Intille, S.S., Rekimoto, J., Tokuda, H., eds.: Ubicomp. Volume 3660 of Lecture Notes in Computer Science., Springer (2005)

An Exploration into Activity-Informed Physical Advertising Using PEST

Matthias C. Sala, Kurt Partridge, Linda Jacobson, and James “Bo” Begole

Palo Alto Research Center (PARC), Computer Science Lab,
3333 Coyote Hill Road, Palo Alto, CA 94304, USA
`salam@alumni.ethz.ch`,
`{kurt,bo,ljacobson}@parc.com`
`http://www.parc.com/csl`

Abstract. Targeted advertising benefits consumers by delivering them only the messages that match their interests, and also helps advertisers by identifying only the consumers interested in their messages. Although targeting mechanisms for online advertising are well established, pervasive computing environments lack analogous approaches. This paper explores the application of activity inferencing to targeted advertising. We present two mechanisms that link activity descriptions with ad content: direct keyword matching using an online advertising service, and “human computation” matching, which enhances keyword matching with help from online workers. The direct keyword approach is easier to engineer and responds more quickly, whereas the human computation approach has the potential to target more effectively.

Keywords: Ubiquitous computing, experience sampling method, human computation, advertising.

1 Introduction

Futurists have often presented visions of extremely personalized advertising. In the motion picture Minority Report (2002), for example, a grinning Gap employee appears on a huge, three-dimensional display at the store entrance. “Hello, Mr. Yakamoto,” the employee says loudly and cheerily. “How did those assorted tank tops work out for you?”

Many find this vision of targeted advertising unsettling on several counts: the system initiates the interaction, causing the consumer to feel out of control; the system appears to be monitoring all of the consumer’s purchases, which many consider a privacy violation; the avatar announces the consumer’s purchase loudly to anyone within earshot; and the cheerful welcome is unmindful of the consumer’s mood and personality.

In contrast to this Orwellian scenario, we believe that future pervasive, targeted advertising systems will treat us, the consumers, with respect. We will not be bombarded by messages; instead, we will receive only the messages we want. We will receive information tailored to our interests, our appetite for

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 73–90, 2007.

information, and our desired level of participation. We will be able to opt in or out, and optionally communicate our reactions to friends, neighbors, and the advertising system itself.

We are optimistic because, from the consumer's perspective, trends have been positive. Digital video recorders have empowered consumers to quickly skip unwanted commercials. Pop-up blockers have reestablished consumer supremacy in online advertising. And Google's highly successful advertising services have maintained a strong separation between paid and unpaid content. Advertisers, too, prefer targeted advertising, as it more likely leads to sales. Progress is being made toward marketeer John Wanamaker's famous phrase, "Half the money I spend on advertising is wasted; the trouble is I don't know which half."

This paper anticipates another technological development: activity-targeted advertising, the combination of advertising and activity recognition. Currently in an early research stage, activity recognition electronically detects a person's real-world activity using ambient and worn sensors, electronic data sources, and artificial intelligence. We believe that activity recognition will someday achieve sufficient accuracy rates to enable many new applications, including highly targeted advertising. Activity-targeted advertising goes beyond current interactive electronic advertising systems such as Cellfire, Inc. [10], Reactrix [19], Freeset Human Locator [13], and P.O.P. ShelfAds [18], which do not present advertisements related to a consumer's activity.

This paper's primary contribution is the exploration of activity-targeted advertising as a novel application. We describe an architecture and our implementation of this architecture, present a preliminary evaluation of our system's effectiveness, and describe what we learned about the issues involved. A particularly novel aspect of our approach is the use of Amazon.com's Mechanical Turk[1]. We used Mechanical Turk as a system component to emulate portions of the system that are not yet mature enough to be performed automatically.

The next section describes related work. Section 3 explains how advances in online advertising has motivated our approach. Section 4 describes our user study design and the system architecture. Section 5 contains the study results, and Section 6 discusses implications. Section 7 covers future work and our conclusions.

2 Related Work

Much attention recently has been directed toward location-based services, including location-based advertisement delivery. For example, Aalto et al. [1] use Bluetooth beacons to determine when to push SMS advertising messages. Location-targeted advertising is related to our goal because location may determine a person's activity. Consider the different activities that happen in work, school, home, restaurant, or church.

However, location is only one of many indicators of user intention [20] and there are reasons to believe targeting activity could be more effective.

---

[1] The authors are not affiliated with Mechanical Turk or Amazon.com.

Some locations are correlated with a certain set of activities, but location-targeted advertising can only target the set as a whole, not the immediately occurring activity. Also, some activities can be done anywhere, and the recent improvements in communication and mobile technologies have further decoupled activity and location.

A few other systems have suggested extending the types of context beyond location to other sensed data [3,22]. However, we are not aware of any that have specifically examined the relationship between advertising and a higher-level model of activity.

Various research groups have investigated approaches for automatically inferring activity by parsing calendar entries [17], collecting data from infrastructure sensors (e.g., cameras and microphones) [15,21,24], and using wearable sensors [12,14]. But to our knowledge, no activity-sensing research has investigated advertising applications.

3 From Online Advertising to Activity-Targeted Advertising

Advertising is undergoing a revolution. In 2006, the online advertising hit $20 billion, reaching a record 12% of all advertising [16]. Much of the market share increase can be attributed to two techniques, *contextual advertising* and *behavioral targeting*, which work especially well online.

Contextual advertising is advertising positioned near content of a similar type. For example, a cooking tool advertisement might be placed alongside a recipe. While contextual advertising developed long before the Internet, keyword analysis has automated the placement of advertisements, making contextual advertising effective and efficient for advertising in narrow markets.

Behavioral targeting makes advertising more personalized. Online, it works by monitoring all sites that a user visits and what they do at each site to choose the best possible advertisement for that person. If a person buys a snowboard and ski boots online, then they might see advertisements for nearby ski venues.

This paper studies “activity-targeted advertising,” which resembles both these techniques. Like contextual advertising, it presents advertisements related to a person’s immediate interest, and like behavioral targeting, it tracks a person’s actions over time. But unlike both of them, it presents advertisements while the person’s main task is something other than consuming media. For example, someone cleaning the bathroom might hear a radio advertisement for soap. Someone driving to the hardware store might see an electronic billboard advertising a store that is open later. Someone sitting in a meeting might see a laptop screensaver advertising time-management products.

The architecture of an activity-targeted advertising system comprises three parts: 1) identifying the person’s activity, 2) appropriately targeting the advertisement to that activity, and 3) presenting the advertisement.

Part 1, identifying the person’s activity, is an active research area in pervasive and ubiquitous computing. As explained in Section 1, we believe that this technology will become accurate enough for activity-targeted advertising, but it is not yet mature. Our system handles this problem by having users self-report their activity.

Part 2, targeting the advertisement, is the main focus of this paper. We employ a keyword-based approach which allows a rich representation of activity and simplifies the mechanism for matching an advertisement.

Part 3, presenting the advertisement, is more complex than it is for online advertising, in which the consumer's activity is always "media consumption." One reason presentation is more complicated is because activity-targeted advertising requires careful timing. The consumer may be engaged in an activity that makes her unreceptive to advertising, or only receptive at specific times during the activity. Consider, for example, an electronic billboard on a tennis court. Showing advertisements while a point is being played is likely to be distracting and ineffective; displaying advertisements between points may be more acceptable.

Another issue concerning the presentation of activity-targeted advertising is privacy protection. Not only are there the traditional privacy concerns that collected data might be misused, but a person might also be embarrassed if a displayed advertisement is related to an activity that they do not want disclosed. It is not enough to screen out content that is publicly inappropriate (e.g., sexual products); the system must also understand the context (e.g., not show advertisements related to a purchased gift when the consumer is with the intended gift recipient).

In this paper, because we focus on Part 2, we take a simple approach to advertisement presentation. Our system presents advertisements on a mobile device. Users of the system are free to configure the mobile device to "vibrate" mode and ignore it at inappropriate times. The small mobile device screen also makes the advertisements effectively private. A commercially viable system is likely to use a different presentation method, but our approach was enough for studying targeting effectiveness.

4 Study Design and System Architecture

To explore activity-targeted advertising effectiveness, we conducted an experience-sampling user study in which a participant's self-reported activity was used to generate an advertisement, which she then rated according to relevance and usefulness. During the study, each participant carried a mobile device that regularly queried their activity throughout the day, displayed an advertisement, and requested their reaction to the advertisement.

The system selected advertisements using one of the following methods:

1. random selection
2a. treating the self-reported activity description as a web search query, whose result pages were searched for advertisements
2b. passing the activity description to a Mechanical Turk task that generated a search query, whose result pages were searched for advertisements

We did not inform participants of the mechanism used to find the advertisements.

The study was performed in two phases. The first phase compared conditions 1 and 2a. The second phase compared conditions 1 and 2b. Each phase had a different group of participants.

4.1 Participants

We recruited 19 participants from our in-house research staff: 17 male, and 2 female, aged 21 to 55. Previous experience with mobile devices was not required. To assess their exposure to advertising, we administered a pre-experiment questionnaire. The questionnaire showed that the participants on average watch 0.8 hours of TV, listen to almost one hour of radio, and spend 2.7 hours online. No participant categorically opposed advertising, and many said that they "like good ads."

As an incentive, we gave participants a small gift (T-shirt, tote bag, etc.), and allowed them the general use (web browser, email, etc.) of the device the study software ran on.

Of the 19 participants, 13 participated in Phase 1, and 6 participated in Phase 2. We would have liked to include more participants in Phase 2, but were not able to do so because of resource constraints.

It is worth noting that our recruiting procedures may have affected the results. Because our recruitment notice indicated that the study involved exposure to advertisements, people who highly disliked advertising might have chosen not to participate.

4.2 Proactive Experience Sampling Tool (PEST)

The Experience Sampling Method (ESM) is often used in research on human activities [4,6,7,9,11]. In ESM, the participant conducts their normal tasks, but is interrupted occasionally to report on what she is doing or has done. The study data can be collected in various ways, such as by phone call, using a timer and paper form, or, as we did, using a device that combines the timer and data collection.

To conduct this study, we developed a tool for performing experience sampling: the Proactive Experience Sampling Tool, or PEST (pun intended). PEST is a .NET library for Windows Mobile devices. PEST supports complex and interactive surveys by providing features for scheduling, logging, automated serialization of the input, and uploading survey responses through a wireless cellular link. PEST is also general enough to support downloading and presentation of the relevant advertisement. In this study, PEST's structure also made it easy to migrate the implementation from Phase 1 (simple keyword targeting) to Phase 2 (Mechanical Turk targeting).

Type	Wording	Form of answer
a Textbox	Where are you right now?	one line free text
b Textbox	What are you doing right now?	one line free text
c Textbox	What had you expected to be doing?	one line free text
d Textbox	What would you rather be doing now?	one line free text
e Button	Send this ad to my email	
f Slider	How relevant was this ad?	0–10
g Slider	How useful was this ad?	0–10
h Slider	When would it be useful?	*a month ago*, *a week ago*, *24 hours ago*, *an hour ago*, *5 minutes ago*, *right now*, *in 5 minutes*, *in an hour*, *in 24 hours*, *in a week*, *in a month*
i Checkbox	Useful at any time	checked/unchecked

Fig. 1. List of questions asked during a single survey

4.3 Survey Administration

Each participant carried a mobile device (an iMate JAM or a Mio DigiWalker) for a 72-hour session, either on weekdays (Monday to Wednesday) or on a weekend (Friday to Sunday). To accommodate personal schedules, a participant could configure the device to recognize sleeping hours, during which no alerts would occur. Participants could also choose between audible or vibrating alerts.

Alerts were scheduled randomly. During waking hours, alerts were scheduled at random intervals chosen with uniform probability between 25 and 65 minutes. These values were determined through pilot testing to balance observation frequency and participant irritation. With this approach, different participants were exposed to differing numbers of alerts, but the randomness also reduced the probability that the participant's activities, when occurring at a similar frequency as the alert schedule, would be over- or under-represented. We chose to use only time to schedule alerts to keep our implementation simple and to broadly cover participants' activities.

After each alert, the participant could chose to take the survey, postpone the survey for 30 minutes, or postpone the survey for 90 minutes. These amounts were also determined by pilot study observation.

4.4 Survey Questions

The survey comprised three parts: 1) asking questions about the participant's location, activity, expected activity, and preferred activity; 2) displaying the advertisement, and 3) asking questions about the appropriateness of the advertisement.

Fig. 1 lists the questions and their format. To reduce the time needed to complete a survey, the survey allowed participants to choose from a list of their previously entered activities for questions *a*–*c* rather than enter a new answer each time. To facilitate quick activity-list searching, and to avoid biasing responses that favored the top entries, we presented the list in alphabetical order, starting at a random point, and wrapping around at the end of the alphabet back to the beginning.

Question *a* asked the participant's location. While our primary purpose was to determine whether activity could lead to more relevant advertising, we also evaluated how location affected self-reported activity, and whether location affected the relevance and usefulness of the advertisement.

Question *b* asked the participant to identify their activity at the time of the alert. Questions *c* and *d* were a variation on this question. Question *c* asked what activity the participant had planned to perform at the current time. This question was designed to collect information like what would be in a typical calendar entry; we wanted to investigate whether targeting one activity was more effective than targeting another. Question *d* measured what the person would have preferred to be doing at the time of the survey alert. Here again, we speculated that advertising might be more effective if it could target something other than the user's exact activity such as what the user desired to be doing.

These first four questions allowed considerable freedom in the participant's response. As a result, participants interpreted and responded to the questions in different ways. For example, in response to the query about current activity, some participants answered in the abstract (such as "working") while others provided specific detail (e.g., "reading Da Vinci code"). This variety made it difficult to taxonomize the activities. However, it also made the descriptions more likely to reflect accurately how individuals conceptualize their activities. We felt that it was important to collect unbiased data, even if it was more difficult to analyze.

Following these questions, the advertisement was shown. The advertisement was fetched from online sources, using one of the methods described in Section 4.5.

Questions *e*–*i* measured the effectiveness of the displayed advertisement. Question *e* caused the advertisement to be emailed to the participant, and was therefore an implicit interest indicator [5].

Questions *f* and *g* were the primary metrics. "Relevance" measures how well the participant determined that the advertisement was activity-targeted—how well it matched their current activity. "Usefulness" was the participant's assessment of how helpful the advertisement was. Although the ideal advertisement would be both relevant and useful, an advertisement could vary in either dimension.

Finally, questions *h* and *i* investigated the timing of the advertisement. In pilot studies we noticed that some participants assigned a potentially useful advertisement a low score if it was not welcome at the immediate time of the survey. These questions helped identify these situations.

4.5 Advertisement Targeting Mechanisms

Phase 1: Simple Keyword vs. Random. The Phase 1 procedure for finding an activity-targeted advertisement started by determining the keyword to use.

Initially, with one-third chance, either a random word from Ogden's Basic English Word List of the 850 most common English words, the current activity description, or the planned activity description was chosen. For the activity descriptions, stop-word removal and word-stemming was performed to increase the chance of an advertisement being associated with the word. The terms were passed to a search engine. If an advertisement was found in the search results, one was selected at random. If an advertisement was not found, the system fell back to random selection. If random selection failed, a new random word was selected. To maintain the correct proportion of the experimental conditions, the system tracked the number of randomly selected versus activity-targeted advertisements and adjusted the probabilities for each condition.

During Phase 1, we observed the following causes of poor matches using simple keyword matching.

Unable to Determine Role in Activity. Advertisements were sometimes useful for a given activity, but only for a person with a specific role. For example, one participant who entered "waiting in the line" for his activity received an advertisement for "queuing solutions." This type of advertisement is unhelpful for those standing in a line, but might be useful for an individual servicing the line. The search engine was not able to know the role the participant was playing in the activity.

Poor Situational Need and Timing. Advertisements related to the current activity sometimes had a short window of time during which they were useful. For example, while "playing tennis," an advertisement promoting tennis rackets might be displayed either too late if the participant already had a racket or too early if the participant was not able to immediately purchase a new racket.

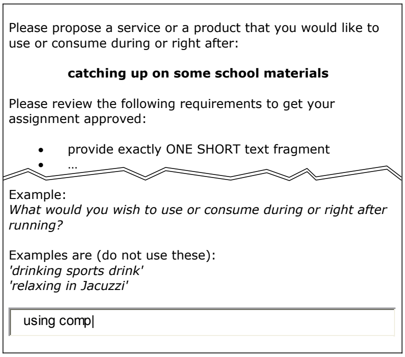
Overly Specific Search Space. If a participant entered an overly-specific activity, such as "waiting for the cashier," the system could not easily infer that the participant was shopping. If it could, then it might have been able to find a more appropriate advertisement.

Overly General Search Space. Conversely, if a participant entered an overly general activity, the system sometimes selected an advertisement appropriate to a time or situation different from what the participant faced in the moment.

Lack of Knowledge of Activity Sequencing. The system also could not easily determine what activity a participant might do next. Being able to predict future activities, even without perfect certainty, would have also improved match quality.

All these aspects are difficult for an automated system to infer, but trivial for a human to understand. This observation led to the approach used in Phase 2.

Phase 2: Mechanical Turk Targeting vs. Random. Mechanical Turk is an Amazon.com web service in which structured tasks can be completed by a member of a pool of workers in exchange for payment. Each task, called a



Please propose a service or a product that you would like to use or consume during or right after:

catching up on some school materials

Please review the following requirements to get your assignment approved:

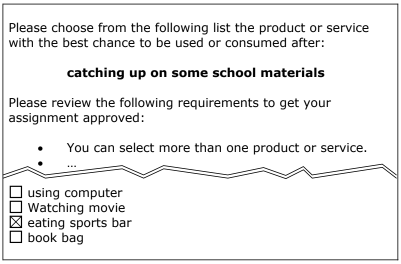
- provide exactly ONE SHORT text fragment
- ...

Example:
What would you wish to use or consume during or right after running?

Examples are (do not use these):
'drinking sports drink'
'relaxing in Jacuzzi'

using comp|

Fig. 2. Idea generating task



Please choose from the following list the product or service with the best chance to be used or consumed after:

catching up on some school materials

Please review the following requirements to get your assignment approved:

- You can select more than one product or service.
- ...

☐ using computer
☐ Watching movie
☒ eating sports bar
☐ book bag

Fig. 3. Voting task

Activity	**Answers**
drinking wine	dark chocolate eat sharp cheese
eating lunch	watching television news chewing gum brush my teeth
catching up on some school materials	using computer drinking Pepsi-Cola Watching movie eating sports bar book bag

Fig. 4. Mechanical Turk answers in response to various activity prompts

Human Intelligence Task (HIT), is invoked as a web service, parameterized upon invocation, and passed to the worker. After completion, the results are sent back to the invoking program. The system is appealing because it allows human effort to be treated in the same way as a procedure invocation. See Barr and Cabrera for other application examples [2].

In our system, each HIT displayed the activity term entered by the participant and asked the worker to suggest an appropriate product or service. Fig. 2 shows the format of the HIT question. Each response was awarded 10 cents. Fig. 4 shows some example responses to Mechanical Turk queries.

In our system, an immediate response was required because the participant was waiting to view the advertisement. Our initial experiments showed most HITs were completed after a few minutes. We felt that this was too long to make participants wait. We experimented with increasing the amount paid to 40 cents, but this did not affect the time-to-completion.

To handle the long delay, we redesigned the experiment to ask participants to evaluate an advertisement based on the activity entered during the previous survey rather than the current one.

In pilot tests we observed that some of the Mechanical Turk workers' responses did not lead to useful advertisements, for the reasons explored in Section 5.6. Mechanical Turk supports oversight by allowing the HIT designer to pay workers only for appropriate responses. However, we felt that this mechanism was inappropriate for our task, because a quick response was necessary, and because having an expert approve all Mechanical Turk responses would not scale up well.

Instead we designed an approach that used a second round of Mechanical Turk tasks to perform oversight. Our implementation worked as follows. Five workers generated suggestions using the HIT shown in Fig. 2. Our system then created a second HIT for ranking the suggestions, shown in Fig. 3. Ten workers completed this HIT for 7 cents each. The suggestions were ranked by counting the number of times they were selected. The activity suggestion with the largest number of votes was chosen as the term to use to find the advertisement. If there were several equally ranked suggestions, one was chosen randomly.

5 Results

Participants filled out a total of 310 surveys. On average, each filled out 16 (the standard deviation was 9.4), giving a response rate of about 35%. This rate is low compared to similar studies [7]. We found several reasons for this during the post study interviews: participants who carried the device during weekdays had a low response rate at work; a few forgot to recharge the device overnight; others struggled with instability problems in the operating system and other installed programs.

5.1 Relevance and Usefulness

Fig. 5 shows the distributions of ratings. The left plots show keyword targeting, the right plots show Mechanical Turk targeting. The upper plots show the relevance, the bottom, usefulness. Within each of the four panes, the upper plot shows the random baseline, and the lower plot shows the results with targeting. Each light-colored point represents the averaged score of one user, and the points are jittered in the vertical direction for better visibility. The dark point indicates the median value, and the gray box marks the 50-percentile region.

One-tailed t-tests for comparing the means of these distributions show a statistically significant improvement in only the relevance score between random and the keyword approach ($p < 0.05$).

Although no statistically significant differences were found for the Mechanical Turk results, there were occasions in which Mechanical Turk performed well where a pure keyword-approach would fail. For example, in response to the activity "reading newspaper," the Mechanical Turk response suggested "drink good coffee," and the system then selected an advertisement for an online coffee retailer. The participant gave this response a 6 for relevance. From examples like this, we believe that an effect might be observed for the Mechanical Turk condition with a larger data set.

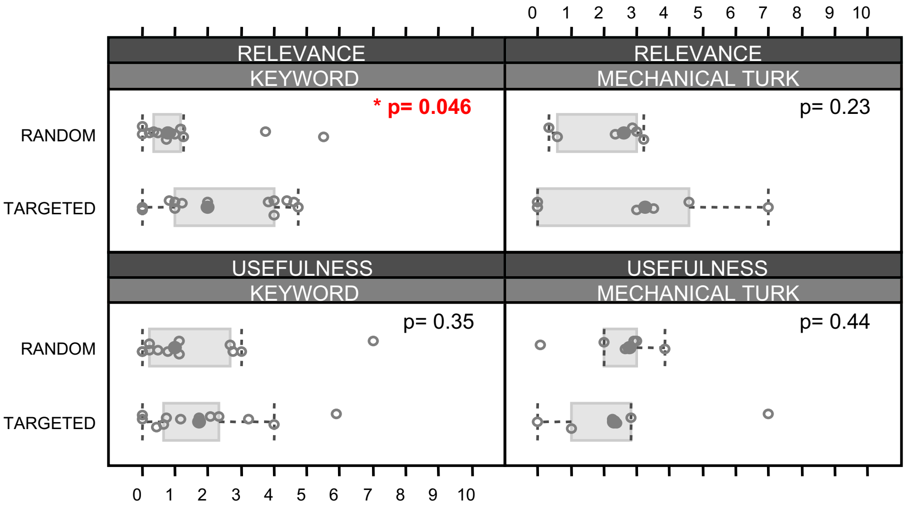


Fig. 5. Distribution of ratings, presented as box-and-whisker plots with the numbers of surveys at each rating. Most advertisements are considered irrelevant and not useful.

Participants did request that several of the advertisements be sent to them by email. While those selected did have on average higher relevance and usefulness scores, the result was not statistically significant.

5.2 Activity

Since participants entered activities using unrestricted text, many different entries represented the same or similar activities. To compare the effects of targeting in activities with different qualities, we categorized the activities into groups according to our observations of the data. See Table 1 for example responses for each category.

Fig. 6 shows the frequency distribution and ratings distribution for each activity class. Frequency distributions are computed separately for current activity, expected activity, preferred activity, and Mechanical Turk-targeted previous activity. We treated Mechanical-Turk-targeted activities separately because the advertisement corresponded to an entry from the previous survey, not the current survey. Most activities have roughly the same distribution, with the exception of "Constructive Mental," which included work, and was not as often chosen as the preferred activity; "Communication," which was less frequently expected or preferred; and "Eating," which was a frequent preferred activity. Both "Media Consumption" and "Eating" rate show better improvement over random than the other activities; we believe this happens because these activities are highly consumptive, so it is easier to directly match advertisements to needs. "Shopping" does not show such an improvement, but the sample size of shopping was small, and several of those advertisements targeted vendors rather than consumers.

Table 1. Our rough categorization of activities based on participant responses. The second column gives an example for each category.

Category	Example
Constructive Mental	"experiment analysis"
Communication	"in a meeting"
Media Consumption	"browsing internet"
Eating	"eating lunch"
Transporting	"going to neighbor"
Manipulating Objects	"laundry"
Live Observation	"attending seminar at work"
Fixing	"trying out new software"
Game Playing	"playing a game"
Thinking and Planning	"thinking about my life"
Shopping	"picking up prescriptions"
Basic Needs	"shaving"
Other	"wait for meeting"
Exercise	"gardening"
Mobile Entertainment	"sightseeing"

We did not observe significant differences between ratings targeted to current, expected, and preferred activity. For the data we collected, this is not surprising, since in most cases the answers to the questions were similar.

5.3 Location

We categorized locations using a similar method to our categorization of activity. Most participants reported their location as at the office or at home. As with the activity breakdown, it is perhaps not surprising that the system performed well in restaurants, at which most participants were engaged in a consumptive activity.

5.4 Timing

On average, participants indicated that "no time" was appropriate for 60% of all advertisements (because they did not reach sufficiently high usefulness scores.) Of the remaining advertisements, participants reported that 25% would be appropriate at "any time," with the remaining 15% equally distributed over the times given in Table 1. Figs. 5–7, when restricted to groups of advertisements with different timing ratings, look roughly the same. Here too, with more data, an effect may emerge.

5.5 Focus Group Discussion

Several participants had technical problems with the study, such as poor connectivity and short battery life. Most found that the device's ergonomics to be adequate, although participants who already carried a cell phone did not enjoy

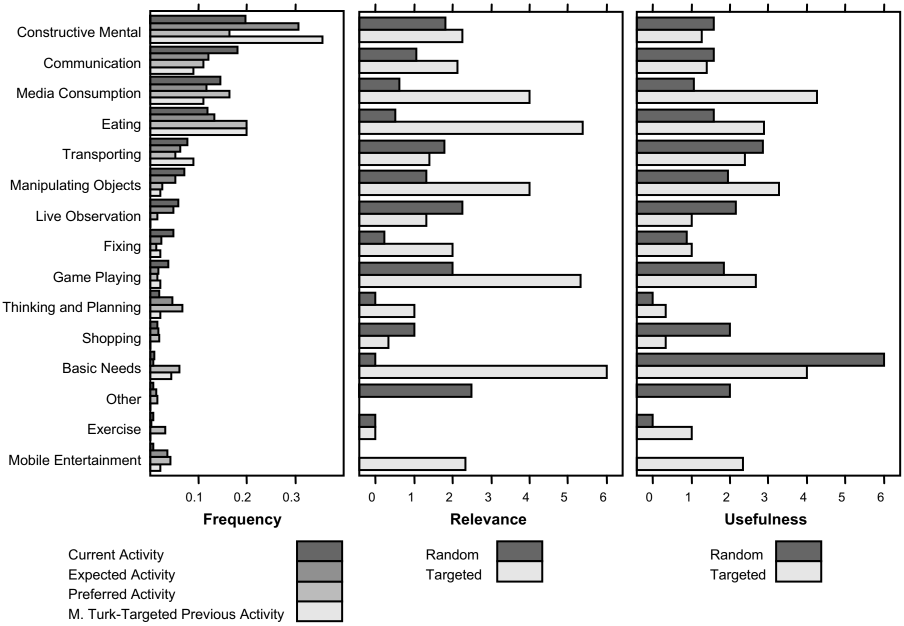


Fig. 6. Breakdown by activity. The leftmost chart shows the frequency of each activity category occurring, sorted by the current activity. The center chart shows the average relevance score for random and targeted advertisements for each activity, and the right chart shows the same measurements for usefulness. Because of small numbers of observations, measurements toward the lower end of these two charts should be treated as increasingly less reliable.

having two devices to carry. Text-entry with the soft keyboard was generally adequate, although participants frequently reused their previous entries. Several participants expressed confusion over the seemingly random alphabetic ordering of the activity list, which we did not explain in advance.

Many participants found the periodic alerts intrusive, and commented that the alert frequency missed some activities and oversampled others. One participant commented, "Normally, I work longer than only one hour on the same task." Participants suggested using other mechanisms for triggering surveys, such as location-based alerting, time-sensitive alerting, and participant-initiated activity reporting.

The advertisements themselves generally seemed to displease the participants. All agreed that most advertisements were irrelevant, but on the other hand, some admitted that they did mail some advertisements back to themselves for personal follow-up or to forward to others. Participants especially seemed to dislike advertisements that were generic if they were shown a similar advertisement more than once.

The survey questions were generally considered straightforward, although some seemed to be slightly confused about the distinction between "relevance"

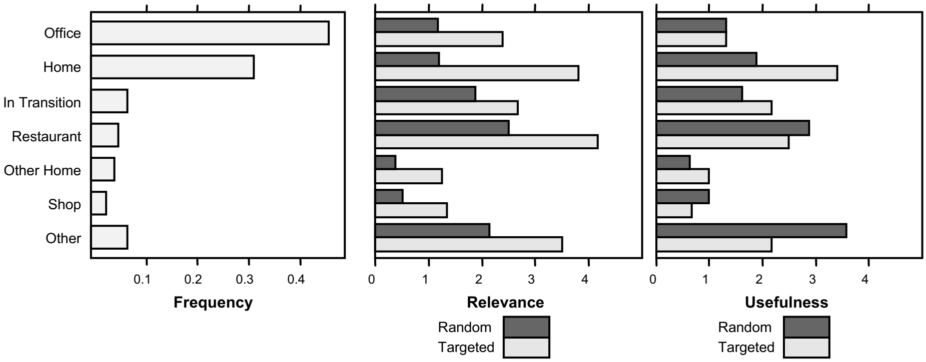


Fig. 7. Breakdown by location. See the caption for Fig. 6.

and "usefulness." The net effect appears to be that our quantitative estimates of usefulness are probably lower than what they would be if the participants had been clear about the distinction. Unfortunately, from the qualitative focus group discussion, we are not able to determine a quantitative difference.

Most participants reported that advertisement exposure was more frequent than they would have preferred. In exploring alternatives, there was disagreement as to whether advertisements should appear for different durations ("one-second advertisements that do not take too much of my time") or whether they should be bunched together ("I like the idea of browsing through coupon booklets").

Participants offered several suggestions for improving the system:

1. Advertising would be more effective if displayed during idle moments rather than during active tasks.
2. Advertisements might be targeted not to the current activity, but to a complementary activity. For example, if a person is unemployed, an advertisement for a job board would be useful.
3. A system could maintain a history of advertisements, which would support later review of advertisements.
4. A system might proactively determine what items a participant is likely to later spend time searching for on the web, and display the search results along with advertising.
5. A system might learn a person's ratings over time and use this information to better target advertising.

Finally, in some cases participants were surprised by PEST's supposed ability to find obscure, but relevant advertisements. However, upon further investigation we discovered that these advertisements were in fact random, and that the participants had attributed more intelligence to the system than it actually possessed.

5.6 Findings Regarding the Use of Mechanical Turk

Although we found our use of Mechanical Turk to be effective, many results failed to adequately help find an appropriate advertisement. The oversight system did help improve response quality, but it did not completely screen out poor responses, and over time its performance degraded slightly. We observed the following categories of problems.

Minimal Responses. Some users entered a meaningless response of only a few characters. We suspect that these users may have realized that we initially rewarded all responses.

Genericity. Many answers were generic, and could be applied to any activity. For example, one worker suggested “drinking coke” in response to every task. We do not know if this response arose from an intentional disregard of the question, or if was generated from an automated script created specifically to answer all our HITs.

Misinterpretations. In response to the activity “meeting,” one worker suggested “having sex.” While semantically appropriate for some meanings of “meeting,” this answer generated a text advertisement that did not fit the work context of our participant, who gave it low ratings. We noticed that humorous responses in general were more likely to pass the second-level filter, regardless of their appropriateness.

Despite these issues, we were able to get many appropriate responses from Mechanical Turk. Overall, we found Mechanical Turk to be a useful complement to computation, and would use it in other projects.

Several modifications to our basic architecture might improve the effectiveness of the Mechanical Turk component. Mechanical Turk workers in the filtering stage might select among a list of advertisements rather than a list of keywords. Amazon.com’s “Qualifications” test might be applied to improve the average worker’s skill level. And finally, the tasks might be redesigned to make them more engaging or game-like [23].

6 Discussion

We were disappointed that usefulness scores were low in both phases. Efforts to improve usefulness through better targeting could combine other knowledge such as a user’s preference profile, his behavior history, his context, and his similarity to other users. Furthermore, an advertisement might be more effective if it targeted not a single activity, but a sequence of predicted future activities. Finally, a system might target advertising to a group of users rather than a single user.

Usefulness might also be improved by working on the presentation of the advertisement. As mentioned in Section 3, this work adopted a simple presentation so we could focus on targeting. The focus group discussions, however, made clear the importance of presentation, especially timing. To be effective, an advertisement must be shown when a consumer is receptive to it, which is not always when the activity is taking place. But, if the user is bored or not especially engaged with their activity, then they are more likely to be receptive. Activity detection may be more useful for determining *when* to advertise than *what* to advertise.

Finally, many hold concerns not only about advertising's invasiveness, as depicted in our opening scenario, but also about its other effects on consumers and on society. Some see advertising responsible in part for creating false needs, distorting social roles and relationships, replacing intellectual discourse with imagery, and weakening democracy [8]. Whether activity-targeted advertising contributes to these problems or helps fix them is not yet clear. We believe that the only way to predict its effects is to understand it better by building prototypes, evaluating them, and publishing the results.

7 Conclusions

We have explored how activity-informed advertising might be implemented, and our exploration has taught us several lessons. First, we have found that activity-targeted advertising can be statistically more relevant than randomly targeted advertisements. Activity targeting therefore does make a difference, although it is still unclear whether that difference benefits either the user or the advertiser, since we did not observe improved usefulness. Our focus-group interviews have led us to believe that using activity information to determine when to target, in addition to what to target, may lead to better results. We plan to study this effect in future work.

Also as part of this exploration we have identified why keyword-based advertising cannot be easily adapted for activity-based targeting. Role, timing, specificity, generality, and sequencing make activity-targeted advertising more difficult. A more detailed study of these factors and search for others would also improve effectiveness.

Additionally, we have demonstrated a new way to use Mechanical Turk: as a "time machine" to simulate a technology that is not available today. This might be useful for applications of other developing technologies, such as speech recognition, face recognition, object tracking, or natural language comprehension. Our approach is similar to the "Wizard-of-Oz" technique. However, because there is a delay between the submission of a question and its response, it is less suitable than Wizard-of-Oz for applications in which an immediate system response is required. The application of Mechanical Turk is also limited to common knowledge questions. However, Mechanical Turk is superior for longitudinal and pervasive applications where the "wizardry" is required in any context at any time of day. And by combining results from multiple individuals, Mechanical Turk can be used to achieve "Wisdom of the Crowds" effects. In our system, for example, we easily achieved a wide coverage of potential products related to a given activity.

Another aspect of using Mechanical Turk and other such systems is that, like the implementation of computer programs, human processes require debugging. It is important to keep in mind that the responders' motivation is to receive as much payment for responses as possible, not necessarily to provide the best possible response. As we experienced, what works well initially may lose effectiveness as loopholes are exploited. While direct human oversight could be used in cases where the difficult task is generation and the easy task is validation of the generated response, in large-scale systems this is not feasible. The Mechanical Turk-based oversight system we developed shows promise, but is not foolproof. More robust oversight mechanisms need to be developed.

Finally, our exploration into activity-based advertising exposes a need for another line of research in activity recognition. Much of the activity-recognition research to date has been bottom-up and technology-driven with the goal of discovering which activities are detectable by sensors. The results have been techniques to detect motor-level human actions such as walking, running, biking, climbing stairs, social engagement, task engagement, object use and other actions. Above this level, researchers have discovered applications that can combine detected actions to infer higher-level activities such as activities of daily living and interruptibility. While successful, this bottom-up approach does not necessarily cover the full range of activities at human cognitive levels. Detecting the user's cognitive state is important for the kinds of context-aware applications that aim to match information to what is on the user's mind, such as advertising and information retrieval. We explored a top-down approach, asking users to describe their activity in their own terms, and then generating a set of labels for activities in natural terms. This approach will allow research into activity-recognition techniques to detect the kinds of activities thought of by users themselves.

Acknowledgments. We would like to thank Pratik Rathod for his assistance collecting and analyzing the data, Diane Schiano for her help designing and analyzing the study, Paul Stewart for his help supporting the backend infrastructure, Elizabeth Churchill for her feedback early on, our paper shepherd Hans Gellersen and the anonymous reviewers for their suggestions and guidance, and Rowan Nairn for insightful discussions during many long train rides.

References

1. Lauri Aalto, Nicklas Göthlin, Jani Korhonen, and Timo Ojala. Bluetooth and WAP push based location-aware mobile advertising system. In *MobiSys*, 2004.
2. Jeff Barr and Luis Felipe Cabrera. AI gets a brain. *Queue*, 4(4), 2006.
3. Rebecca Bulander, Michael Decker, Gunther Schiefer, and Bernhard Kölmel. Comparison of different approaches for mobile advertising. In *Second IEEE International Workshop on Mobile Commerce and Services*, 2005.
4. Scott Carter and Jennifer Mankoff. Momento: Early stage prototyping and evaluation for mobile applications. Technical report, University of California, Berkeley, April 2005. CSD-05-1380.
5. Mark Claypool, Phong Le, Makoto Wased, and David Brown. Implicit interest indicators. In *IUI*, 2001.
6. Sunny Consolvo and Miriam Walker. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing*, 02(2), 2003.

7. Jon Froehlich, Mike Y. Chen, Ian E. Smith, and Fred Potter. Voting with your feet: An investigative study of the relationship between place visit behavior and preference. In *Ubicomp*, 2006.
8. John Harms and Douglas Kellner. Towards a critical theory of advertising. http://www.uta.edu/huma/illuminations/kell6.htm.
9. James M. Hudson, Jim Christensen, Wendy A. Kellogg, and Thomas Erickson. "I'd be overwhelmed, but it's just one more thing to do": Availability and interruption in research management. In *CHI*, 2002.
10. Cellfire Inc. http://www.cellfire.com/.
11. Stephen S. Intille, John Rondoni, Charles Kukla, Isabel Ancona, and Ling Bao. A context-aware experience sampling tool. In *CHI*, 2003.
12. Jonathan Lester, Tanzeem Choudhury, and Gaetano Borriello. A practical approach to recognizing physical activities. In *Pervasive Computing*. Springer-Verlag, 2006.
13. Freeset Human Locator. http://www.freeset.ca/locator/.
14. Paul Lukowicz, Jamie A. Ward, Holger Junker, Mathias Stäger, Gerhard Tröster, Amin Atrash, and Thad Starner. Recognizing workshop activity using body worn microphones and accelerometers. In *Pervasive Computing*. Springer-Verlag, 2004.
15. Anant Madabhushi and J. K. Aggarwal. A bayesian approach to human activity recognition. *Second IEEE Workshop on Visual Survelliance*, 1999.
16. Joe Mandese. Online ad spend predicted to reach $20 billion. *Online Media Daily*, June 2006.
17. Nuria Oliver, Eric Horvitz, and Ashutosh Garg. Layered representations for recognizing office activity. In *Fourth IEEE International Conference on Multimodal Interaction*, 2002.
18. P.O.P. ShelfAds. http://www.popbroadcasting.com/main/index.html.
19. Reactrix. http://reactrix.com/.
20. A. Schmidt, M. Beigl, and H-W. Gellersen. There is more to context than location: Environment sensing technologies for adaptive mobile user interfaces. In *Workshop on Interactive Applications of Mobile Computing (IMC)*, 1998.
21. Emmanuel Munguia Tapia, Stephen S. Intille, and Kent Larson. Activity recognition in the home using simple and ubiquitous sensors. In *Pervasive Computing*. Springer-Verlag, 2004.
22. Wessel van Binsbergen. Situation based services: The semantic extension of LBS. In *3rd Twente Student Conference on IT*, 2005.
23. Luis von Ahn. Games with a purpose. *IEEE Computer Magazine*, June 2006.
24. D.H. Wilson and C. Atkeson. Simultaneous tracking and activity recognition (STAR) using many anonymous, binary sensors. In *Pervasive Computing*. Springer-Verlag, 2005.

Evaluating a Wearable Display Jersey for Augmenting Team Sports Awareness

Mitchell Page and Andrew Vande Moere

Key Centre of Design Computing & Cognition
University of Sydney, NSW 2006, Australia
mpag5967@mail.usyd.edu.au, andrew@arch.usyd.edu.au

Abstract. This paper introduces a user-centered design process and case study evaluation of a novel wearable visualization system for team sports, coined *TeamAwear*. TeamAwear consists of three basketball jerseys that are equipped with electroluminescent wires and surfaces. Each jersey can be wirelessly controlled to represent game-related information on the player in real-time, such as the amount of individual fouls, scores and time alerts. A participatory user-centered approach guided the development process towards a more meaningful, ethically and ergonomically valid design. The system aims to enhance the awareness and understanding of game-related public information for all stakeholders, including players, referees, coaches and audience members. We initially hypothesized that such increased awareness would positively influence in-game decisions by players, resulting in a more interesting and enjoyable game play experience for all participants. Instead, the case study evaluation demonstrated TeamAwear's perceived usefulness particularly for non-playing stakeholders, such as the audience, referees and coaches, supporting more accurate coaching assessments, better understanding of in-game situations and increased enjoyment for spectators. The high amount of game-related cognitive load on the players during game-play seems to hinder its influence on in-game decisions.

Keywords: wearable computing, visualization, design, evaluation.

1 Introduction

Wearable visualization focuses on the design of small computers than can be continuously worn on the human body and enable the representation of abstract data, either to the wearer themselves, or to other people in the wearer's vicinity. It differs from more common visual applications on mobile devices in that *wearables* are specifically designed to be unobtrusively integrated within the user's clothing. By merging visualization with fashion, clothing is considered as a sort of public display that is meant to 'signal' an interpretable meaning [1]. By representing specific information, and thus making people 'aware' of aspects that were normally hidden from view, a wearable display can potentially alter the experience of persons present in its immediate vicinity, including that of its wearer. Accordingly, could a wearable display, worn in a sports context, alter the subjective experience of its typical

A. LaMarca et al. (Eds.): PERVASIVE 2007, LNCS 4480, pp. 91 – 108, 2007.

stakeholders, such as the athletes, coaches, referees, and spectators involved? Would athletes change their game-play when they have access to more relevant game-related information in real-time?

TeamAwear (so named for Team Sports Awareness Wearable Display) is a wearable display in the form of an electronically augmented sports jersey. It is capable of representing game-related information sources in real-time via a small micro-electronic board that controls a series of light-emitting display panels. Its intended application setting is within a team sports environment, where it is simultaneously worn by multiple athletes during the game itself. By perceiving the worn displays, it is expected that the different sports stakeholders become more consciously 'aware' of otherwise hidden or fast-changing crucial game-play related information in the periphery of their attention, hereby positively influencing their sports experience. This approach addresses the relative lack of information that sports stakeholders (especially athletes) receive during fast-paced live games, as opposed to remote television spectators.

We hypothesized that an increased awareness of game-related information would allow athletes to make improved in-game decisions, leading to more interesting game-play. Athletes could exploit opportunities to benefit the individually displayed information in a natural way. For instance, poorer players might be inclined to always pass to top scorers, hereby shifting the collaborative team efforts. Opposing players may deliberately target only those with a high degree of fouls. This is a familiar problem in computer supported collaborative work, as highlighted by the 'work versus benefit' paradigm [2]. Although we found that players did not operate in this way, a different evaluation approach involving a greater number of wearable display jerseys, might still uncover aspects of this kind of intentional exploitation.

In addition, we anticipated that the system could support more accurate coaching assessments, lead to a better understanding of game situations and an increased enjoyment of the game for spectators. We have chosen basketball as our initial application domain for its relatively rapid pace, high amount of relevant information, and high reliance on team-based strategies. Although TeamAwear's technical implementation is not as complex as most other (sensor-embedded and thus input-oriented) wearable computing systems, its main contributions instead focus on a participatory development process within a sports context, and its usage of wearable computing as a intuitive output medium of useful and real-time information.

The development, implementation and use of technology in the sports domain inherently encounters several ethical considerations, which depend on their purpose and potential influence on the sport experience itself [3]. For instance, access to technology should be equal, to ensure no team has a perceived advantage. Technological aids that potentially enhance sports performance historically have received a very slow acceptance, or none at all. Accordingly, although the Australian Basketball Association has permitted this research, no real future commitment for allowing awareness-enhancing jerseys has been made. Even so, the TeamAwear system research is significant in its potential use as a training device for both team and individual sports, as an aid for physically disabled or auditory impaired athletic game-play or as awareness-enhancing clothing outside of the sports domain, such as during team-oriented high-demanding collaborative tasks or emergency situations.

2 Related Work

A *wearable display* extends the fields of *visualization* and *wearable computing,* both of which play significant roles in the sports area. Graphical visualization has already been used in a sports context, in general to enhance the spectator experience. For instance, the *TennisViewer* system allows spectators to explore and understand the vast amounts of data occurring during a competitive tennis match [4]. Visualizations can also facilitate more accurate judgments by referees. The *Piero* graphics analysis tool allows athletes and their actions to be virtually tracked and analyzed by referee officials during a rugby match [5]. Only few visualization examples exist that aim to inform the athletes themselves. The *Cyclops Auto Serve Line Detector* uses sound cues to convey judgmental information to athletes during a tennis match [6]. Similarly, *TeamAwear* conveys relevant game information to support the activities and understanding of common team sports stakeholders.

The merging of wearable visualization and fashion design is a recent trend in which clothing becomes electronically enhanced to reveal information about wearers, or their surrounding environment. Several projects have investigated different technical means to convey sensor data publicly through electronic fashion. These approaches tend to experiment beyond the use of simple LCD or pixel-based displays, and have utilized LED lights [7], [8], so-called ‘e-textiles’ [9], electroluminescent wire (e.g. [10]), thermo-chromatic inks, shape-changing materials such as shape memory alloys (e.g. [11]), inflatables (e.g. [12]), and so on. In contrast, the TeamAwear system aims to communicate information in an efficient and immediately understandable way.

Wearable visualization in a sports context can be used to convey information to athletes and other stakeholders in a variety of ways. For instance, wearable devices can be used to provide critical information about athletes to referees. The *SensorHogu* force-sensing vest for martial arts improves the accurate detection of contact between athletes, allowing for more fairly judgments [13]. A number of wearable-sensor based devices aim to provide information to improve coaching, such as pressure sensors for high impact sports [14], and accelerometers for movement-oriented sports such as skiing [15]. Similarly, wearable GPS modules, named *FitSense*, are fitted to professional athletes during football competitions to track and visualize their movements during game-play [16]. It is well documented, that when feedback is provided in an appropriate manner, sport athlete performance tends to increase [17]. A tactile display embedded within a vest, named *TACT*, improves the athlete’s performance in sports such as ice skating, by communicating relevant real-time information to athletes via minor vibrations targeting their upper body [18]. The *Houston* wearable mobile display shares a person’s physical activity data to encourage them to physically exercise outside of the competitive sports context [19]. Athletic wearable devices for the everyday individual have become fashionable and socially acceptable, and form the focus of new commercial ventures, such as the *Running Computer* from *Polar* and *Adidas* [20], and the *Nike + iPod Sport Kit*, consisting of an embedded foot sensor developed by *Nike* and *Apple* [21]. In contrast to most approaches, TeamAwear communicates relevant sports information in a team focused way, and considers all surrounding stakeholders, as opposed to just a coach or referee. The *PingPongPlus* ‘athletic-tangible interface’ digitally augments the inbuilt dynamics of a Ping Pong game by providing feedback in the form of a visual and auditory visualization, which encourages athletes to keep playing while still maintaining the competitive nature of the sports game [22]. Similarly, the TeamAwear jerseys respond to the intrinsic performances of the players and teams, which can ultimately persuade and increase in these performances.

In general, technology in sports must adhere to a few basic guidelines, in particular safety and fairness. Any technology which attempts to generate risk to stakeholders, or provide an unfair advantage to any one team or athlete will not be accepted by the sports community [23]. This aspect has been addressed during development of TeamAwear, which included extensive investigation into user needs and requirements specifically for the team sports context.

3 User-Centered Design

Because of the relative novelty of wearable computing, and wearable visualization for sports in particular, a user-centered design approach was chosen. By incorporating insights and comments from relevant stakeholders, it was expected that the system's usefulness would increase, as the design process would inevitably encounter critical ethical, ergonomic and usability issues that required insight from experienced users. The design process consisted of three consecutive stages: Evaluative Ethnography, User-centered Discussion and User-centered Participatory Design.

3.1 Evaluative Ethnography

Evaluative ethnography is a fieldwork-based approach that aims to detect relevant information in order to establish the 'work-ability' of a proposed design system [24]. The observations can be facilitated with the aid of specific documentation tools, such as written field notes and proxemic diagrams. An evaluative ethnography study was carried out over a period of approximately eight weeks to examine the intended users during their normal 'sporting state', observed during scheduled training sessions and competitive basketball games. As a result, several primary system design tasks were discerned, such as regarding the environmental parameters (e.g. the typical amount of light, noise and distances on a sports court), the cognitive load on human senses, the understanding of each stakeholder's role during game-play and their typical actions. Accordingly, these insights allowed us to make several preliminary design decisions regarding the choice of technical parts, including the wireless communication, power supply and display media materials. This research also revealed the relatively high demands and constraints on the potential design of the wearable device, in the context of its physical design (i.e. relatively high amount of physical contact between players, so that even falling is not uncommon), its efficiency of information display (i.e. rapidly changing game development), perceptual abilities (i.e. players dedicate full cognitive attention to game-play), game-related restrictions (i.e. devices are not allowed to protrude or advantage any team or team member) and environmental parameters (i.e. bright and loud physical setting).

3.2 Participative Design Study

A set of user discussion sessions were organized for which representatives for each of the typical basketball stakeholders (i.e. athlete, coach, referee and spectator) were invited. The goal of these sessions was to identify potential usage scenarios as well as critical design requirements. The initial recruitment occurred via distributed flyers and handouts. We specifically appealed for female basketball players, due to an assumed reduced amount of physical aggression during game-play, and less risk of potentially harming the device or its wearer. The recruitment resulted in a limited but sufficient response due to the relative scarcity of female basketball players, the technological character of the device, and the fact that the subjects did not receive any reward for their significant time investment. The numbers of volunteer users was based on the relative importance of their roles to the intended application: each of the sessions involved at least 6 semi-professional athletes, 1 coach and 1 referee, and 3 spectators.

The initial design discussions were intentionally left open-ended, 'allowing' the concept of a sports-related wearable display to be discovered and developed by the focus group itself. During the sessions, it became clear that there was a genuine interest for real-time information access during game-play. However, no participant wanted to display personal physiological or environmental data (e.g. stress level, heartbeat, amount run) to the outside world, for potentially risking unfair tactics by opposing teams. Instead, it was agreed to display 'publicly available data', related to the actual game itself. Several game-related information sources, normally hidden from view, were identified that were considered as useful during a typical competitive basketball game, listed in Table 1.

Table 1. Game-related information sources selected to be displayed by the jersey

Information	Stakeholder	Rationale
Time Limits	*athlete* *coach* *referee*	Athlete: a reminder to hasten teams actions to take a shot (shot clock), or before end of game Coach: strategic decisions depending on game clock Referee: identify time clock related errors more effectively
Individual Fouls	*athlete* *coach* *spectator*	Defensive Athlete: in-game decisions and playing style, i.e. provoke offensive faults. Offensive Athlete: in-game decisions and playing style, i.e. attempts to provoke defensive faults. Coach: reminder of team members with fouls (for both teams), i.e. determining exchange of players, game strategies to protect players with high amount of fouls, or influence competing strategy. Spectator: increased awareness of influencing factors, 'catch up' when part of the game has been missed
Individual Score	*athlete* *coach* *spectator*	Defensive Athlete: in-game decisions and playing style, i.e. more attention and status to top scorers. Offensive Athlete: affects in-game decisions and playing style, i.e. more passes to actual top scorer. Coach: choice of strategies and exchange of players. Spectator: focus on top versus poorly performing players
Winning versus Losing Team	*spectator* *athlete*	Athlete: increased awareness, especially in a tight competition. Spectator: increased engagement in game

Fig. 1. Group design proposal sketching (left); determining appropriate positioning of wearable computing box (middle); hands-on physical design to determine display locations (right)

Participants were divided into small groups containing a variety of each stakeholder type. Following a typical constructivist studio-based approach, each group creatively designed a set of proposals, which were presented, critiqued and refined by the whole group. These design sessions revealed a general consensus on the relative importance of specific information sources, a wish for continuous unobstructed visibility of the display, and a desire to make the jerseys fashionable and aesthetically appealing. These user participation sessions also determined the ergonomic position of the wearable computing box (see Figure 1), which was chosen to be attached below the lower chest or waist. Female participants preferred the waist location, whereas males preferred the lower chest. The participants also decided on the physical design of the system (also see Figure 1), including the material, color and positioning of the displays themselves.

3.3 Prototype Development

After the participative design sessions, three fully working jersey prototypes were implemented. These allowed for the evaluation of different solutions for attaching the displays, and tested the wearability of the jersey during extensive use. The following game-related information sources were selected to be displayed (see Figure 2).

- **Time Limits.** Consisting of the ‘game clock’ on the upper left chest (shown when only 1 minute of game-play remains for a particular round), and the ‘shot clock’ on the upper right chest (shown when only 10 seconds remain for an athlete to make a goal attempt). This information was considered important by the participants as it acts as an urgent warning signal to take action, which is often neglected or inefficiently communicated verbally. For this reason, the displays were placed on the upper chest for maximum visibility towards own team members**.** The displays have contrasting colors (i.e. violet and yellow) so they are easily distinguishable.
- **Fouls.** The foul displays operate symmetrically on the shoulders from the inside out, according to the accumulation of fouls by the respective player. As the maximum amount of fouls is five, only four wires are shown. The foul displays were placed on the shoulder area, so the wearer themselves could more easily refer to and even peripherally detect their own number of fouls.
- **Individual Points.** Embedded symmetrically on each side of the jersey, are three panels representing the player’s individually scored points. These displays are

illuminated sequentially upwards as more points are scored over the entire period of game-play, in gradations of 10 points. In this way, all participants can perceive who the game's top scorers are. These displays were placed along the sides to maximize visibility, especially during passing/shooting scenarios at which an athlete's arms are often raised.

- **Winning/Losing.** A simple on/off display is embedded in the back of the jersey to indicate which team is winning ('on' for winning, 'off' for losing). This display is visible at quite far distances from the court, such as for spectators who might have missed parts of the match. The athletes themselves thought that it would be useful to be reminded of who was winning when moving between goal areas, a time when often only the athletes' backs are visible.

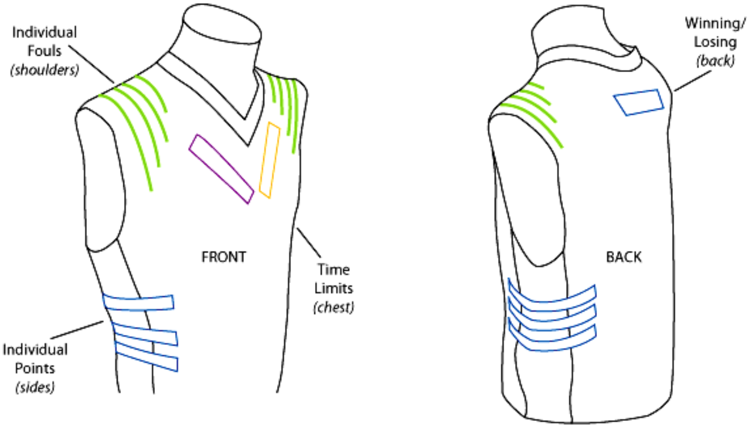


Fig. 2. Diagram showing the exact display locations and their specified meanings

One should note that the display layout of the jerseys shown in Figure 2 is a product of a creative design process, and depends on its intended application. It has been tailored specifically for use in basketball play, where display elements are positioned in visible areas not restricting the traditional basketball jersey format. While its intentions would remain the same, the display layout (and information content) of such a jersey within other team-based sports would unquestionably differ.

3.4 Technical Development

The technical development was mostly determined by safety and usability considerations. Safety is of the utmost importance for any sports-related activity, and the most important reason why technology is not readily accepted in a sports context. Typical safety risks include physical contact with electricity-conducting elements, accidentally hitting or being caught in protruding parts, or falling on top of the wearable device. Inevitable technical failure was considered by deliberately including 'weak spots' that allow quick and safe detachment instead of uncontrollable breakage.

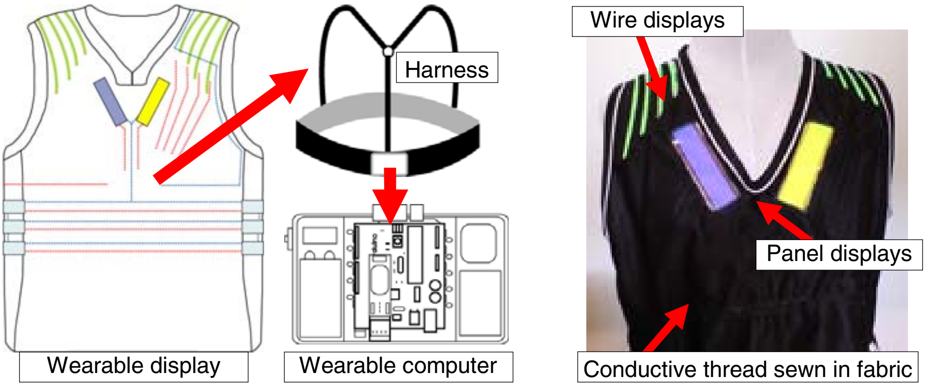


Fig. 3. Diagram of system parts including wearable display jersey (dotted blue and red lines represent conductive thread circuit), harness, and wearable computer (left); Final prototype TeamAwear jersey (right)

In addition, the jerseys were designed so that they do not disturb the normal physical activity of the players. Adverse influences range from simple long-term tiredness due to the extra weight, to subtle alterations in physical movements potentially affecting the player's overall sports performance. Lastly, due to the heavy perspiration during sports, all parts were protected against moisture and allowed for easy washing.

The TeamAwear system consists of three separate parts which conveniently link together during use, as illustrated in Figure 3:

- **Wearable Display Jersey.** Each jersey is equipped with nine detachable light illuminating display panels (5 x 1 inch) and eight wires (1/8 inch diameter). Both displays are made from *electroluminescent* (EL) material, containing a phosphor-based substance that emits light in response to a high-frequency AC electrical signal. The panels and wires are flexible, allowing for an unobtrusive connection to textile fabric. All displays are sealed within transparent vinyl pockets sewn beneath the jersey surface, so they can be easily removed if the jersey needs to be washed. Each jersey contains an additional fabric mesh layer sewn inside, adding an extra protective layer between the electronic components and the wearer's body.
- **Conductive Thread.** An electrical circuit is sewn into the jersey fabric with *conductive thread*, a type of silver-coated synthetic fiber able to pass current with little resistance (<0.2 Ohms). The thread is used to sew a precise pattern that creates a complete circuit onto the jersey itself, as it acts as a power conduit to each display. The jersey circuit includes both positive and negative (ground) threads. Each conductive thread is sealed using washable fabric gel to avoid direct contact and short-circuiting. Conductive press studs allow the jersey to be easily (dis-) connected with the EL displays so the jersey can be washed (Figure 4, right).
- **Wearable Computer.** A micro-electronic board controls each of the displays based on the instructions it receives via a wireless connection with a central computer. The wireless communication is accomplished via a Class 1 Bluetooth serial module, which was selected for its fast communication speed and far

physical range, good reliability, and easy hardware availability. However, this choice especially suffered in terms of power supply demands, as Bluetooth was originally developed for high-bandwidth communication streams. The system is capable of transmitting data from a central location to up to 10 different jerseys simultaneously, which can receive this data up to a distance of over 94 feet, the standard length of a basketball court (see Figure 5). The electronic board is based on *Arduino* [25], a community-driven physical computing platform that aims to reduce the technical complexity for the enthusiastic interaction designer. The setup is powered by a single standard 9 Volt battery. The computing components are embedded in a non-breakable plastic box measuring about 4.7x3x1 inch, shown in Figure 4 (left). The box features a safety switch to disconnect the electrical power in case of emergency. The resulting setup weighs approximately 8.8oz, similar to an average mobile phone.

- **Harness.** Each system includes a unisex harness to which the wearable computer is attached, worn underneath the jersey (Figure 4, left). It assures that the device is fixated even during brisk movements, and can cause minimal physical harm to the wearer's body after unintentional contact. It has a Y-shape front for female athletes and can be worn in the reverse for male athletes. It is made from an elastic ribbon, and can be adjusted for any body size or preferred wearing position.

Fig. 4. The wearable computer box attached to athlete's body by harness under the jersey (left); wearable display jersey during use (middle); attachment of displays beneath jersey surface using conductive thread and conductive quick-snap studs for easy detachment (right)

The application software is based on Processing [26], a community-driven Java-based platform that can easily communicate to Arduino boards. The application sends real-time instructions from a laptop computer to each of the jerseys via a standard Class 1 'cable-replacement' Bluetooth USB adapter (range: 300 feet). The interface can be used by a game official at the bench, and is shown in Figure 5 (right).

4 Evaluation

The case study evaluation took place in a normal basketball hall on our university campus, within the context of a socially competitive match. Plays were limited to a half court, to ensure all stakeholders were in close viewable distance from one another. All participants were informed beforehand as to the meaning of each display on the jersey. The evaluation consisted of two stages: an *awareness* investigation involving pairs of game-like scenarios, with and without wearing the TeamAwear system; and a *game-pla*y simulation during which the TeamAwear system was worn by athletes during a socially competitive basketball game, followed by questionnaires and interviews.

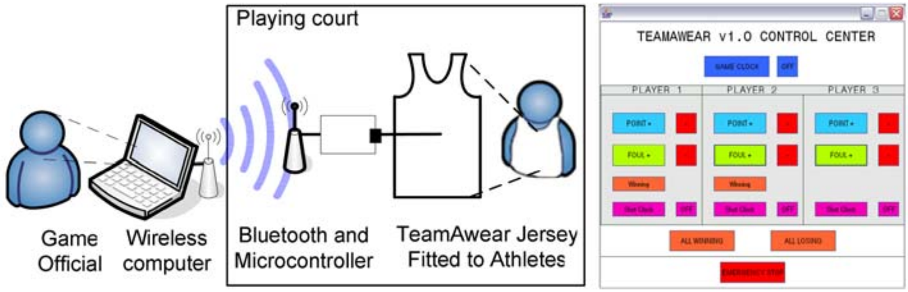


Fig. 5. When data is registered during game-play, instructions are sent in real-time to all connected TeamAwear jerseys via Bluetooth (left); Control interface running on a laptop computer during game-play maintains wireless connections with all TeamAwear jerseys (right)

4.1 Participants

The case study evaluation focused on all team sports stakeholders, including athletes, coaches, referees and spectators. The evaluation participants were recruited using flyers and handouts, which were distributed in sports halls and handed out after basketball games. Eleven participants took part, including: five athletes (female: four; age range 19-25), one referee (age 42), one coach (age 23), and four spectators (age range 16 to 55). All athletes had amateur to semi-professional basketball experience, and spectators indicated beginner to amateur basketball knowledge. One should note that as the subjects were aware of the system's purpose beforehand from the content on recruitment flyers and received no incentive for their participation, their attitude might be positively biased towards the use of technology in sports. The athlete category in particular initially seemed very enthusiastic, some stating they considered it to be completely acceptable *"for people of their generation to … be fitted with small computers*". Before the evaluation, players were allowed to wear the jerseys for 5 minutes during training to become accustomed and learn its visual use.

4.2 Awareness Evaluation

This evaluation was designed to test the ability of the jerseys to increase the stakeholders' awareness of game-related information sources. Players followed four different game-play scenarios, first without, and then while wearing the TeamAwear jerseys. Each scenario simulated a change in one of the game-related information sources represented by the displays. For instance, a scenario could consist of assigning fouls to a random number of athletes, or observing their individual scores. At the conclusion of each scenario, a survey queried each stakeholder about their in-game awareness in regard to the information source changed. In the cases which used the TeamAwear system, all displays would be completely switched off immediately after the play so users could not see them at the time of answering.

The results of the surveys showed a slight increase in awareness for the players, of which approximately average 4 out of 5 could correctly recall game-play data with jerseys, average 3 out of 5 without. For the non-playing stakeholders viewing the game from the court side, the difference was more significant (average 4 out of 4 versus 2 out of 4). This result seems to suggest a difference in use between the in-game experience and non-player experience of the display. Due to the fast-paced nature of a typical basketball game-play, with its need to focus on the ball, competitors and teammates, a player can only rely on quick glances. Therefore, the error rate was slightly more pronounced (1 out of 5 versus 2 out of 5) with the wire displays on the shoulders, due to a lower visual perceptibility. No errors were recorded in regard to the winning versus losing team scenario (all participants responded correctly), as players have an already intrinsic strong interest in this information. Non-playing stakeholders tended to have less cognitive effort to observe the jerseys, becoming more aware of the information communicated on the displays.

4.3 Game-Play Evaluation

A 15 minute social competitive basketball game was carried out and recorded on video, with two players in each team and one team continuously wearing the jerseys. Small teams naturally limit any 'strategic' influences, but are considered sufficient for the goals of this prototype case study, which focused on detecting the fundamental usefulness and usability of this highly technological intervention. The game was followed by a referee, a coach and an audience. An *experience questionnaire* was issued directly after the game, querying issues of wearability, visibility, and user experience via both binary and open-end questions.

- **Wearability.** The issue of wearability was only applied to the athletes, who felt that the jersey and wearable computer were comfortable to wear and did not restrict their normal physical movements during game-play ("*I actually forgot I was wearing it*"). The flexible and user-adaptable harness assured that the ergonomic influence of the system was minimized. A small number of players (2 out of 5) were bothered by the heat emission from the wearable computer box, and noted that the jersey made them feel slightly hotter than usual. However, some condensation did build up underneath the plastic surfaces protecting the displays, which not only limits the natural dissipation of sweat, but ultimately might become a security risk in terms of short-circuiting the electrical parts on a long-term basis.
- **Visual Display.** All participants reported that they could clearly perceive the displays during use, and were able to understand and remember what each display represented. Several responses referred to display arrangement as "*intuitive*" and "*easy to understand*". All participants agreed that the displays were not overwhelming or distracting, even when multiple displays were activated. All players agreed that the panel displays were significantly easier to perceive than the glowing wires, which were sometimes confusing as the difference between multiple active wires was not easily discernable. The athletes revealed that they

were not able to perceive their own displays, but instead often relied on other jerseys for general game-related information. Players sometimes deliberately stretched the shoulder part of the jersey in front of their face to double-check its actual state, an undesirable action that is possibly related to self-assurance and not trusting the technology: "*I wanted to be sure my fouls were represented correctly*".
- **Game Experience.** In terms of the overall game experience, it was found that the athletes might benefit the least from the TeamAwear system. Spectators claimed the game was "*more interesting to watch*" and "*easier to follow*" as a result of the wearable displays. In particular, the spectators with only a rudimentary knowledge of basketball indicated they were able to "*understand the play better*" and that this "*made the experience richer*". All participants unanimously agreed that there exists a definite use for this technology in the future.

Because of the fast-paced nature of team sports, it is often difficult for players to remember their mental considerations for specific game-play related actions. Therefore, the participants were shown a video recording of their game-play on a television screen, and then *retrospectively interviewed* as a group. This approach aimed to allow the participants, specifically the athletes, to perceive their actions from a third-person viewpoint and 'recall' eventual in-game reasoning, and possible influences caused by the TeamAwear system, similar to a *post think-aloud* session.

- **Influencing Decisions.** The feedback provided by athletes during the video playback reflected only subtle changes in some aspects of their playing style. Most players described that during the majority of the game, the jerseys did not influence their in-game decisions. One athlete mentioned that when she saw her team mate had a high points score, "*it made me want to pass to her more, as she had a 'hot hand*". The players indicated that they mainly glanced to the jerseys during time periods when the pace of game-play naturally slowed or stopped, such as after a goal was scored or a foul was called, "*whenever the ball was reset, I would take the time to look at the displays before starting the play again*". The athletes reported that the time display had the most influence: "*I played harder [more aggressively] when I saw the game had little time left*", and "*I noticed the game-clock, and I yelled this to my team mates*". Some players indicated that even when they perceived some useful information, such as the foul amount of their competitors, such knowledge inherently would not change their game-play anyway. These results suggest that the intensity of the game itself supersedes any consideration of perceiving and interpreting the display: the players were not able to pay enough attention to the displays as they were too preoccupied with the game situation itself, and even when they perceived the displays during calmer moments, they did not consider (or forgot) this knowledge when making in-game decisions.
- **Confidence.** An awareness of one's own and the team's performance during game-play seemed to boost a personal feeling of self-belief. When one player noticed her particularly high score indicated by the displays on the side, "*being reminded of my high score gave … a general feeling of confidence*". Also, athletes felt more confident in their own actions during game-play, as one athlete stated that "*I felt more comfortable passing to a team member I could see had scored more points than myself*". This feeling relieved some pressure placed on

athletes during game-play, with one athlete exclaiming that when she could see her team was winning "*it made me feel more relaxed*". Although the jerseys were originally intended to enhance the athlete experience indirectly through improved decisions and actions, it seemed they especially boosted self-assurance and morale, as scores are represented on an individual level, not on an anonymous scoreboard.

- **Non-Player Comments.** Although initially not intended, spectators seemed to benefit the most. The interview responses indicated they were able to follow the game more closely and accurately. One spectator who was inexperienced with the rules and plays of basketball said "*I didn't have to think a great deal to know what was going on*". Another novice spectator indicated that the jerseys "*did enhance the experience for me*" as she could see and understand when a player received a foul. This opinion was also reflected by other novice spectators, one saying the jerseys "*gave me the basic information I otherwise wouldn't have known*". Overall this contributed to a more enjoyable experience for spectators who were able to better understand game-play, even finding it *"more interesting than normal"*. The coach indicated she had a better overview over the game flow, as *"you can keep track of your players and more easily make decisions of which players to keep on and which to take off"*, "*you can tell your players which opponents to 'double-up on' based on observing their fouls or points*". This knowledge can normally only be accessed by a dedicated person scouting the players, or by verbally requesting information from the official bench. The referee indicated "*it could really speed up the game from the referee's perspective, having to consult the bench less*", as much of the useful information is literally represented 'on' the player themselves.

Fig. 6. User evaluation study during a social competitive match. From Left to Right: high scoring athlete (blue horizontal strips) clearly identifiable by all stakeholders, a high foul athlete (green shoulder wire displays), a high and low scoring athlete in close proximity, and the wireless laptop computer steering the wearable display jerseys (foreground).

4.4 Key Design Requirements for Wearable Visualization Devices in Sports

Based on the experiences and outcomes encountered during the development of the TeamAwear system, a number of design requirements became apparent.

Ethical Considerations

This study revealed several ethical considerations for a wearable visualization application in the sports domain. Firstly, there is the aspect of potential injury of the wearer: any additional object worn by a player might directly or indirectly inflict personal harm, or cause a deficiency in sporting performance. Naturally, one needs to consider whether the advantage of being more aware of information weights up against wearing a less comfortable jersey or an extra weight on the chest. For instance, the Human Ethics Committee had difficulty to assess this project's true risk due to its perceived technological novelty and its high-risk 'competitive' sports application domain. Secondly, one needs to consider the potential change in competitiveness when only a portion of the players allow their information to be exposed to the competing team. From a moral perspective, either all or no players should wear a potentially performance deteriorating device. Wearing externally controlled information on one's body necessarily results in giving up the right for self-expression: the wearer has no direct control to what information is shown, in what context and during what time. More particularly for our display, the top players, but also the least productive ones, become rapidly visible. This may be seen as a clear benefit for non-expert players who could use the display to make decisions for improving their game-play, yet may also represent a disadvantage as expert players might already make exactly the same decisions without the aid of such display. This immediate visual disclosure to external critique might not be desirable, depending on the wearer's self-confidence level. This aspect might also explain why no participant agreed on displaying individual physiological data, such as heartbeat, stress level, tiredness or energy usage. Participants were quick to point out early in the design stage that any display of such personal data would be damaging to the individual athlete's and team's performance. It could easily result in players losing motivation, as well as opposing teams taking advantage of one's obvious weak points. However, this issue was approved by spectators, who agreed this would contribute to a more interesting and intriguing spectator experience. This raises the question whether the athletes should wear such revealing display solely for the sake of the spectators? One should also note that some typical game-related aspects considered of high value in a basketball game are not explicitly shown on the jerseys, such as shot accuracy, defense actions, steals, rebounds or assist passes. Consequently, by highlighting only a subset of information, the wrong emphasis might be created towards players, coaches and spectators, to what a good 'basketball performance' really is about. Lastly, there is the issue of inherently 'relying' on technology, especially in this application where there was no technology required before. For instance, what happens if the system breaks, causing players, coaches or referees to rely their actions on wrong information: should the match be halted, the end results altered, or should the jerseys only be treated as non-essential and non-trustworthy gadgets?

Wearability

The inherent characteristics of the sports domain highly restrict technical design possibilities. From miniaturization of all computing elements to the washability of clothing, each design decision is primarily determined by strict wearability and usability requirements.

Perceptibility

The evaluation study made apparent that the display type itself plays a significant role. Athletes revealed that they often made decisions based only on their peripheral vision. For instance, players can immediately decide to pass a ball to others based only on seeing a glance of clothing color, posture, body height or hair color. Therefore, displays with a large surface area are favored, as they are more easily distinguishable from any distance. There also needs to be a substantial contrast between the “visibility of states” that the display embodies [27]. For instance, it was found that the EL-wire displays could be quite deceiving, as they glow bright green when turned on, but still appeared greenish when turned off. In contrast, the panels could be discerned more easily due to their absolute color change (white to bright blue). It is important to keep displays subtle: flashing or fading displays can be confusing or show false information when in an ‘intermediate’ state. Spacing between displays is crucial, as it was found that displays placed close together were perceived as merged. To assure that no faulty information is provided, the state of the whole system should be visible at all times: for instance, is a display turned ‘off’ because of the actual data, or because it is broken?

Intuitiveness

Unlike ambient displays and electronic fashion pieces, which are intended to be learnt over time, wearable visualization should be discernable in as little time as possible. Although all participants of this study were briefly informed about the meaning of each display, it was found they did not need to ‘think’ about what the displays were showing. For instance, the wire displays representing the fouls were intentionally placed on the shoulder to ‘warn’ players when they had many fouls, similar to the aggressive raising of shoulder blades. Similarly, the displays representing athletes’ individual points were arranged in an upward fashion, reminiscent of a music equalizer or data visualization bar chart. An intuitive layout design requires input from end-users, who may be familiar with different representations for concepts such as ‘big’, ‘small’, ‘good’, ‘bad’, or ‘warning’. In this study, the participants seemed to be inspired by alternative sport visualizations, such as on-screen graphics shown during sports broadcasts, and graphical visualizations on video game avatars.

Design Process

This project has demonstrated the advantage of user participation, which provided insights that would be impossible to gather in a different way. User contributions were particularly useful in the system’s physical and ergonomic design, which performed successfully during the game-play evaluation. It was also helpful in identifying the most important information sources, and their most suitable positioning on the jersey. The largest problem encountered, however, was recruiting enough suitable and committed people for a technological application that was virtually unknown.

5 Conclusion and Future Work

This paper introduced the design and case study evaluation of a novel wearable display system for team sports, that augments the mutual awareness of game-related information sources in real-time. Although its technological implementation is relatively simple compared to other sensor-based wearable systems, its design was constrained by the complex demands of the high-risk sports application context. Although its technological design proved to be successful, several unforeseen conceptual insights were discovered. The influence of the displays on the players in-game decisions, as originally hypothesized, showed to be very limited due to the fast-pace and highly demanding cognitive load of the game itself. As the information was shown onto the players themselves, the displays rather seemed to influence their self-confidence level. In contrast, referees, coaches and spectators seem to benefit the most of an increased game awareness: to make decisions more effectively and efficiently, to understand the actual game dynamics more profoundly, and to make the game more enjoyable to watch. Therefore, we propose wearable visualization as an effective awareness-enhancing device for non-players, which is conceptually similar to superimposed, explanatory infographics during television broadcasts or webcasts.

We consider our case study evaluation as successful for further investigation in the area of wearable visualization in the sports domain. Even when a similar system would not be accepted by the relevant national sports federations for competitive use, the system can still be used for training, disability sports, collaboration or social competition purposes. For instance, the jerseys can be worn to reflect the player's performance during training, can visually indicate strategic 'plays' (e.g. which player has to move when and where), or communicate strategies from the coach to the players (locally or remotely). A similar system could convey strategic information or increase situational awareness for emergency services during crisis operations. All participants indicated to understand the system's usefulness and even agreed to use it in the future when it would become more user-friendly. Therefore, future research could include fine-tuning the design of the jerseys, evaluating the system during a match with 'real' parameters (i.e. five against five, full time), and the addition of more complex information sources that cannot be easily discerned or remembered by the stakeholders (e.g. shot accuracy and assists, or aggregated data such as defensive and offensive statistics).

The apparent success of TeamAwear among the spectators of a basketball game could be seen as desirable outcome for the highly visual medium of television, inspiring further development as a media application to promote audience interest and enjoyment. Placed on the players themselves, such jerseys might be a more natural solution that also appeals to the players, and not only the audience. It might also overcome the typical performance problems of infographical computer graphics overlays in fast-paced sports.

Acknowledgements

We would like to thank all participants for their time and effort, and our family, friends, and fellow staff for their ongoing assistance during development.

References

1. Liu, C.M., Donath, J.S.: Urbanhermes: social signaling with electronic fashion. Proceedings of the SIGCHI conference on Human Factors in computing systems (2006) 885-888
2. Grudin, J.: Groupware and social dynamics: eight challenges for developers. Communications of the ACM **37** (1994) 92-105
3. Loland, S.: Technology in sport: Three ideal-typical views and their implications. European Journal of Sport Science **2** (2002) 1 - 11
4. Jin, L., Banks, D.C.: Visualizing a tennis match. Proceedings of the 1996 IEEE Symposium on Information Visualization (INFOVIS '96). IEEE Computer Society (1996)
5. BBC News: Piero gives rugby new perspective. Accessed 2 October
6. Hibbert, L.: Decisions you can't argue with. Professional Engineering **12** (1999) 26-27
7. Berzowska, J.: Memory Rich Clothing: Second Skins that Communicate Physical Memory. Proceedings of the 5th conference on Creativity & Cognition, ACM Press, New York (2005) (2005)
8. Iso-Ketola, P., Karinsalo, T., Myry, M., Hahto, L., Karhu, H., Malmivaara, M., Vanhala, J.: A Mobile Device as User Interface for Wearable Applications. The 3rd International Conference on Pervasive Computing (2005) 5-9
9. Martin, T., Jones, M., Edmison, J., Shenoy, R.: Towards a design framework for wearable electronic textiles. Wearable Computers, 2003. Proceedings. Seventh IEEE International Symposium on (2003) 190-199
10. Rosella, F., Genz, R.: Cute Circuit Kinetic Dress. SIGGRAPH Cyber Fashion Show (2005)
11. Berzowska, J., Coelho, M.: Kukkia and Vilkas: Kinetic Electronic Garments. Wearable Computers, 2005. Proceedings. Ninth IEEE International Symposium on (2005) 82-85
12. Almeida, T.: Modes for Urban Moods. Interactive Telecommunications Program Thesis. NYU, New York (2005)
13. Chi, E.H.: Introducing wearable force sensors in martial arts. IEEE Pervasive Computing **4** (2005) 47- 53
14. Sherriff, R., Sherriff, L., Forscutt, O., Ho, S., Mann, S.: Smart-top: Athlete Monitoring for High Impact Sport. 14th Annual Conference of the National Advisory Committee on Computing Qualifications, Napier, NACCQ **457** (2001)
15. Michahelles, F., Schiele, B.: Sensing and Monitoring Professional Skiers. IEEE Pervasive Computing **4** (2005) 40-46
16. Wisbey, B., Montgomery, P.: Quantifying AFL Player Game Demands Using GPS Tracking. FitSense Australia (2005)
17. Liebermann, D.G., KATZ, L., HUGHES, M.D., BARTLETT, R.M., McClements, J., FRANKS, I.M.: Advances in the application of information technology to sport performance. Journal of Sports Sciences **20** (2002) 755-769
18. Vannieuwland, R.: Tactile technology: the skin as untapped sense. TNO Magazine (2005) 20-21
19. Consolvo, S., Everitt, K., Smith, I., Landay, J.A.: Design requirements for technologies that encourage physical activity. Proceedings of the SIGCHI conference on Human Factors in computing systems (2006) 457-466
20. Adidas-Polar: http://www.adidas-polar.com/. Accessed January 28 2007
21. Nike + iPod: http://www.apple.com/ipod/nike/. Accessed January 28 2007

22. Ishii, H., Wisneski, C., Orbanes, J., Chun, B., Paradiso, J.: PingPongPlus: design of an athletic-tangible interface for computer-supported cooperative play. Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit (1999) 394-401
23. Busch, A.: Design for Sports: The Cult of Performance. Princeton Architectural Press. (1998)
24. Hughes, J., King, V., Rodden, T., Andersen, H.: The role of ethnography in interactive systems design. interactions **2** (1995) 56-65
25. Arduino: What is Arduino? http://www.arduino.cc/. Accessed 12 April 2006
26. Fry, B., Reas, C.: Proccessing.org. http://www.processing.org/. Accessed 12 October
27. Mankoff, J., Dey, A.K., Hsieh, G., Kientz, J., Lederer, S., Ames, M.: Heuristic evaluation of ambient displays. Proceedings of the SIGCHI conference on Human factors in computing systems (2003) 169-176

Using Location, Bearing and Motion Data to Filter Video and System Logs

Alistair Morrison, Paul Tennent, John Williamson, and Matthew Chalmers

Department of Computing Science, University of Glasgow, Glasgow, UK
`{morrisaj, pt, jhw, matthew}@dcs.gla.ac.uk`

Abstract. In evaluating and analysing a pervasive computing system, it is common to log system use and to create video recordings of users. A lot of data will often be generated, representing potentially long periods of user activity. We present a procedure to identify sections of such data that are salient given the current context of analysis; for example analysing the activity of a particular person among many trial participants recorded by multiple cameras. By augmenting the cameras used to capture a mobile experiment, we are able to establish both a location and heading for each camera, and thus model the field of view for each camera over time. Locations of trial participants are also recorded and compared against camera views, to determine which periods of user activity are likely to have been recorded in detail. Additionally the stability of a camera can be tracked and video can be subsequently filtered to exclude footage of unacceptable quality. These techniques are implemented in an extension to *Replayer*: a software toolkit for use in the development cycle of mobile applications. A report of initial testing is given, whereby the technique's use is demonstrated on a representative mobile application.

Keywords: Video, auto-classification, analysis toolkit, log synchronisation, visualisation.

1 Introduction

Evaluating and analysing a mobile application, especially one involving multiple participants and locations, can be a difficult task. Multiple observers may be needed, often recording multiple streams of video that complement system logs from the application running on multiple devices. The sheer volume of recorded information can make detailed analysis a time-consuming and labour-intensive task. This can be the case where an analyst employs an exploratory approach, for example using ethnographic techniques to reveal patterns of use, or when the analyst carries out a hypothesis-driven experiment. Much of the data may not be relevant or complete, and so examining all the information looking for periods of interest can consume much of an analyst's time; considering the specific case of video data, an analyst may be required to watch tens or even hundreds of hours of footage in search of evidence. Here we present a new technique designed specifically to aid in this pursuit. The technique is implemented as part of Replayer [12]: an evaluation tool for the

A. LaMarca et al. (Eds.): PERVASIVE 2007, LNCS 4480, pp. 109 – 126, 2007.

combined analysis of video data and recorded system logs. Replayer already temporally synchronises quantitative log data with mixed media recordings, so that recorded system and interaction events in an experiment can be analysed within the same tool. Analysts can therefore pinpoint specific events in the timeline of an experiment, and jump to the periods of video showing the timeframes at which these events occurred. However, there is no guarantee that all or any of the video streams captured at these instants will have captured the event of interest, or if they have, whether the quality of the video will be acceptable. This is a particular issue in the evaluation of multi-user mobile applications, where a roaming camera will likely struggle to capture all the participants' movements.

The technique presented here augments video recordings with the location and heading of each camera as well as data about the cameras' motion. With this information, Replayer is able to inform an analyst on which events are likely to have been captured, and automatically tailor video playback to show only these periods. A further application of this technique is in identifying all the periods of video footage that capture a particular person, as he or she moves in and out of the visual fields of multiple cameras. Finally, this technique allows the system to automatically discard all the video in which the camera is shaking excessively – a common problem when recording mobile systems as roaming camera operators are often forced to run. In this paper we explain the implementation of these new techniques, and use a mobile application to demonstrate how effective these facilities can be.

The following section provides an overview of Replayer: the software toolkit to which the described techniques are an extension. This is followed by a description of related work in the field. The four main benefits offered by the ability to classify data in this manner are outlined in Section 4, before the technical details of the process are described in Section 5. An experimental trial is described in Section 6 and the results are analysed. This is followed by a description of future directions for this work and finally our conclusions.

2 Overview of Replayer

The Replayer toolkit [12] has been developed to support the evaluation and development cycle of mobile computing systems. It can be used in usability testing or by computing or social scientists in studies into the use of mobile applications. As in the simple case shown in Figure 1, logged system data as well as video and audio recordings can be examined in a synchronised manner, along with textual notes recorded either during a system trial or post-hoc. In the example of the figure, selecting an area of a graph of accelerometer readings also selects video corresponding to the times of those readings.

The work of qualitative methods of video analysis, such as ethnography, is time-consuming but affords rich detail of the user experience of systems – detail that may be unavailable from quantitative data such as system logs. In a complementary way, quantitative methods allow for rapid indexing and for overviews such as statistical distributions and visualisations, but may abstract away from the subjective experience of users. The analytic practices associated with these complementary approaches are

Fig. 1. A simple case of analysing heterogeneous data in Replayer. The analyst has selected a period of time (colouring the selection blue) in graphed accelerometer logs (top left), which triggers Replayer to highlight in green the corresponding section of the video timeline (bottom) and to cue the video of that section (top right). (Colours mentioned here and in other parts of the paper refer to images available in the digital version of this paper).

often carried out separately, or even in opposition to each other, but Replayer is intended to allow for tighter coupling and integration of these different forms of data and these different forms of analysis.

By mixing quantitative and qualitative analysis techniques, Replayer is a powerful tool for examining data recorded about a system and its use, providing many different techniques for synchronising, visualising and understanding the data. The example of Figure 1 shows how interactions performed in one component are reflected in another. In fact, each visualisation component in Replayer is linked to every other in this way to support brushing [4]; any selection made in one immediately makes a corresponding selection in another. For example, we may have a graph showing all the system events for a given participant on a timeline, and a map showing a spatial distribution of those events. Selecting one event on the timeline would highlight the location on the map at which the event occurred. As shown above, this is also applied to video data – selecting the event on the timeline also shows any video captured at that time by each camera, jumping to the appropriate part of each recording. Similarly, in the example in Figure 1, selecting a section of the video timeline (bottom of the figure) would highlight the accelerometer data logged during the time period of the section.

An early prototype demonstrating the ideas of the project was described in [14]. The full Replayer system was introduced in [12], which provides further details on both the system capabilities and implementation that are not covered in this summary. The techniques described in this paper extend the synergistic combination of quantitative and qualitative analysis that Replayer is based on. Faced with large volumes of video data, we have chosen to use quantitative data about location and motion to assist existing qualitative analysis techniques, such as ethnography, rather than replace them. In this paper, we outline techniques that use location and motion data to focus on particular sections of video that might benefit from detailed qualitative analysis, and to allow those doing such analysis to better relate their usual material – video – to visualisations of quantitative data that are made more usable by their tight coupling in interaction. Findings and results from our user trials of Replayer, intended to help people carrying out their own user trials in this hybrid or synergistic way, are the subject of forthcoming work. In this paper we focus on techniques to allow classification and filtering of media data: a novel extension to Replayer that itself shows useful synergy in approaches.

The following discussions involve a number of different roles in using and evaluating applications, and it is worth clarifying vocabulary at this stage. Replayer is a desktop tool for data analysis. It is intended that Replayer be used by *analysts* looking into the results of user trials of mobile applications. *Participants* in these trials will have their activity logged by code within the mobile application and be filmed on video by *camera operators* (often part of the analysis team).

3 Related Work

Replayer grew out of work on evaluating multi–user ubiquitous computing systems, in particular mobile multiplayer games. More traditional tools for analysis tend to focus on parts of the analysis task, such as Transana (www.transana.org) for transcribing audio and video. Others are limited in the tools for analysis and visualisation that they support. For example, The Observer (www.noldus.com) allows synchronised playback of up to four video streams alongside sensor data (e.g. physiological data), but relies on simple graphs and tables of quantitative data. Replayer allows more complex interactive filtering and selection of data, and also integrates statistical tools such as mutual information, and sophisticated visualisation tools such as force–based layouts of multidimensional data. DRS [9] is beginning to move beyond systems such as The Observer in its support for complex categorisation schemes and shared repositories of both raw data and analytic results, and also supports a timeline-based tool to display video, but its tools for interaction and visualisation are currently very limited. An early system synchronising video was developed by Badre et al. [2] which used a video tape/CD based system and made use of captured event streams to synchronise time-stamped events with the time-code on a video.

Our system takes advantage of the trend towards positioning systems such as GPS being integrated with devices such as mobile phones and cameras. Previous work with GPS–enabled cameras includes RealityFlythrough by McCurdy et al. [11], which placed images and video streams from camera-enabled mobile phones in 3D space to create an amalgamated panoramic scene. Beeharee and Steed [5] have used occlusion information to filter dynamically generated content provided based on users' locations, removing that which cannot be seen due to visual occlusion. Using hardware similar to that described in this paper, Sawahata et al [13] describe a video capture system augmented with an inertial sensing platform and a GPS unit. The time-series from the accelerometers and gyroscopes is classified with a hidden

Markov model; the system can distinguish walking, running and standing behaviour. The video stream is indexed with these activity levels, and the physical location at which they occurred. Conversely, Aiwaza et. al. [1] use a combination of shot and pan detection (from the video stream itself) and sensed brain data to segment video streams from wearable cameras to produce summaries of activity. Brain activity in the alpha and beta bands is used to estimate attention, so that regions where the user actively paid attention can be extracted from the video stream.

Employing a completely different form of hardware and one that limits the system to use in pre-prepared areas, de Silva et. al. [8] use pressure sensitive floor tiles to sense the location of people within a building. A number of video cameras are present within the test location. The floor tile data is used to produce a video which automatically hands-over between cameras as the user goes in and out of shot. The hand-over algorithm minimises camera switches while maintaining good coverage. In a commercial rather than research area, GeoVector (www.geovector.com) have produced a number of applications which dynamically deliver content to mobile devices based on a GPS reference and a heading from an electronic compass.

4 Benefits of Classifying Video

When performing an evaluation of a mobile system, video capture becomes a challenge. In traditional lab-based experiments, one or two cameras would generally be able to record everything, and typically these cameras could be affixed to tripods and subsequently ignored while the experiment was captured. In a mobile experiment, such as those we have run to evaluate Treasure [3] and Feeding Yoshi [6], it is common to use both fixed and roaming cameras. With a fixed camera, participants will move in and out of the camera's field of vision. A fixed camera at some distance may provide an overview of the entire experiment, but typically the range makes this less than ideal for detailed analysis of a participant's actions. A roaming camera is one carried by a dedicated camera operator for an experiment. Generally the footage from roaming cameras is of better quality than that of fixed cameras as they are able to follow participants around; however in order to capture every participant's actions the ratio of camera-operators to experimental participants must be 1:1. This may not be feasible, so it is likely that there will not be continuous video data for every participant. Additionally, it is often the case that the camera-operator is unable to keep the picture steady – especially in fast paced experiments. Large amounts of footage may therefore be unusable even if the camera was pointing towards interesting activity. Therefore, based on our experience of user trials, and our observations of others, we developed our requirements for the system presented here, which offers four specific benefits:

4.1 Automatically Find Video of a Logged Event

Replayer allows an analyst to visualise a timeline of all recorded events. By recording information on each camera's location and in which direction it was pointing, it can be established whether or not each of these events is likely to have been captured on video. On selecting an event of interest, Replayer is able to return only sections of video in which that event may be seen to take place.

4.2 Compiling All the Video for a Single Participant

It is common in video analysis to use a more exploratory approach to investigating a given dataset. In this form of analysis, an analyst is not searching for specific events, but rather closely examining a participant's activity across many hours of video. In this case, Replayer allows the analyst to select a single user to examine and can skip playback of the multiple streams of video to show only the periods where the participant of interest is in view.

4.3 Filtering Out Participants

With the current trend of privacy concerns, particularly when performing experiments with children and teenagers, it may be the case that some participants withdraw their consent for the use of their video footage. In a similar manner to the example in Section 4.2, Replayer allows us to filter videos for periods containing such participants, returning only video *excluding* them, and thus allows the presentation of only 'safe' data. Another application of this technique might be to exclude trial coordinators from videos. Multiple camera operators might appear in each others' recorded footage. Such periods could be filtered out from a presentation should this be desirable.

4.4 Filtering Out Video of Unacceptable Quality

In some cases, roaming camera-operators are unable to maintain a steady image when capturing in the field. They may need to run to keep up with participants, or have to pay attention to external factors such as traffic or perhaps another evaluator. In these cases where the operators' attention is withdrawn from the camera, this often results in unusable footage. Replayer now allows an analyst to automatically check for such footage and automatically skip past it.

The following section describes the processes by which these techniques operate.

5 Auto-classifying Video Content

A novel component has been added to the Replayer toolkit to automatically classify video to detect the users present in shot in any particular frame, the user activity being performed and the stability of the camera while the shot was being taken. This is achieved by logging during the experiment not only participant location, but also the camera's location and bearing. To analyse the experiment data, Replayer parses the collection of log files and automatically detects the periods of participant activity that have been recorded. The classification can then be viewed using Replayer's existing visualisation components, as illustrated in Section 6.2.

The log files for each camera contain information required to calculate the field-of-view at any particular instant. Timestamped location and bearing information are required, as are lens width and range. With static cameras that will stay in one fixed position during the course of an experiment, it is a simple matter to record this information at any time before or after the experiment. Each participant's recorded button clicks and sampled locations are checked against the camera logs to assess which periods of participant behaviour have been captured. For each user log value, the last recorded position and bearing of each camera is checked to calculate the estimated field of vision. Figure 2 illustrates how the location of the event is then checked to see if it is within one or more of these fields.

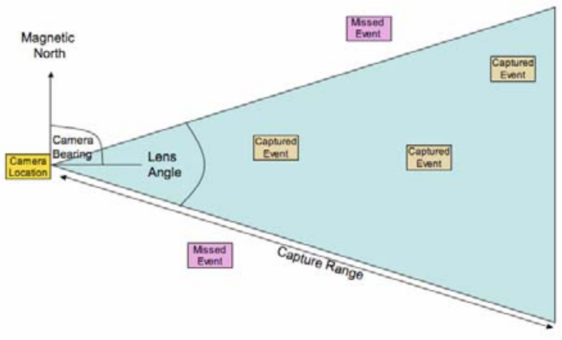


Fig. 2. Spatial distribution of events in a mobile application experiment. A triangle is created for a static camera showing its field of capture. Each event is tested to see if it is within view.

Figure 3 shows an example of two participants' logged GPS trails that have been recorded during a trial and classified using this technique. In the image on the right, having processed camera location and bearing information, the events that fall into each cameras' sights are highlighted in red. The fixed camera locations are also rendered on the map, with their fields of view shown as semi-transparent triangles. Although the examples presented here use GPS locations, any positioning system can be used, so the technique could be used on data recorded indoors.

In the trial described in the following section, the roaming camera was augmented with a PDA interfaced to the MESH inertial sensing platform [10], which provides GPS tracking along with tri-axis accelerometer, gyroscope and magnetometer sensing capabilities. Replayer is not restricted to this particular hardware, and will operate on any logged data representing geometric camera operator positions and headings..

In addition to classifying recorded activity by camera bearings, the quality of footage can be gauged by the stability of the camera over short periods of time. Accelerometers record motion in the X, Y and Z axes sampled at a rate of 100Hz. By averaging the derivatives of motion in each dimension over short windows of time, the level of camera instability for that window can be judged:

$$\sum_{a\in\{1,2,3\}} \frac{1}{k}\sum_{t=T}^{T+k}\left(\frac{x_a(t)-x_a(t-1)}{\Delta t}\right)^2 \tag{5.1}$$

over a window length k for each axis a.

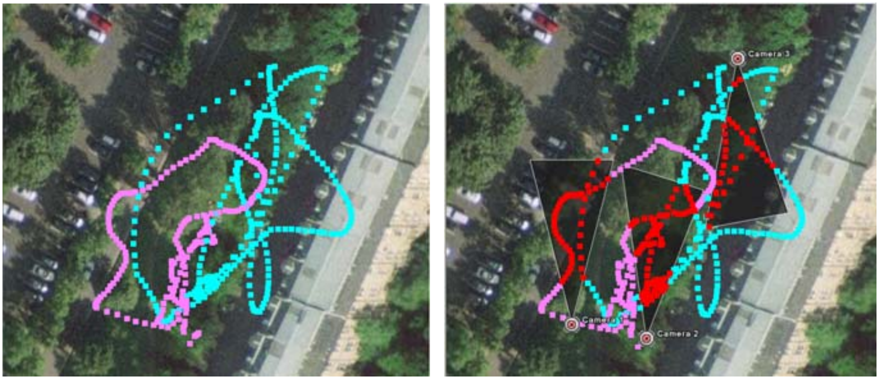


Fig. 3. The image on the left shows the GPS trails for two participants. The visual fields of the fixed position cameras are added to the image on the right, and the objects captured by the cameras are calculated and highlighted in red. Note that the participants did not enter the building, and the slightly misleading impression that they did is due to the angle from which the aerial image was captured.

Visualisation components in the Replayer interface can give overviews of the stability over time. Figure 4 illustrates, where the instability of the augmented camera is graphed. The graph shows clear periods where the camera was fairly still, others where the operator was likely to be walking and periods of high motion where the operator appears to be running. Such periods of high instability (when the footage is likely to be blurry) can then be filtered out at the analyst's discretion. Conversely, periods where the camera operator was running might indicate an occurrence of interest, which the operator was keen to capture. By selecting the high values (as was illustrated in Figure 1) or those immediately preceding or following, the video playback can instantly jump to this period.

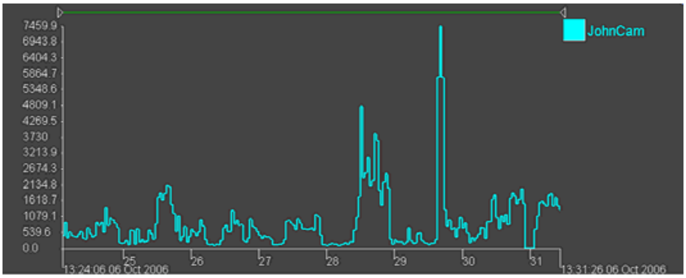


Fig. 4. Level of motion detected in augmented camera over time

The following section details an experiment where these techniques were put into practice in the evaluation of a simple mobile application.

6 Example of Use

In initial trials to demonstrate the effectiveness of these technique, we developed ColourLogger, a simple representative application for mobile devices. The application's interface shows three buttons, marked red, green and blue. Participants were asked to walk around a small area looking around for objects of these colours and, on discovery of such an item, to press the appropriate button. Although ColourLogger is a simple application it is perfectly adequate to fully illustrate the benefits of the presented techniques.

6.1 Data Capture from the ColourLogger Experiment

The trial was conducted in an area of approximately 1000m^2. The trial zone constitutes mainly of grass, with roads on either side. Trees lined around the grass meant that some areas had more overhead cover than others. Two participants (both computing science researchers) walked around this area with the ColourLogger application for ten minutes and were asked to record a button press for anything on the ground they encountered. ColourLogger was run on a Hewlett Packard iPaq hx2410 running Microsoft's PocketPC 2003 framework and using a SysOnChip compact flash GPS receiver. The time of each of the participants' button presses was recorded in a system log. A second continuous log was maintained recording timestamped GPS positions for the participant, along with the number of GPS satellites currently available. The latter value helps to determine the quality of the GPS fix.

Five video cameras were used to record activity in the experiment. Four cameras were set up in fixed locations and the fifth was carried by a camera operator who roamed around following the participants. The fixed cameras were in this case mobile phones, capturing video at 176x144 resolution. While this is not particularly high quality, it serves to demonstrate that many low cost cameras can be used to add to the variety of video streams for a given experiment.

The roaming camera used was a more traditional CCD-based digital video camera, augmented using a Hewlett Packard iPaq 5550 interfaced to the MESH inertial sensing platform [10], which logs accelerometer, gyroscope and magnetometer data at 100Hz (1Hz for the GPS device). Onboard hardware filtering is applied to the inertial readings, rolling off at around 20Hz. The device is attached to the base of the camera so that the position and orientation of the camera and sensing platform are correctly linked. The physical location of the camera is logged via the GPS, while the orientation is obtained via the magnetometers and accelerometers. The magnetometers measure the yaw angle, and the accelerometer readings are used to estimate the roll and pitch from the effect of the Earth's gravitational field. Knowledge of the roll and pitch is used to correct for variations in the magnetic field as the device is tilted, and thus obtain accurate yaw estimates. Standard strapdown inertial sensing techniques [15] are used to perform this tilt-compensation.

To ensure the veracity of the heading magnetometer readings, the camera mount shown in Figure 5 was constructed and used in the filming process. This rigidly fixes the roaming camera and MESH sensor pack together – maintaining a fixed relationship between their orientations – while providing a convenient grip for the operator. Critically, the mount also magnetically isolates the magnetometer from the disturbances induced by the battery and other metallic components within the camera, which would otherwise have significantly distorted the heading data. The accuracy of the heading data was confirmed by cross-checking the magnetometer readings against a standard magnetic compass during a calibration phase at the start of the field trials. An additional advantage to this mount is the fact that the weight of the PDA, which controls and records data from the MESH sensors counterbalances the camera, increasing the stability of the footage. The handheld camcorders typically used in such trials for their light weight and relative inexpensiveness are notorious for producing shaky footage simply because of the manner in which they are held. This mount system serves to reduces this, with the trade-off of adding more weight.

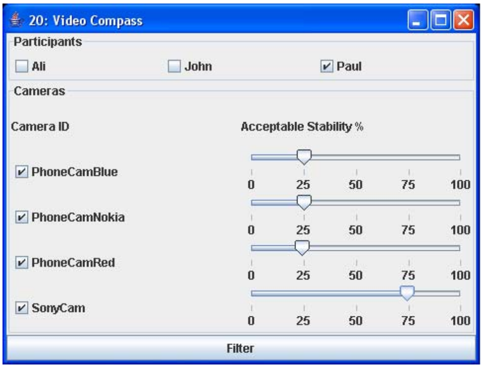
Fig. 5. The camera mount used in the trial. A PDA logging GPS locations, bearings and motion is attached to the base of the mount at a sufficient distance from the camera to be isolated from its magnetic effects.

The field-of-view of the roaming camera is a given by a cone extending from the measured location of the camera along the yaw angle estimated from the combined accelerometer and magnetometer readings. This is used to estimate the potential visibility of targets. In order to get locations and bearings for the fixed cameras, the augmented camera was positioned next to them and an annotation was made in the logs, allowing for post-hoc synchronisation. The range of each camera was estimated to be 20 metres. Although participants are visible in the footage beyond this distance, the intention was to classify the periods where they were sufficiently close to observe in reasonable detail. This value may be altered as desired.

Once complete, the data captured from the experiment consisted of the following: five video recordings at a variety of qualities and stored in a variety of codecs; one log from the augmented camera showing timestamped locations and bearings; one single-line log from each of the stationary cameras showing location and bearing; one log from each participant showing location and another showing timestamps of button events. These data logs and video files were subsequently read into the Replayer toolkit and synchronised using the QCCI [14] technique: an efficient method based on recording video footage of PDA screens that display system times, and using this information to calculate offsets for video files. To verify synchronisation, a whistle was blown at the beginning and end of the trial and recorded by all the cameras. Using Replayer's 'Play All' feature to play all the videos simultaneously verifies that the whistle is heard in each stream at the same time.

6.2 Analysing the Data

Once these logs have been parsed by Replayer, several options for examining the data are available. A number of examples of use are demonstrated below. The interface shown in Figure 6 can be used to filter the data. The checkboxes at the top allow an analyst to show or exclude data captured at times where the selected participants were in view. Similarly, checkboxes on the left allow this filtering process to involve any or all of the cameras. The sliders on the right set the minimum allowed stability for each camera, so that periods where there is a lot of camera movement can also be filtered out.



☐ Ali ☐ John ☑ Paul

☑ PhoneCamBlue
☑ PhoneCamNokia
☑ PhoneCamRed
☑ SonyCam

Fig. 6. The interface for the filtering controls

Visualising Which Events have been Captured

The graph in Figure 7 shows logged events over time. The *x*-axis covers the time from the beginning until the end of the experiment. Glyphs are placed on the *y*-axis dependent on event type (in this trial, each of the three button clicks) and are coloured by participant. The analyst can zoom or pan this graph, with the green box in the top right corner giving a context of how the current view corresponds to the full graph.

An analyst may be interested in one type of event. Existing Replayer functionality allows one row of this graph (one type of event) to be selected, which would instruct the video component to show only the corresponding time periods. Of course, there is no guarantee that all video recorded at these times have caught the events. The techniques presented in this paper allow the graph to be further filtered, so that events uncaptured on video are coloured grey in this view. Events captured but above the threshold for acceptable instability are similarly filtered. This shows the analyst exactly which logged data can be enhanced by the context provided by video footage.

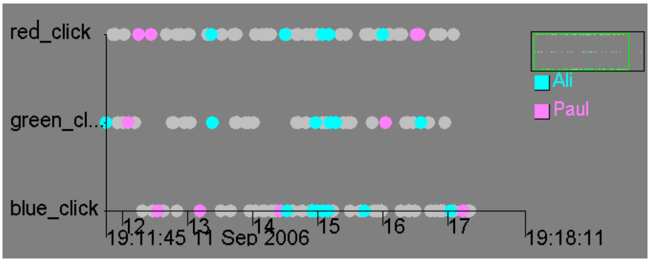


Fig. 7. An *Event Series* component showing logged events over time (*x*-axis), and the named events along the *y*-axis. Events are coloured by participant ID, and because of the selection made in another component the events that were not captured on any cameras are greyed out.

Spatial Distribution of Captured Events

Replayer shows spatial information by plotting points in Google Earth (http://earth.google.com/). The screenshot in Figure 8 shows logged GPS trails for one participants as he moved around the area in which the experiment took place. The locations and fields-of-view of the fixed cameras are also shown. From this view, it is easy to see the participant activity that has been captured, and the distances from the cameras. Those at closer range are likely to be more clearly visible than those just on the periphery of a given camera's view. Such an image would also be useful in re-positioning cameras for future experiments to capture more data.

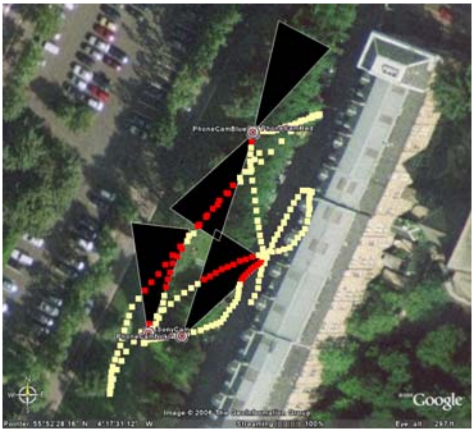


Fig. 8. GPS trail for one participant during the trial. The four static cameras are shown, with the portions of activity captured on camera coloured red.

While this particular form of visualisation is very effective when dealing with fixed cameras, it becomes more complex when dealing with roaming cameras. Figure 8 shows data recorded over time, but when handling a roaming camera, there is not a single place to draw the visual field of the camera. In such cases, an animation can be displayed: the Replayer mapping component can be set to 'replay' the recorded data, displaying only one icon for each participant and camera at a time, and showing how they moved in real-time during the trial. The roaming camera can then be shown to move around the trial area, with the triangle rotating to show the bearing at that time. This can be viewed in synchronisation with the video, to show in real time both the location of the roaming camera and the video recorded at that time.

Playback Video of a Single Participant's Activity

A third visualisation, shown in Figure 9, displays five streams of video footage and a timeline for each. An analyst has selected to see data from one of the two participants and the timelines have been automatically highlighted in green over the periods where the participant has been calculated to be in view. As can be seen, the green areas around the playback markers show periods of participant presence in the timelines labelled PhoneCamBlue, SonyCam and a short period in RoamingCam. This matches with the footage from the cameras.

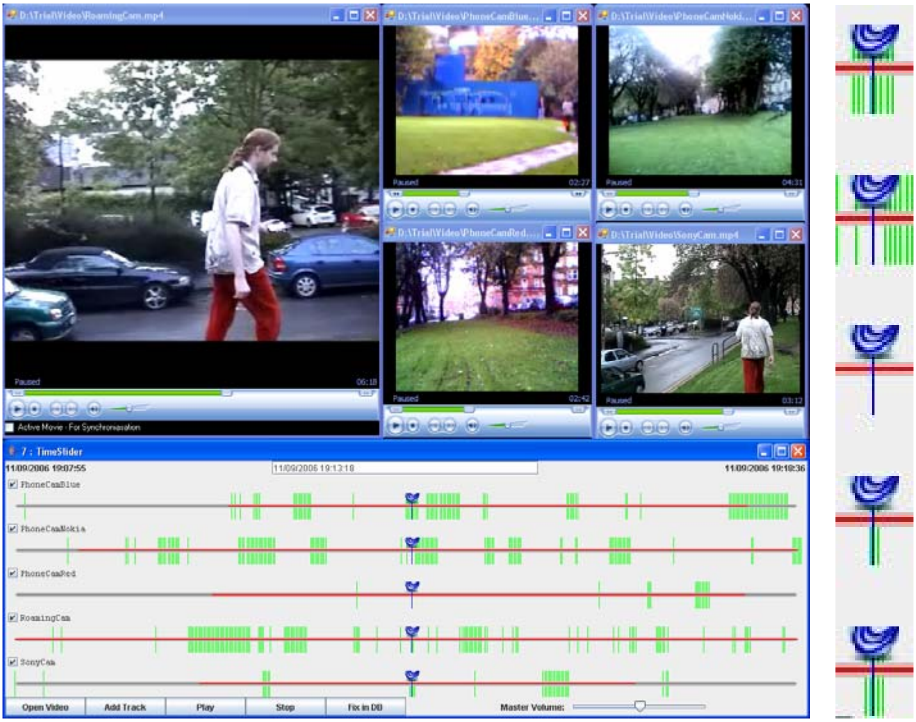


Fig. 9. Five video streams playing in synchronisation, with a timeline for each shown underneath. The analyst has selected to view all the video for a specific participant and the green highlighting on the timelines shows the periods in each of the videos where the participant is judged to appear. As can be seen, the playback marker is within the green area for three of the timelines, indicating that the participant should be in view of three of the cameras at this point. Larger images of the five video playback markers are shown on the right for clarity. Video playback can be set to skip periods where the participant does not appear in any video streams.

In the figure, all the streams are being played together. The lower part of the image shows a timeline for the experiment, which is shown to run from 19:08 until 19:18. A separate timeline is shown for each video stream, with the areas of each drawn in red showing the periods of the trial for which that camera has footage. As all the cameras were switched on and off at different times, these do not match, but Replayer synchronises the streams for concurrent playback. The thumb markers drawn on each timeline show the current position of the footage in each stream and the green areas of highlighting show the periods at which the participant was in view.

On playback, Replayer can be set to play all footage in real time or to show only the periods where the participant was recorded. In the latter mode, the system will display only those video streams that contain the participant, hiding the streams where no data is being displayed. Periods where the participant is outwith the view of all the cameras will be skipped entirely and multiple streams will be shown concurrently in the periods when the participant was captured by more than one camera.

The playback is also synchronised with other components so that, for example, glyphs on event graphs are highlighted as they occur.

Filter Video by Camera Stability

The final visualisation, shown in Figure 10, demonstrates the ability to analyse video data by camera stability. Replayer has processed the accelerometers logs from the augmented camera and created a graph of camera motion over time, as shown in the top of the figure. The video component at the bottom left shows footage from a static camera, whereas the one on the right shows that recorded from the augmented roaming camera whose stability has been graphed.

The analyst has decided to select a period of high instability from this graph. This is achieved by dragging the triangular markers at the top of the tool, which highlights the selected period in blue. As a result of this selection, the two video components below the graph jump to the beginning of the selected period. The static camera has actually filmed the roaming camera operator at this time, and it can be seen that he is filming something to his right, but appears unsteady, perhaps running or losing his balance on the slope. A participant is visible in the footage recorded from the roaming camera at this point, but, as suggested by the graph, this is occurring at a period of high motion and the video stream is unstable and blurred.

The discovery of this footage is perhaps not of great benefit to an analyst, but this example serves to show how periods of camera instability can be successfully detected automatically, so that they may be filtered out of trial playback to leave only reliable footage, should that be the analyst's desire.

Accuracy of Results

To assess the accuracy of the system on this trial data a brief evaluation was performed. Video data was filtered to show only those periods where the system had judged participants to be in view. The video was then manually reviewed to verify the system's findings. Several such comparisons were made for different cameras and filtering on different participants, with the system found to have correctly identified the periods at which participants were in view with results ranging between 61% and 82% accuracy.

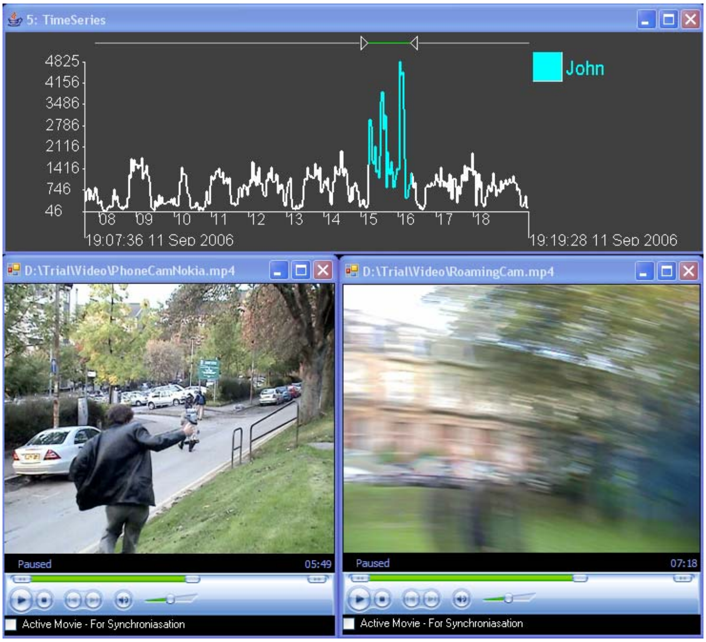


Fig. 10. The graph at the top of the figure shows instability over time of a roaming camera in the experiment. A period of high instability has been selected by an analyst, highlighting it in blue and selecting this period to view in the video components. The video at the bottom right shows the footage recorded from the roaming camera, while the video in the bottom left shows footage from a static camera that happens to have recorded the roaming cameraman at this instant. While it would be more likely that an analyst would want to filter *out* such footage, this example serves to demonstrate that periods of camera instability can be reliably identified.

These results are very dependent on the technology being employed. As GPS positioning is generally not guaranteed to provide pin-point locations, the results are not expected to be perfectly accurate in all cases. Subjective impressions of position accuracy were made by examining GPS trails and logs showing the number of satellites a participant's PDA could see at any given time. Smooth GPS trails probably indicate a good representation of a participant's actual route, whereas a more scattered display of points in the trail suggests more noisy data. From these impressions it appeared that one participant's position had been logged more accurately than the others. Of all the evaluation comparisons, the classification of this user's appearance on static cameras was found to be performed with the greatest accuracy. The example that gave percentage accuracy in only the low 60s was analysing the classification with the roaming camera of the participant who had the poorer GPS logging. In this case, inaccurate position logging of either the participant or the camera operator could lead to errors.

This analysis was not performed in a carefully managed lab environment, but has demonstrated that the presented techniques can operate with a fair degree of accuracy under realistic usage conditions. For example, the trial area was lined with trees (as can be seen in Figure 3), meaning that GPS positioning was more accurate in some areas than others, as would be the case in many real system trials. Additionally, as one would expect, the hardware used in the trial did not all perform to the same standard, and the more precise sensors led to better classification of video. It is also worth noting that the presented techniques are not dependent on GPS technology; as positioning technologies continue to improve, so will the accuracy of this technique.

Summary

In performing qualitative evaluation of a pervasive computing system, an analyst may wish to examine many hours of video recorded on several cameras for each of several trials. Such a volume of data is very time consuming to work through looking for specific events or well-filmed actions. Although an analyst may wish to examine all the available footage in a thorough evaluation, there is clear benefit in being able to quickly find the most interesting periods without resorting to a linear search.

With the techniques presented here, an analyst can select a particular logged event of interest (for example, a specific type of user interaction, or a recorded system event such as entering into wireless network range) and be shown every occurrence of these on a timeline. This collection of events can be filtered to leave only those that were captured on video and those recorded when the camera was sufficiently steady to provide a clear image. Should the analyst wish, the recorded events can be further analysed on a map to see how far they occurred from the camera, and only those that occurred close to a camera can be selected. Having made this selection, the analyst can watch the events of interest across several concurrent video streams.

The simple ColourLogger application demonstrated here was intended to be representative of many mobile applications, and to allow for our experimentation and development to take place. Of course, the value of examining the data collected about the use of ColourLogger is minimal, but it has served as a stepping stone towards use of the new Replayer extension in the trials of new mobile applications that are now in development. It is anticipated that these techniques will play an integral part in the running of system trials involving multiple analysts and multiple participants using our systems over a large geographic area and over a significant period of time. Evaluation of the system is also still ongoing, and the next step will be to give the system to analysts other than ourselves to assess how it assists their work practices and to determine whether video can be classified with acceptable accuracy. Of particular interest would be comparing alternative methods of evaluation: judging the benefit of the presented techniques by comparing analyses performed with and without Replayer. Other future work is described in the next section.

7 Future Work

As we continue to work on Replayer, one particular use we envisage is in the orchestration of a system trial. The sort of data described here could be uploaded and analysed in real time, something not implausible given the prevalent availability of wireless communications in modern devices; indeed all the devices used in our example were equipped with 802.11 wi-fi and could have been sending data directly to a server. It would then be possible, for example, to show a histogram of the amount of activity captured for each participant. An orchestrator seeing inequality in this histogram could direct roaming cameras to concentrate on particular participants as required. Of course this brings up issues of assured continuous connectivity, something that [5] points out may not be as simple to achieve as it seems.

Several possibilities exist for further refining the techniques, with a view to increasing accuracy or providing additional filters for unusable video. The degree to which participant and camera positions can be correctly logged is obviously important in successfully classifying the video, so it would be interesting to incorporate hybrid techniques that combine GPS with wi-fi and other beacon types such as Place Lab (http://www.placelab.org) and Navizon (www.navizon.com) to see what benefits these could offer to the classification accuracy. We also aim to examine the issue of occlusion. A feature being integrated into Google Earth is a 3D representation of all the buildings in a given city; we hope to use this information to limit the modelled fields of view from each camera to reflect this occlusion. Additionally, sensor information is available in three axes, so, in the case of roaming cameras, we also intend to allow analysts to quickly discard footage where the camera is pointing at the ground or sky. Another approach would be to incorporate computer vision techniques to further classify the contents of the video. Properties such as poor lighting or lens flare could be identified, and the analyst would be given the opportunity to filter out such periods of footage.

8 Conclusion

We have presented a system for analysis of data recorded in mobile application evaluations. Specific recorded user activity can be queried, to automatically skip irrelevant video footage among the volumes of data recorded by multiple cameras, and to focus on that which is salient. This reduction can include, for example, showing only the video in which a particular participant appears, or showing only that area of video in which a particular system event has been captured. Additionally we demonstrated a technique by which video of an unacceptable image stability can be automatically discarded. When performing an evaluation of a larger scale system, potentially hundreds of hours of video may be captured. The technique described in this paper allows an analyst to quickly locate specific data of interest within that footage. We devised, implemented and recorded a representative application to demonstrate how this method might be used, and showed some of the capabilities provided by this extension to the Replayer toolkit. This extension of Replayer does require some additional hardware; however, the hardware used is becoming increasingly available in commodity devices. We believe this is a valuable addition to the already versatile Replayer toolkit, and suggest that it could be widely applicable in the evaluation of mobile computing systems. It serves as another example of the benefits arising from combining quantitative and qualitative data, and from combining associated analytic approaches in synergistic ways.

Acknowledgements

We thank Malcolm Hall, Jose Rojas, Marek Bell, Louise Barkhuus, Andrew Crossan and John Ferguson for operating cameras during our initial trials. This research was funded by UK EPSRC (GR/N15986/01), UK ESRC (RES-149-25-1046) and by the EU PASCAL Network of Excellence.

References

1. Aizawa, K. Ishijima, K. Shiina, M.: Summarizing Wearable Video, In Proceedings of International Conference on Image Processing, Thessaloniki, Greece (2001) pp. 398-401.
2. Badre, A. N., Guzdial, M., Hudson, S. E., and Santos, P. J.: A User Interface Evaluation Environment Using Synchronized Video, Visualizations and Event Trace Data. Software Quality Journal, Volume 4, Number 2 / June, 1995, Springer.
3. Barkhuus L., Chalmers, M., Hall, M. et al. Picking Pockets on the Lawn: The Development of Tactics and Strategies in a Mobile Game, Proc. Ubiquitous Computing (Ubicomp), Tokyo, LNCS 3660, pp. 358-374 (2005).
4. Becker, R.A., Cleveland, W.S.: Brushing Scatterplots. Technometrics, Vol. 29 (1987) pp.127-142.
5. Beeharee, A., Steed, A.: Filtering Location-Based Information Using Visibility. In Proceedings of Location- and Context-Awareness, Munich, Germany (2005) pp. 306-315.
6. Bell, M., Chalmers, M. Barkhuus L. et al., Interweaving Mobile Games with Everyday Life, Proc. ACM CHI, Montreal (2006), pp. 417-426.
7. Chalmers, M., Galani, A.: Seamful Interweaving: Heterogeneity in the Theory and Design of Interactive Systems. In Proceedings of ACM Symposium on Designing Interactive Systems (DIS 2004), Massachusetts, USA (2004), pp. 243-252.
8. de Silva, G. C., Ishikawa, T., Yamasaki, T and. Aizawa, K.: Person Tracking and Multicamera Video Retrieval Using Floor Sensors in a Ubiquitous Environment. In Proceedings of Interbational Conference on Image and video retrieval (CIVR2005), Singapore (2005) pp 297-306.
9. French, A., Greenhalgh, C., Crabtree, A., Wright, M., Hampshire, A. and Rodden, T.:Software Replay Tools for Time-based Social Science Data. Presented at the 2nd International Conference on e-Social Science, Manchester (2006).
10. Hughes, S., Oakley, I., O'Modhrain, S.: MESH: Supporting Mobile Multi-modal Interfaces. Presented at the Seventeenth Annual ACM Symposium on User Interface Software and Technology (UIST 2004), Santa Fe, New Mexico, USA (2004).
11. McCurdy, N.J., Carlisle, J.N., Griswold, W.G., Harnessing Mobile Ubiquitous Video. In: Proceedings of the ACM Conference on Human Factors in Computing (CHI2005), Portland, Oregon, USA (2005) pp. 1645-1648.
12. Morrison, A., Tennent, P., Chalmers, M.: Coordinated Visualisation of Video and System Log Data. In: Proceedings of the Fourth International Conference on Coordinated & Multiple Views in Exploratory Visualization (CMV2006), London, UK (2006) pp. 91–102.
13. Sawahata, Y., Aizawa, K., Bakker, E. M. et al (Eds.).: Indexing of Personal Video Captured by a Wearable Imaging System. In: Proceedings of International Conference on Image and Video Retrieval (CIVR 2003). Urbana, IL, USA (2003) pp. 342–351.
14. Tennent, P., Chalmers, M.: Recording and Understanding Mobile People and Mobile Technology. In: Proceedings of the First International Conference on e-Social Science, Manchester, UK. (2005).
15. Titterton, D. H., Weston, J.L.: Strapdown Inertial Navigation Technology, 2nd edition. The Institution of Electrical Engineers (2004).

Inference Attacks on Location Tracks

John Krumm

Microsoft Research
One Microsoft Way
Redmond, WA, USA
`jckrumm@microsoft.com`

Abstract. Although the privacy threats and countermeasures associated with location data are well known, there has not been a thorough experiment to assess the effectiveness of either. We examine location data gathered from volunteer subjects to quantify how well four different algorithms can identify the subjects' home locations and then their identities using a freely available, programmable Web search engine. Our procedure can identify at least a small fraction of the subjects and a larger fraction of their home addresses. We then apply three different obscuration countermeasures designed to foil the privacy attacks: spatial cloaking, noise, and rounding. We show how much obscuration is necessary to maintain the privacy of all the subjects.

Keywords: location, privacy, inference attack, location tracks.

1 Introduction

Location is an important aspect of context in pervasive computing. As location-sensitive devices pervade, it becomes important to assess privacy threats and countermeasures for location data. Privacy advocates worry that location data can be used to harm a person economically, invite unwelcome advertisements, enable stalking or physical attacks, or infer embarrassing proclivities[7, 28]. Except for isolated incidents, these threats have remained largely hypothetical, as have the proposed countermeasures. This paper is a first attempt at actually testing and quantifying one type of privacy threat using real location data: we try to identify individuals based on anonymized GPS tracks. With an attack in place, we are also able to quantify the effectiveness of some countermeasures that have been proposed in the literature.

Despite the potential harm, people generally do not place a high value on the privacy of their location data. Danezis *et al.*[6] found that 74 students set a median price of £10 (about US$ 18 at the time of publication) to reveal 28 days of personal location tracks for research purposes. The median set price doubled if the data was going to be used for commercial purposes. In our own GPS survey (described below), we easily convinced 219 people from our institution to gives us two weeks of their driving data for a 1 in 100 chance to win a US$ 200 MP3 player. A survey of 11 participants with a mobile, location-sensitive message service found that privacy concerns were fairly light[16]. In our GPS study, only 13 of 62 (21%) whom we

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 127 – 143, 2007.

asked insisted on not sharing their location data outside our institution. In 55 interviews with subjects in Finland, Kaasinen[17] found that "… the interviewees were not worried about privacy issues with location-aware services." However, he adds, "It did not occur to most of the interviewees that they could be located while using the service."

It may be that the implications of leaked location data will not be adequately understood until there is a widely publicized incident of an innocent victim being seriously harmed. A recent story[8] in the New York Daily News describes how a suspected killer was caught via cell phone tracking, but the tracked person was not one of the "good guys". There have been at least two incidents in the U.S. where a man tracked an ex-wife or ex-girlfriend by secretly installing a GPS in her car[25].

This paper takes a different approach to exposing the risks of leaked location data by quantitatively assessing the threat using real location tracks to infer a person's identity. Tracks such as these can be used benevolently to assess traffic[22], train a system about a user's habits[26], create customized driving routes[20], help predict where a user is going[19], or create a travelogue[10]. To protect this data from malicious inferences, researchers have proposed pseudonymity[27], which attaches a persistent ID to the GPS data but that does not link the ID to the identity of the user. Pseudonymity was the same scheme used to protect the identities of AOL search users when their search query logs were released and subsequently retracted by AOL. The identity of searcher pseudonym "4417749" was uncovered from the search logs by a reporter[3]. In this paper, we assess the effectiveness of pseudonymity on GPS logs.

Based on two-week (or longer) GPS tracks from 172 known individuals, we developed four heuristic algorithms to identify the latitude and longitude of their homes. From these locations, we used a free Web service to do a reverse "white pages" lookup, which takes a latitude and longitude coordinate as input and gives an address and name. We report the fraction of the individuals we were able to correctly identify and the fraction whose home address we found based on our four home-finding algorithms. We go on to assess the effectiveness of three obscuration algorithms that attempt to alter the GPS data in a way to prevent our privacy attacks. This is the first paper we know of to assess quantitatively the risk of identifying the persons associated with leaked, pseudonymized location tracks.

Analyzing data in order to illegitimately gain knowledge about a subject is known as an "inference attack". Our tests are intended to mimic what an attacker would do with a large volume of location data from several individuals, assuming he or she has defeated any encryption or access control on the data. We assume the attacker's goal is to identify the subjects after which he or she would nefariously profit from a multitude of associated identities and location tracks. The large volume of data necessitates an automated approach of the type we implement. Clearly an attacker with a smaller set of potential victims could afford more time-consuming means of identifying them by physically staking out their neighborhood or manually inspecting their location tracks. Our attacks are limited to computation.

Our tests are based on GPS data gathered from volunteer drivers, which we describe in the next section.

2 Multiperson Location Survey

Our Microsoft Multiperson Location Survey (MSMLS) is an ongoing survey of where people drive. We loan subjects a Garmin Geko 201 GPS receiver, capable of automatically recording 10,000 time-stamped latitude and longitude coordinates. The GPSs are powered from the car's cigarette lighter, and a simple hardware modification ensures that the GPS turns on whenever it detects available power. This is necessary for the cars whose cigarette lighter is powered only when the car is on. Subjects are instructed to leave the GPS on their car's dashboard. We set up the GPS in an adaptive recording mode so it ceases to record when the car is stopped. This prevents the memory from filling up while the car is parked. With this recording mode, we found that the median separation between points is 64.4 meters in distance and 6 seconds in time. Each subject recorded data for at least two weeks.

We recruited subjects from our institution and allowed their adult family members to participate as well. Subjects are compensated by being entered in a drawing from 100 subjects to win an MP3 player worth about US$ 200. Before receiving the GPS receiver, each subject fills out an online survey, whose questions include the subject's name, home address, and other demographic information. The subject's name and home address data serve as the ground truth for assessing our privacy attacks and countermeasures. At the time of the study, we had data from 172 drivers whose addresses were recognizable by our reverse geocoder. These are the subjects we used for the tests in this paper. From the demographic data, 72% were male, 75% had a domestic partner, 37% had children, and the average age of drivers was 37.

Other location-gathering efforts include Ashbrook & Starner's[2] two studies of subjects with wearable GPS recorders. One had a single subject for 4 months, and the second had six users for 7 months. Their GPS recorders could hold 200,000 points, compared to our 10,000. Liao *et al.*[21] gathered GPS data from one person for four months and subsequently five people for one week. As of this writing, the OpenStreetMap[15] project has 5511 GPS traces contributed by volunteers in an effort to produce copyright-free maps.

3 Inferring Home and Identity

Given a set of time-stamped latitude and longitude coordinates, the first step in our privacy attack is to infer the coordinates of the subject's home. This section describes how we first computed the location of a subject's home and then the subject's identity from pseudonymous GPS data.

3.1 Related Efforts

The general problem of extracting significant places from location data has received much attention. Marmasse and Schmandt's comMotion[24] system designated as significant those places where the GPS signal was lost three or more times within a given radius, normally due to a building blocking the signal, after which the user was prompted for a place name. Marmasse's subsequent work[23] looked at a combination of dwell time, breaks in time or distance, and periods of low GPS accuracy as potentially significant locations. Ashbrook & Starner[2] clustered places where the GPS signal was lost and asked users to name such locations. Using locations generated from Place Lab, Kang *et al.* [18] used time-based clustering to identify places that the user would likely find important. Hariharan & Toyama[11] created a time- and location-sensitive clustering technique to hierarchically represent "stays" and "destinations". Liao *et al.*[21] used this algorithm to find a user's frequent destinations for higher-level machine learning about a user's habits. Hightower *et al.*'s BeaconPrint[12] algorithm finds repeatable sets of GSM and Wi-Fi base stations where a user dwells. This is interesting in that it does not use spatial coordinates as a location indicator, but instead sets of consistently heard radio transmitters. Subramanya *et al.*[30] used a dynamic probabilistic model on inputs from GPS and other sensors to classify the user's motion state (*e.g.* stationary, walking, driving, *etc.*) as well as the type of location from among indoors, outdoors, or vehicle.

Of the work above, only Liao *et al.*[21] made an attempt to automatically determine which of the important places are the subject's home. They used machine learning on labeled place data to achieve 100% classification accuracy in finding locations of their five subjects' home and work places.

The work most closely related to ours is from Hoh *et al.*[13] who used a database of week-long GPS traces from 239 drivers in the Detroit, MI, USA area. Examining a subset of 65 drivers, their home-finding algorithm was able to find plausible home locations of about 85%, although the authors did not know the actual locations of the drivers' homes. Our study is based on drivers' self-reported home addresses, and we also attempt to infer the drivers' names as well as home locations.

3.2 Finding Homes in GPS Traces

Our first challenge is to find the coordinates of each subject's home based on their GPS data. For each subject, we have a list of time-stamped latitude and longitude points. We tested four algorithms for picking out the location of the subject's home, two of which depend on segmenting the GPS data into discrete trips. Our segmentation is simple: we sort the list of points by time and split it into candidate trips at points which are separated by more than five minutes. We then retain only those trip segments that meet three criteria:

1. The trip must have at least ten measured points.
2. The trip must be at least one kilometer long.
3. The trip must have at least one pair of points during which the speed was at least 25 miles/hour. This helps eliminate walking and bicycle trips which we are not trying to analyze.

The first two criteria tend to eliminate noise trips that result from random data gathered from parked vehicles. The final point in each trip segment is the trip's destination, which gives us a list of latitude and longitude points, one for each trip, some of which are likely the location of the subject's home.

These are our four heuristic algorithms for computing the coordinates of each subject's home:

Last Destination – This algorithm is based on the heuristic that the last destination of the day is often a subject's home. For each day of the survey, we found the destination closest to, but not later than, 3 a.m. We computed the median latitude and longitude of these destinations for our estimate of the home location.

Weighted Median – We assume that the subject spends more time at home than at any other location. Each coordinate in the survey (not just the destinations) is weighted by the dwell time at that point, *i.e.* the amount of time until the next point was recorded. The weighted median latitude and longitude is taken as the home location. The weighted median can be thought of as a regular median of a set of values, but values are repeated in the set proportional to a second set of corresponding weights. Thus, if a point is recorded at 8 p.m. as the subject parks his car at home, and if and nothing else recorded until 8 a.m. when the subject leaves home, the point recorded at 8 p.m. will have a much higher weight than points recorded at more frequent intervals during travel. This method implicitly accounts for the variable recording rate of our GPS receivers and it avoids the need for segmentation into trips.

Largest Cluster – This heuristic assumes that most of a subject's coordinates will be at home. We build a dendrogram of the subject's destinations, where the merge criterion is the distance between the cluster centroids. The dendrogram is a common, agglomerative, hierarchical, clustering technique. We stop clustering when the nearest two clusters are over 100 meters apart. The home location is taken as the centroid of the cluster with the most points.

Best Time – This is the most principled (and worst performing) algorithm for finding the subject's home. It learns a distribution over time giving the probability that the subject is home. For each measured location (not just the destinations), we reverse geocoded the latitude and longitude coordinates into a street address. Reverse geocoding takes a (latitude, longitude) and returns a street address or other symbolic representation of the location. We used the MapPoint® Web Service (MPWS) as our reverse geocoder. From our survey, we took each subject's self-reported home address and normalized it to the same format used by MPWS. Looking at 30-minute intervals in time, we computed the frequency with which the reverse geocoded points matched the subject's actual address. In order to compensate for the GPS's adaptive sampling times, we resampled all the measured location traces at one-minute intervals to avoid biasing the distribution away from times when points were recorded infrequently. The relative probability of being at home *vs.* time of day is shown in Figure 1. As expected, people are more likely to be home at night than during the day. Applying this distribution, we compute the relative probability of being home for each measured latitude and longitude for each subject. We extract those coordinates for each subject that have the maximum relative probability and take the home location as the median of those points.

The work most similar to ours, Hoh *et al.*[13], used heuristics similar to ours for finding homes based on GPS traces. They first dropped GPS samples recorded at speeds greater than one meter/second and then applied agglomerative clustering until

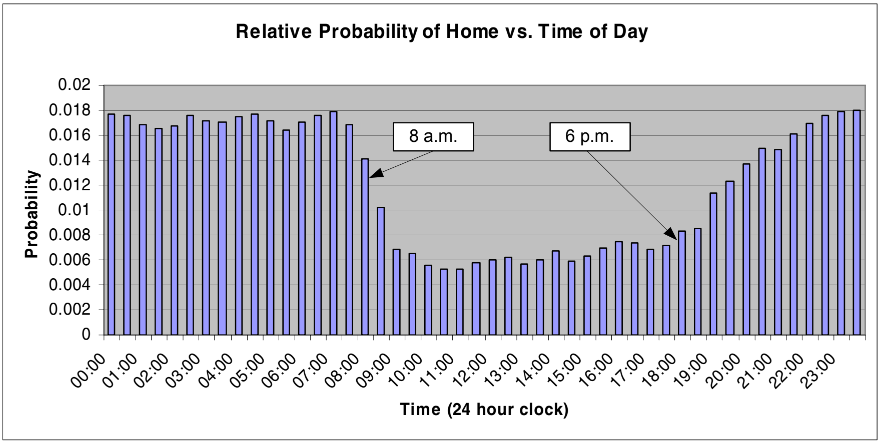


Fig. 1. The relative probability of a subject being at home varies depending on the time of day. We used this distribution in our "Best Time" algorithm to help determine the location of a subject's home.

the clusters reached an average size of 100 meters. They eliminated clusters with no recorded points between 4 p.m. and midnight as well as clusters deemed outside residential areas by manual inspection of maps.

Although our ultimate goal is to infer identities, we can assess the performance of our algorithms at an intermediate step by evaluating how well they locate each subject's home. For this evaluation, we used MPWS to geocode the location (*i.e.* find the latitude and longitude coordinates) of each subject's home based on their reported addresses. We then computed the errors between the geocoded locations and the inferred locations given by our four algorithms. The best performing algorithm, in terms of median error, was "Last Destination", whose median error was 60.7 meters. "Weighted Median" and "Largest Cluster" had nearly the same median errors, at 66.6 meters. "Best Time" was significantly worse with a median error of 2390.2 meters. Figure 2 shows a histogram of home location errors from the best-performing "Last Destination" algorithm. Based on these results, we can conclude that an attacker, using data like ours, could computationally locate a subject's home to within about 60 meters at least half the time.

The "Best Time" algorithm reveals an interesting characteristic of our reverse geocoding solution. Reverse geocoding is an integral part of our privacy attack, because it is the link from a raw coordinate to a home address and ultimately to an identity via a white pages lookup. In developing the probability distribution for "Best Time", we found only 1.2% of the measured points were reverse geocoded to the subjects' self-reported home addresses. This is after we resampled our data at a constant one minute interval to compensate for our GPS's adaptive recording mode. This is why Figure 1 shows only relative probability (normalized to a sum of one), not the absolute probability of a subject being at home, because the computed absolute

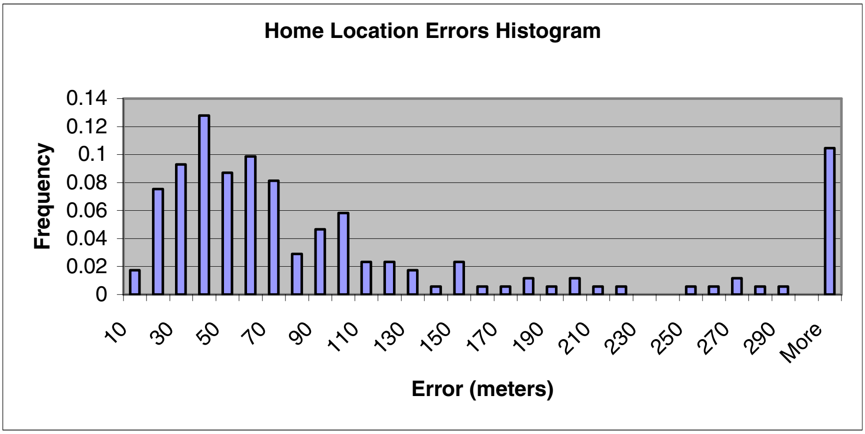


Fig. 2. This is a histogram of the home location errors from the "Last Destination" algorithm. The median error was 60.7 meters.

probabilities are clearly too small. For the purposes of “Best Time”, it is enough to know only the relative probabilities. In evaluating our reverse geocoder, it seems extremely unlikely that our subjects collectively spend only 1.2% of their time at home, which makes us suspicious that our reverse geocoder was usually not giving the correct address corresponding to the measured coordinates. This is one weak point in the type of attacks we are examining.

3.3 From Home Coordinates to Identity

Armed with a likely coordinate of a subject’s home, the final step in our privacy attack is to find the subject’s identity. We accomplished this with a Web-based, white pages lookup. Windows Live™ Search has a freely downloadable API[14] that allows no-cost, programmatic access to its search capabilities. In “phone book” mode, the search engine can be set up to return street addresses and associated names within a given radius of a given coordinate. There are several paid services available on the Web which give the same information.When our search engine returned multiple results, we took the one physically nearest the given coordinate based on the search engine’s returned latitude and longitude fields.

3.4 Attacker Summary

Summarizing our assumptions about the attacker, we assume the following:

- The attacker has access to about two weeks of time-stamped GPS data recorded from 172 unknown drivers. The GPS receivers are in the drivers’ vehicles, not carried on the drivers themselves. The GPS data is recorded at a median interval of 6 seconds and 64.4 meters.

- The GPS data points for each driver are tagged with a common pseudonym such that all the GPS data for each driver can be easily grouped together and distinguished from data for the other drivers.
- The attack consists of first trying to computationally identify the latitude and longitude of each driver's home based on the GPS data. Then these coordinates are used to find the driver's name using a Web search.

3.5 Results

We applied the four algorithms in Section 3.2 to each of the subjects in our study. Each algorithm gives a single coordindate as a guess for the subject's home location. We submitted these locations to our search engine and manually compared the subject's name to the name(s) returned. Sometimes the search engine returned the names of two people living at the same address. We counted the return as a success if it contained at least the subject's name. We also counted the return as a success if it returned just the subject's first initial and last name, *e.g.* "G. Washington" was considered a valid match for "George Washington".

Of the 172 subjects, the four inference algorithms performed as follows:

Algorithm	Number Correct Out of 172	Percent Correct
Last Destination	8	4.7%
Weighted Median	9	5.2%
Largest Cluster	9	5.2%
Best Time	2	1.2%

This shows that there is a legitimate danger in releasing pseudonymized location data to potential attackers. However, the number of successful identifications was not high. We speculate that these low rates were caused by three main types of problems:

- Measurements
    - Inaccurate GPS. GPS may not report a location near enough to a subject's house due to its inherent inaccuracies.
    - Missing GPS. Our adaptive recording mode may not have captured points close enough to the subject's home, especially if a subject, upon arriving at home, drove immediately into a parking structure that blocks GPS.
    - Inaccurate home location heuristics. As shown above, our best home location algorithm has a median error of 60.7 meters.
- Database
    - Inaccurate reverse geocoding. This is apparent from the reverse geocoding we did for the "Best Time" algorithm, in which only 1.2% of the measured GPS points were coded to the subject's self-reported home address. Reverse geocoding normally works by linearly interpolating to a house number based on addresses at the street intersections. The reverse geocoder is generally unaware of different sized land lots and house number gaps.
    - Outdated and/or inaccurate white pages data. We performed a white pages search for the self-reported addresses of our subjects. As shown

in Figure 3, only 33% of the subjects' names were found listed with their addresses. 11% of the address listings had different names listed, possibly because the subject had moved. 43% of the addresses were not found in the white pages.

- Subject behavior
    - Parking locations distant from home locations. Some subjects may park their cars at a distance from their actual homes. Tracking the subjects themselves rather than their vehicles may have compromised more identies.
    - Multiunit buildings. The coordinates of a parked vehicle are not a good clue to the exact housing unit of subject who lives in an apartment building or condominium. Using the analysis from the Figure 3, we found that 13% of our subjects lived in multiunit buildings.

Figure 3 shows that, based on the white pages, 33% of the subjects could have been found, while the remainder were masked in one way or another. Despite these vagaries, we can say that there is a least a 5% chance that an attacker can infer a subject's identity based on two weeks of GPS data like ours using a fairly simple algorithm and free, Web-based lookups.

Clearly there are countermeasures available, such as encryption and strict privacy policies. Our study is intended to highlight the need for such countermeasures by

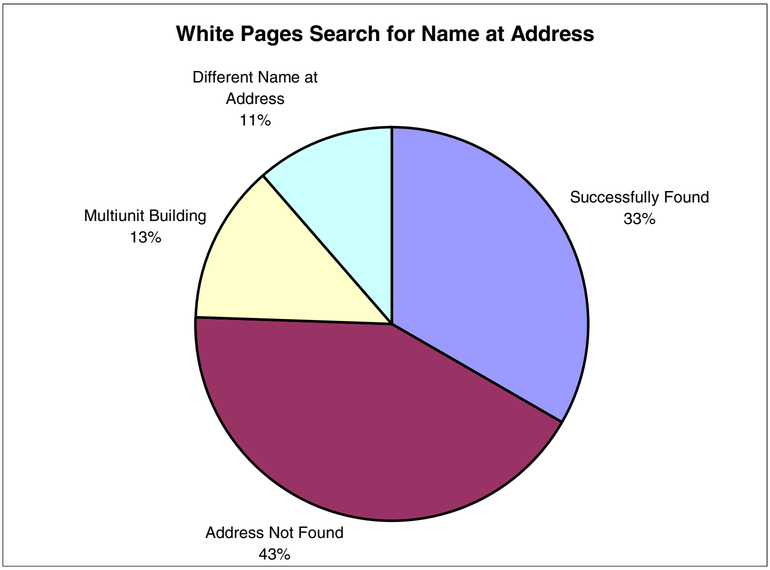


Fig. 3. We used our white pages search to look up the name(s) associated with our subjects' self-reported addresses. Of all the addresses, only 33% yielded a set of names that included the corresponding subject. The rest were either addresses in a multiunit building giving several possible names (for which the subject usually did not appear), a different name given for the address, or the address was not found.

showing how vulnerable location data is to simple privacy attacks. In the next section we test some computational countermeasures that are designed to obscure the location of the subject's home by corrupting the data.

4 Countermeasures

Pseudonymity is one countermeasure to protect the identity of people if their location history is exposed. In this section, we test three additional countermeasures that have been previously proposed in the research literature. These could be applied to pseudonymized location data to foil the attack presented in the previous section. In addition to regulatory and privacy policy methods, Duckham and Kulik[7] describe a variety of computational approaches to protecting location privacy:

Pseudonymity – This is the technique we examined above, which consists of stripping names from location data and replacing them with arbitrary IDs.

Spatial Cloaking – Gruteser and Grunwald[9] introduce the concept of spatial cloaking. A subject is k-anonymous if her reported location is imprecise enough for her to be indistinguishable from at least k-1 other subjects. Scott *et al.* [29] speculate that software agents could be used to implement spatial cloaking. Beresford and Stajano[4] introduce a related concept called "mix zones". These are physical regions in which subjects' pseudonyms can be shuffled among themselves to confuse an inference attack.

Noise – If location data is noisy, it will not be useful for inferring the actual location of the subject. This technique is called "value distortion" in Agrawal and Srikant's work on privacy-preserving data mining[1].

Rounding – If the location data is too coarse, it will not correspond to the subject's actual location. This is called "value-class membership" in [1].

Vagueness – Subjects may report a place name (*e.g.* home, work, school, mall) instead of latitude and longitude. In their study on disclosure of location to social relations, Consolvo *et al.*[5] found that vagueness was not popular for mobile users communicating their locations to family and friends, although it would likely be more acceptable for disclosure to strangers.

In addition, Hoh *et al.*[13] describe one other computational technique:

Dropped Samples – Hoh *et al.* found that reducing the GPS sampling interval from one minute to four minutes reduced the home identification rate from 85% to 40%.

We have already tested the effectiveness of pseudonymity in the previous section. In this section, we measure the effectiveness of noise, rounding, and one type of spatial cloaking applied on top of pseudonymity.

These computational countermeasures contrast with information-preserving countermeasures like encryption and access control, which are beyond the scope of this paper. While these other techniques may be better, we speculate that drivers would be more comfortable releasing corrupted data than they would accepting an authority's promise to be careful with the uncorrupted data.

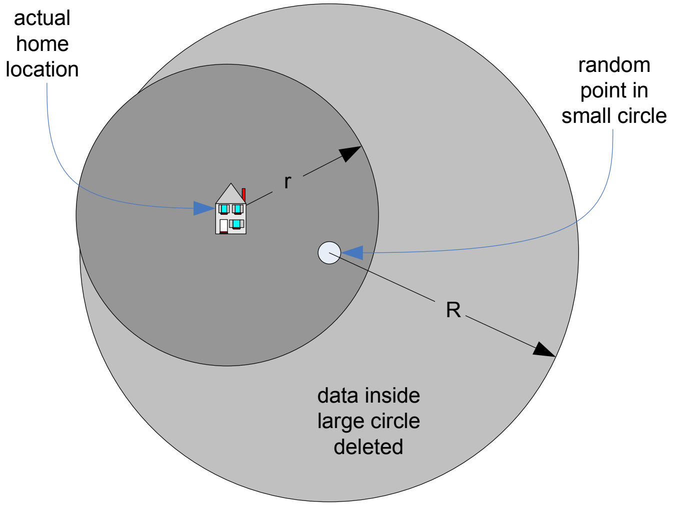


Fig. 4. We delete data inside the large circle to help conceal the home's location

4.1 Countermeasure Specifics

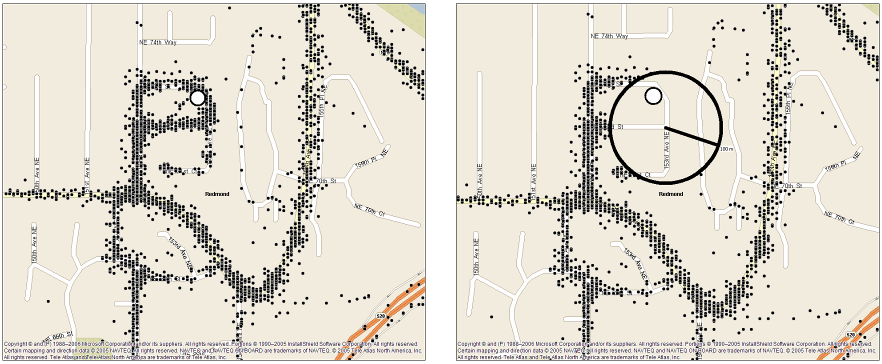
Our three countermeasures apply to the raw latitude and longitude data. In a real setting, they would be applied to the data near the source before it is transmitted anywhere an attacker could access it. The specifics of the three methods are:

Spatial Cloaking – Previously described spatial cloaking is applied to groups of people in the same region. We implemented an alternative that uses only a single user's data. It simply deletes coordinates near the subject's home, creating ambiguity about the home's actual location. A simple version of the algorithm would delete all points in a circle centered at the subject's home. However, a geometrically-minded attacker might be able to guess the circle's radius and find the center. Instead, we center the "invisibility circle" at a random point inside a smaller circle which is centered on the home's location. Specifically, we have a circle of radius r centered on the home. We pick a uniformly distributed, random latitude and longitude coordinate inside this circle as the center of a larger circle of radius R, $r < R$. We delete all the measured coordinates inside the larger circle. This is illustrated in Figure 4. This process is applied to each subject with a different random point for each one. This ensures that points at and near the home will be deleted, and the randomness for each subject makes it more difficult for the attacker to infer the geometry of the deletions.

Noise – We implement noise by simply adding 2D, Gaussian noise to each measured latitude and longitude coordinate. For each point, we generate a noise vector with a random uniform direction over $[0,2\pi)$ and a Gaussian-distributed magnitude from $N\left(0,\sigma^2\right)$. A negative magnitude reverses the direction of the noise vector.

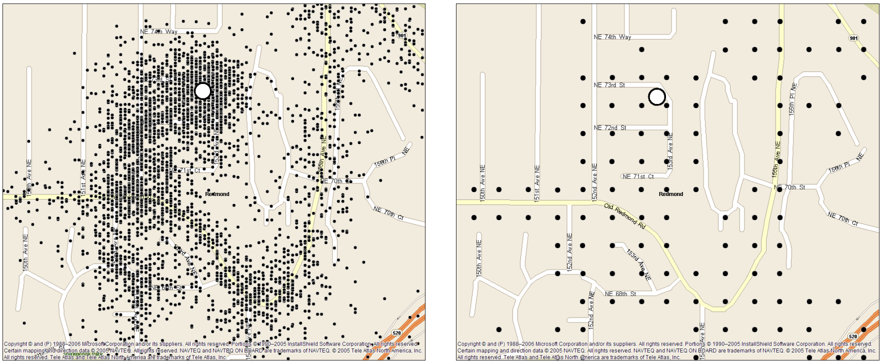
Rounding – We snap each latitude and longitude to the nearest point on a square grid with spacing Δ in meters.

Figure 5 shows the effect of these three methods on data from one of our survey's subjects.



Uncorrupted GPS data

Spatial Cloaking – Points inside circle deleted. 100 meter radius circle was centered at a random point near the home.



Noise – Added Gaussian noise with 50 meter standard deviation

Rounding – Each point snapped to nearest point on 50 meter × 50 meter grid.

Fig. 5. This demonstrates the effect of our three methods of corrupting GPS data to enhance privacy on a set of data from one of our subjects. The upper left image shows the raw, uncorrupted GPS data. The white circle in each image shows the location of the subject's house.

4.2 Countermeasure Results

We evaluated the effectiveness of our three countermeasures as a function of their various parameters. In evaluating pseudonymity above, we considered what fraction of names we could correctly identify. However, this is highly dependent on the quality of our white pages lookup, which we demonstrated as poor. For evaluating the three countermeasures, we instead measured how many correct home addresses we could find, which eliminates the uncertainty caused by poor white pages. To find the address associated with the inferred coordinates of a home, we used the MapPoint® Web Service. As a baseline, when run on unaltered coordinate data, the four inference algorithms correctly find these fractions of the subjects' home addresses:

Algorithm	Number Correct Out of 172	Percent Correct
Last Destination	22	12.8%
Weighted Median	20	11.6%
Largest Cluster	19	11.0%
Best Time	6	3.5%

We are trying to find how much we have to corrupt the GPS data for the three countermeasures to significantly reduce the number of correct address inferences.

Spatial Cloaking – To simplify presentation, we present only the best-performing "Last Destination" algorithm for spatial cloaking. The results for varying values of R and r (see Figure 4) are shown in Figure 6. It is not until a deletion radius R of 2000 meters that the inference rate for home addresses dropped to zero. Changing the size of the smaller circle, radius r, did not have a noticeable effect.

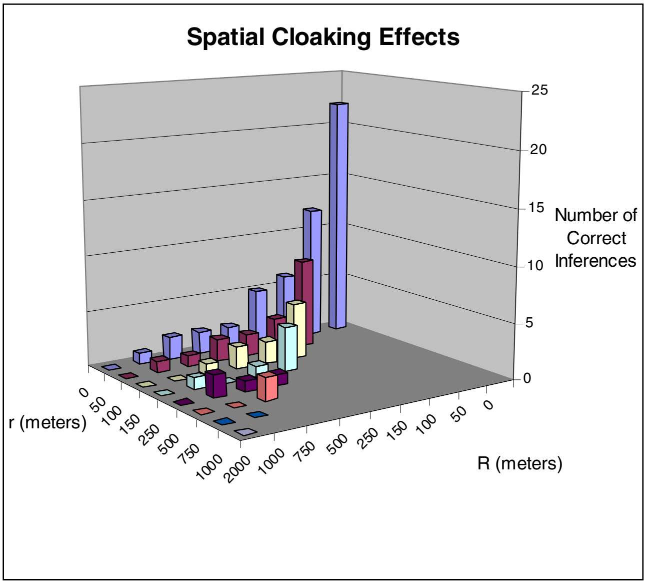


Fig. 6. Based on 172 subjects, spatial cloaking was not 100% effective until all data within 2000 meters of the home was deleted

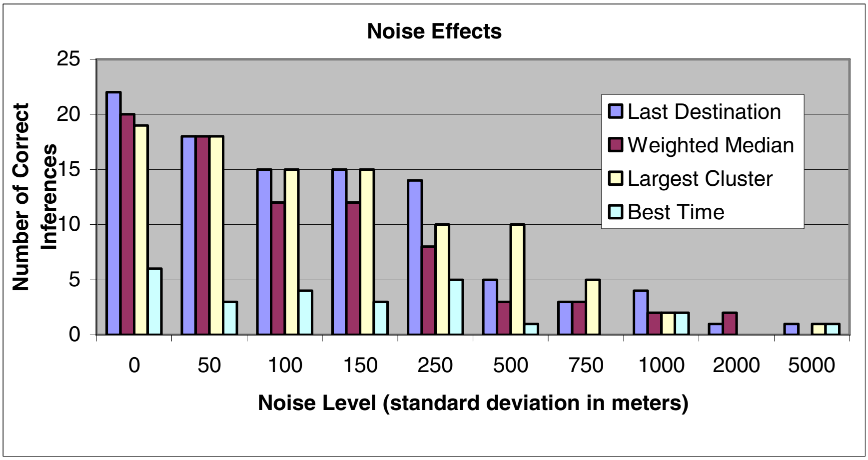


Fig. 7. Adding noise with a standard deviation of five kilometers reduced the effect of inference attacks to near zero

Noise – The number of correct address inferences as a function of σ is shown in Figure 7. As expected, the number of correct addresses found decreases with increasing noise, although the amount of noise required for a significant reduction in performance is perhaps surprising. Noise with a standard deviation of 5 kilometers reduced the number of found addresses to only one out of 172 for three of the inference algorithms and to zero for "Largest Cluster".

Rounding – Coarser discretization reduced the number of correct address inferences, which dropped to zero for all algorithms with a Δ of 5 kilometers, as shown in Figure 8.

None of the home-finding algorithms stood out as uniquely robust at resisting the countermeasures. Likewise, none of the countermeasures proved uniquely effective at resisting a privacy attack. The exact points at which the countermeasures become effective (*e.g.* noise level, spatial coarseness) likely do not generalize well to other types of location data due to variations in the time between measurements, distance between measurements, the sensor's intrinsic accuracy and precision, and the density of homes. From a qualitative perspective, however, the level of corruption required to completely defeat the best inference algorithms is somewhat high.

In choosing a countermeasure, it would be important to assess not only its effectiveness but the effect on the intended application. For instance, a traffic-monitoring application may be relatively unaffected by cloaking a few hundred meters around drivers' homes, because the application uses only aggregate statistics from multiple drivers, and because road speeds in residential neighborhoods are relatively unimportant. On the other hand, noise or rounding could easily overpower the map matching techniques that the traffic application would use to match GPS

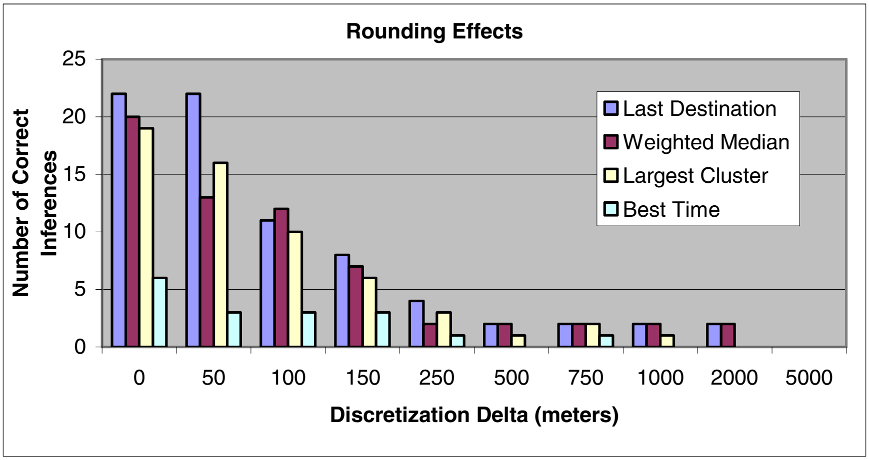


Fig. 8. Discretizing latitude and longitude to a five-kilometer grid was enough to protect all addresses from our inference attacks

points to actual roads. Similar arguments apply to the applications of making customized driving routes or travelogues, because cloaking the home location would have only a minor effect compared to noise or rounding.

One unanswered question is the point at which a countermeasure becomes practically effective. Even if a few addresses or identities can be compromised, the attacker may not have any way to determine which inferences are correct. However, more sophisticated attacks could estimate their own uncertainty, highlighting which inferences are most likely correct.

5 Summary

This is the first paper we know of to make a thorough, experimental assessment of computational inference attacks on recorded location data, as well as the effectiveness of certain countermeasures. We showed that it is possible, using simple algorithms and a free Web service, to identify people based on their pseudonymous location tracks. Using GPS data from 172 subjects, we can find each person's home location with a median error of about 60 meters. Submitting these locations to a reverse white pages lookup, we were able to correctly identify about 5% by name. This number is low partly due to our inaccuracy in finding the home locations, but also due to the vagaries of Web-based white pages. If we tried to identify only the home addresses, our accuracy rose to almost 13% using a commercially available reverse geocoder. Both the white pages and the reverse geocoder proved to be weak links in the inference attack, but both these technologies will improve as their benevolent applications become more important.

We tested three different countermeasures: spatial cloaking, noise, and rounding. We quantified their effectiveness by how well they prevented our inference algorithms from finding the subjects' home addresses. Our results show how much the location data needs to be corrupted to preserve privacy. The best of our home-finding algorithms proved somewhat robust to these techniques. The high degree of corruption required when using noise or rounding means that several location-based services could become unusable.

Future work on this problem should expand the experimental matrix with additional attack algorithms and countermeasures. It would be useful to create a quantitative assessment of the effect of the countermeasures on benevolent applications as well as on the attack algorithms as a guide to privacy policymakers.

References

1. Agrawal, R. and R. Srikant. *Privacy-Preserving Data Mining*. in *ACM SIGMOD Conference on Management of Data*. 2000. Dallas, TX, USA: ACM Press.
2. Ashbrook, D. and T. Starner, *Using GPS to Learn Significant Locations and Predict Movement across Multiple Users.* Personal and Ubiquitous Computing, 2003. **7**(5): p. 275-286.
3. Barbaro, M. and T.Z. Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, in *New York Times*. 2006: New York, NY USA.
4. Beresford, A.R. and F. Stajano, *Location Privacy in Pervasive Computing*, in *IEEE Pervasive Computing*. 2003. p. 46-55.
5. Consolvo, S., et al. *Location Disclosure to Social Relations: Why, When, & What People Want to Share*. in *CHI '05: Proceedings of the SIGCHI Conference on Human factors in Computing Systems*. 2005. New York, NY.
6. Danezis, G., S. Lewis, and R. Anderson. *How Much is Location Privacy Worth?* in *Fourth Workshop on the Economics of Information Security*. 2005. Harvard University.
7. Duckham, M. and L. Kulik, *Location Privacy and Location-Aware Computing*, in *Dynamic & Mobile GIS: Investigating Change in Space and Time*, J. Drummond, et al., Editors. 2006, CRC Press: Boca Raton, FL.
8. Gendar, A. and A. Lisberg, *How Cell Phone Helped Cops Nail Key Murder Suspect*, in *Daily News*. 2006: New York.
9. Gruteser, M. and D. Grunwald. *Anonymous Usage of Location-Based Services through Spatial and Temporal Cloaking*. in *of First ACM/USENIX International Conference on Mobile Systems, Applications, and Services (MobiSys 2003).* 2003. San Francisco, CA USA.
10. Hariharan, R., J. Krumm, and E. Horvitz. *Web-Enhanced GPS*. in *First International Workshop on Location- and Context-Awareness (LoCA 2005)*. 2005. Oberpfaffenhofen, Germany: Springer.
11. Hariharan, R. and K. Toyama. *Project Lachesis: Parsing and Modeling Location Histories*. in *Third International Conference on GIScience*. 2004. Adelphi, MD.
12. Hightower, J., et al. *Learning and Recognizing the Places We Go*. in *UbiComp 2005: Ubiquitous Computing*. 2005.
13. Hoh, B., et al., *Enhancing Security and Privacy in Traffic-Monitoring Systems*, in *IEEE Pervasive Computing*. 2006. p. 38-46.
14. http://www.microsoft.com/downloads/details.aspx?FamilyId=C271309B-02DE-42A7-B23E-E19F68667197&displaylang=en.
15. http://www.openstreetmap.org/.

16. Iachello, G., et al. *Control, Deception, and Communication: Evaluating the Deployment of a Location-Enhanced Messaging Service*. in *UbiComp 2005: Ubiquitous Computing*. 2005. Tokyo, Japan.
17. Kaasinen, E., *User Needs for Location-Aware Mobile Services.* Personal and Ubiquitous Computing, 2003. **7**(1): p. 70-79.
18. Kang, J.H., et al. *Extracting Places from Traces of Locations*. in *2nd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH'04),*. 2004.
19. Krumm, J. and E. Horvitz. *Predestination:Inferring Destinations from Partial Trajectories*. in *Eighth International Conference on Ubiquitous Computing (UbiComp 2006)*. 2006. Orange County, CA USA.
20. Letchner, J., J. Krumm, and E. Horvitz. *Trip Router with Individualized Preferences (TRIP): Incorporating Personalization into Route Planning*. in *Eighteenth Conference on Innovative Applications of Artificial Intelligence (IAAI-06)*. 2006. Cambridge, MA USA.
21. Liao, L., D. Fox, and H.A. Kautz. *Location-Based Activity Recognition using Relational Markov Networks*. in *Nineteenth International Joint Conference on Artificial Intelligence (IJCAI 2005)*. 2005. Edinburgh, Scotland.
22. Lorkowski, S., et al. *Towards Area-Wide Traffic Monitoring-Applications Derived from Probe Vehicle Data*. in *8th International Conference on Applications of Advanced Technologies in Transportation Engineering*. 2004. Beijing, China.
23. Marmasse, N., *Providing Lightweight Telepresence in Mobile Communication to Enhance Collaborative Living*, in *Program in Media Arts and Sciences, School of Architecture and Planning*. 2004, MIT: Cambridge, MA, USA. p. 124.
24. Marmasse, N. and C. Schmandt. *Location-Aware Information Delivery with comMotion*. in *HUC 2K: 2nd International Symposium on Handheld and Ubiquitous Computing*. 2000. Bristol, UK: Springer.
25. Orland, K., *Stalker Victims Should Check For GPS*. 2003, Associated Press on CBS News at http://www.cbsnews.com/stories/2003/02/06/tech/main539596.shtml.
26. Patterson, D.J., et al. *Opportunity Knocks: A System to Provide Cognitive Assistance with Transportation Services*. in *UbiComp 2004: Ubiquitous Computing*. 2004. Nottingham, UK: Springer.
27. Pfitzmann, A. and M. Kohntopp, *Anonymity, Unobservability and Pseudonymity -- A Proposal for Terminology*, in *Designing Privacy Enhancing Technologies*, H. Federath, Editor. 2001, Springer-Verlag. p. 1-9.
28. Schilit, B., J. Hong, and M. Gruteser, *Wireless Location Privacy Protection*, in *IEEE Computer*. 2003. p. 135-137.
29. Scott, D., A. Beresford, and A. Mycroft. *Spatial Security Policies for Mobile Agents in a Sentient Computing Environment*. in *6th Fundamental Approaches to Software Engineering (FASE)*. 2003. Warsaw, Poland: Springer-Verlag.
30. Subramanya, A., et al. *Recognizing Activities and Spatial Context Using Wearable Sensors*. in *21st Conference on Uncertainty in Artificial Intelligence (UAI06)*. 2006. Cambridge, MA.

Shake Well Before Use: Authentication Based on Accelerometer Data

Rene Mayrhofer and Hans Gellersen

Lancaster University, Computing Department, South Drive, Lancaster LA1 4WA, UK
`{rene,hwg}@comp.lancs.ac.uk`

Abstract. Small, mobile devices without user interfaces, such as Bluetooth headsets, often need to communicate securely over wireless networks. Active attacks can only be prevented by authenticating wireless communication, which is problematic when devices do not have any a priori information about each other. We introduce a new method for device-to-device authentication by shaking devices together. This paper describes two protocols for combining cryptographic authentication techniques with known methods of accelerometer data analysis to the effect of generating authenticated, secret keys. The protocols differ in their design, one being more conservative from a security point of view, while the other allows more dynamic interactions. Three experiments are used to optimize and validate our proposed authentication method.

1 Introduction

Applications envisioned for ubiquitous computing build upon spontaneous interaction of devices, such that a device can make serendipitous use of the services provided by peer devices that may not be known a priori. In many scenarios, it will be desirable to verify and secure spontaneous interactions in order to ascertain that devices become paired as intended and protected against attacks on their wireless link. In a managed network environment, device-to-device authentication would be based on prior knowledge of each other or access to a trusted third party, but neither can be assumed to be available in wireless ad hoc networks for ubiquitous computing. As a consequence, secure device pairing requires the user to be in the loop, for example to enter a shared secret such as a PIN code into both devices. A challenge is to find mechanisms for users to pair devices that are not only secure but also scale well for use in ubiquitous computing. Specific challenges are that devices will, in many cases, be too small to reasonably include key pads and displays, and that required user attention must be minimal to be acceptable for spontaneous and short-lived interactions.

Pairing of a mobile phone with a headset for interaction over a wireless channel is a familiar example: we would like to achieve such interaction in a spontaneous manner (i.e. not requiring pre-configuration of phone and headset for each other) but also ensure that it is secure. The wireless communication channel between the devices is susceptible to attacks ranging from eavesdropping to

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 144–161, 2007.

man-in-the-middle (MITM). If an attacker were successful in establishing themselves between, in this case, phone and headset, during the pairing process, then they would obtain complete control over all phone calls. To safeguard against such attacks, a so-called *out-of-band channel* is used during pairing in order to authenticate communication over the primary channel. The out-of-band channel must be limited such that it is user-controllable that only the intended devices can communicate over it for the purposes of authentication. Note that authentication and the subsequent pairing can be anonymous or "ephemeral" [1], i.e. based on information only shared over the out-of-band-channel rather than actual device identities.

In this paper we contribute a method for device-to-device authentication that is based on shared movement patterns which a user can simply generate by shaking devices together. Using embedded accelerometers, devices can recognize correlation of their movement and use movement patterns for authentication. From a user perspective, jointly shaking is a simple technique for associating devices [2]. In our method, it simultaneously serves as out-of-band mechanism. Shaking has a number of characteristics on which we can build for our purposes:

- It is *intuitive*. People are familiar with shaking objects as manual interaction that does not require learning, for instance from shaking of medicine, or musical instruments. This means that shaking is unobtrusive in the sense that it does not require the user's full attention while being performed.
- It is *vigorous*. While there are many motion patterns that could be performed with two devices, shaking tends to produce the highest continuous acceleration values. While bouncing will produce larger accelerations, they only occur as short spikes. Shaking provides acceleration larger than most activities – and can thus be detected by simple thresholding – for as long as necessary to pair devices (and as long as the user will not get tired).
- It is *varying*. As we will show below in our first experiment (in section 7.1), the activity of shaking can be surprisingly different for different people. We do not use shaking patterns as identification, but still benefit from large differences in acceleration values, because this generates high entropy from an attacker's point of view.

It is important to note that users do not have to follow a particular pattern of shaking but that they can shake as they like; we do not attempt to identify people by their shaking patterns, but use it as a source of shared device movement.

We contribute two protocols that combine cryptographic primitives with accelerometer data analysis to establish secure wireless channels by creating authenticated secret keys. The two protocols achieve this aim differently: the first is based on Diffie-Hellman key agreement and authentication of this key, uses a conservative and better known design, provides better security and allows more flexibility in comparing accelerometer time series; the second generates cryptographic key material directly out of accelerometer data streams, is computationally less expensive and thus easier to implement on resource limited devices, and allows more dynamic interactions and group authentication.

Both protocols use standard techniques of sensor data processing and time series analysis: sampling, alignment, and feature extraction. After extracting appropriate features, our cryptographic protocols ensure that authentication is only possible if both devices have access to the same feature values. Specifically, they protect against MITM attacks on the wireless communication channel by using additional information gathered from the extracted features. This approach is general, so that other sensors than accelerometers can be used with similar methods, apart from changes in domain-specific heuristics. Sensor-based authentication offers potential benefits to small, mobile devices that communicate wirelessly and do not have traditional user interfaces. Examples are mobile phones, smarts cards, key fobs, and generally accessories like headsets, watches, or glasses.

2 Related Work

First concepts on secure device pairing suggested direct electrical contact [3], while other suggestions to implement an out-of-band channel include a “physical interlock” and the “Harmony” protocol [4], ultrasound [5], visual markers and cameras [6], audio messages [7], the GSM short message service (SMS) [8], key comparison, distance bounding and integrity codes [9], or manual input [10,1]. The DH-DB protocol proposed in [9] might also be applicable to an interactive challenge-response scheme based on sensor data such as accelerometer data. These approaches, with the exception of using camera phones, have in common that they scale poorly from a user point of view. That is, they tend to be obtrusive and require the user’s attention. In our approach, we implement a low bandwidth private channel over similar accelerometer readings, and use it for authenticating a device pairing.

The idea of shaking two (or multiple) devices together to pair them has first been described as “Smart-Its Friends” [2]. We use the same interaction technique but extend it to include secure authentication. Castelluccia and Mutaf presented a protocol for pairing CPU-constrained wireless devices under the assumption of anonymous broadcast channels [11]. To achieve this property of source indistinguishability, they argue that devices engaging in this authentication protocol should be shaken and rotated randomly around each other. This shaking serves to prevent signal strength analysis, but is, in contrast to our work, not used directly as input to the authentication protocol. Hinckley presented an implementation of “synchronous gestures” [12] as a means of user interaction. By correlating accelerometer time series on devices connected via WLAN, bumping them together or tilting them can be detected and used as user input. Bumping is one possible user interaction for starting the pairing process, i.e. a trigger for our authentication method. Another closely related work was presented by Lester et al. [13] and describes how to determine if two devices are carried by the same person.

3 Design of the Acceleration-Based Pairing Method

Figure 1 shows our architecture for authenticating device pairings with shaking patterns. Both protocols make use of the same three pre-processing tasks 1 to 3. They are executed locally on each device and result in “active” time series segments of equidistant samples. Our two protocols differ in tasks 4 and 5, which can both be interactive, i.e. communicate with the remote device to which the pairing is in process.

For protocol 1, tasks 4.1 and 5.1 are actually executed in parallel: after generating a secret key with standard Diffie-Hellman (DH) key agreement (which is the first phase of task 5.1), the devices exchange their time series segments via an interlock protocol. Then they compare their locally generated segment with the one received from the remote device to check if they are similar enough. If they pass this check, the second phase of task 5.1 derives the secret session key that will be used for consecutive secure communication. This design is conservative from a security point of view and, due to the non-interactive feature extraction and comparison, allows the devices to use different means of verification. The disadvantage of splitting task 5.1 into two phases is potentially a larger delay for authentication, and the disadvantage of using DH is higher computational load.

Protocol 2 executes its tasks 4.2 and 5.2 in order: discrete (in contrast to the real-valued samples) feature vectors are extracted in task 4.2, which act as input to the interactive key agreement in task 5.2. This is an iterative process. In each time step, feature vectors generated by 4.2 are checked for matches in task 5.2. After sufficient iterations, a secret shared key can be generated out of the collected matching feature vectors in task 5.2. This design has the advantages of more dynamic key agreement, with devices being able to “tune into” other device’s key streams, and of being less computationally expensive. On the other hand, it does not provide forward secrecy and protection against offline attacks as protocol 1 does, and is more unconventional and thus less well studied from a security point of view.

For both protocols, there is a trade-off between usability and security that can be exploited by applications and users depending on their requirements. Tasks 4 and 5 are described in more detail in sections 5 and 6, respectively.

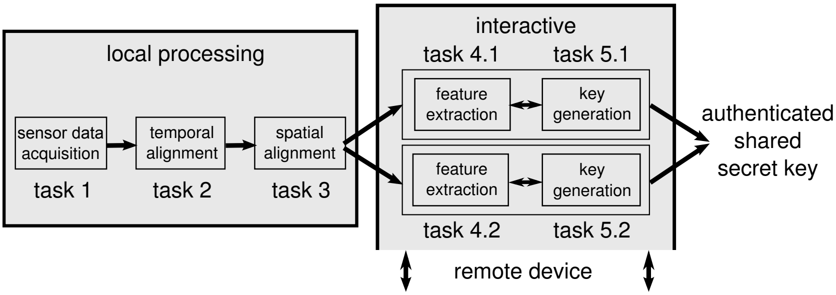


Fig. 1. Architecture for both authentication protocols

4 Pre-processing of Accelerometer Data

The three pre-processing tasks, executed as consecutive steps, are used to sample and segment the sensor data so that feature extraction can build on normalized time series.

Task 1: Sensor data acquisition. This first task is conceptually straight forward, but requires careful implementation. Sensor data is assumed to be available in the form of time series of acceleration values in all three dimensions, sampled at equidistant time steps. These must be taken locally and not be communicated wirelessly — for security purposes, it is critical not to leak any of this raw data, which can be difficult considering the possibility of powerful side-channel attacks (see e.g. [14]). Our practical experience shows a sample rate between 100 and 600 Hz to be appropriate.

Task 2: Temporal alignment. As the two devices sample accelerometer time series independently in task 1, we require temporal synchronization for comparison. We assume that devices are equipped with sufficiently accurate real-time clocks, so that differences in sampling rates and drift will not be issues. This reduces temporal alignment from an arbitrarily complex problem to *triggering* the authentication procedure and to *synchronizing* the starting points for time series comparison.

Triggering can be *explicit* by direct user input, e.g. pressing an "authenticate now" button on both devices within a short time frame or bumping both devices against the table or each other, or *implicit*, simply by starting to shake both devices. We prefer the second protocol due to its ease of use, although it is more difficult to implement. Synchronization can be at a *sample level*, i.e. within less than half the sample width, or at an *event level*, i.e. based on the onset of detected (explicit or implicit) events with the respective device. We use the latter, because it does not require time synchronization between the devices — shaking events can be detected locally at each device without communication, which is beneficial from a security point of view.

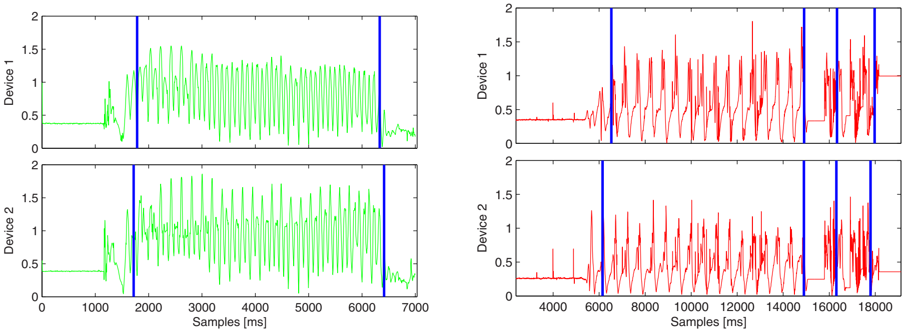
For both triggering and synchronization, we detect motion and align those parts of the time series where shaking is detected, which we call *active segments*, by their start times. Segments are considered active when the variance of a sliding window exceeds a threshold. Practical experiments show good results at a sample rate between $f = [128; 512]$ Hz with a sliding window of $v = f/2$ samples, i.e. 1/2 second, and a variance threshold around $T_\sigma = 750$.

Task 3: Spatial alignment. Shaking is inherently a three-dimensional movement. In addition to the need to capture all three dimensions, the alignment between the two devices is unknown. This means that the three dimensions recorded by the two devices will not be aligned, which is a hard problem in itself. Lukowicz et al. describe how to calibrate three-dimensional accelerometers without user interaction during stable periods [15]. However, since we are interested in the active phases and have to assume that the alignment of the devices changes during the transition,[1] we can not directly apply this result. Instead, we reduce the three dimensions to a single: by taking only the magnitude over all normalized dimensions, i.e. the length of the vector, we solve the alignment problem. This approach requires considerably less resources than other methods such as principal component analysis (PCA) or modeling using domain-specific knowledge.

The result of these steps is that, when shaken together, both devices will extract active segments of one-dimensional acceleration magnitude vectors. Even without synchronized clocks, the start times of these independent time series are typically synchronized within a few samples (on the event level).

5 Feature Extraction for Authentication Purposes

Two devices that are shaken together will experience similar, but not exactly the same movement patterns. Even assuming noise-free sampling of accelerations, the two accelerometers must have physically separate centers. Whenever rotation is part of the movement, these separate centers will necessarily experience different accelerations, thus causing different sensor time series even if the devices remain fixed in relation to each other. The problem of verifying that two devices are shaken, or more generally, moved together therefore becomes a classification problem. Figure 2 shows examples of spatially aligned sensor time series used as input to feature extraction with detected borders of active segments.



(a) Two devices shaken by one person in the same hand

(b) Two devices shaken by two people, one each

Fig. 2. Example time series after spatial alignment with detected active segments

In deciding if time series are similar enough for authentication, the aim of the feature extraction task is twofold: a) to extract feature values that are robust to small variations in the shaking patterns and to sampling noise and b) to extract

---

[1] When a user picks up the two devices to shake them, they will most probably be aligned differently in their hand than they were before picking them up.

a sufficiently large feature vector for use in the authentication protocol. In our approach, the feature vector will be used to authenticate a key or to directly generate a key, and thus it needs to be of high entropy from an attacker's point of view, i.e. involve a large amount of uncertainty.[2] As indicated in section 1, we argue that shaking is an appropriate movement for creating entropy: it creates varying sensor readings, because it is one of the human movement patterns that includes the highest frequency components. Slower movements will intuitively not generate as much entropy.

There is an extensive body of literature on feature extraction from accelerometer data. Particularly relevant to our problem are the described uses of the coherence measure by Lester et al. [13] and cross-covariance by Aylward et al. [16]. Both suggest sliding, windowed variances on each device for activity detection, as used in our current implementation for task 2. Huynh and Schiele compare different features for activity recognition and suggest the use of quantized FFT coefficients [17]. For task 4, we select the most promising of the recently suggested features: coherence and quantized FFT coefficients.

5.1 Coherence

We adopt the approach that was previously used by Lester et al. to distinguish between two devices worn by the same person (on different parts of the body) and two devices worn by two people walking in-step. They used coherence averages and showed that simple, non-calibrated, cheap accelerometers are suitable for analyzing human motion. Coherence is approximated by the magnitude squared coherence (MSC) as

$$C_{xy}\left(f\right)=\frac{P_{xy}\left(f\right)}{P_{xx}\left(f\right)\cdot P_{yy}\left(f\right)}$$

with (cross-) power spectra

$$P_{xy}\left(f\right)=\frac{1}{n}\sum_{k=0}^{n-1}x_k\left(f\right)\cdot\bar{y_k}\left(f\right)$$

computed over FFT coefficients $x_k\left(f\right) = FFT\left(a_k\left(t\right)\cdot h\left(t\right)\right)$ and $y_k\left(f\right) = FFT\left(b_k\left(t\right)\cdot h\left(t\right)\right)$ using the standard von-Hann window $h\left(t\right) = \frac{1-\cos(2\pi t/w)}{2}$ That is, it is computed as the power spectrum correlation between two signals split into n (optionally overlapping) averaged slices a_k and b_k of the signals a and b, respectively, normalized by the signal power spectra. Note that, although the signals a and b in time domain are real, their FFT coefficients x and y are complex. By using squared magnitudes, C_{xy} is also real-valued. By $\bar{x}$ we refer to the conjugate complex of x. Because the significance of coherence values depends on the number of averaged slices n – the more slices, the lower the coherence

---

[2] The authentication protocol is said to be computationally secure if an attacker's entropy of the key approaches the key length, which is typically 128 bits.

values are for the same signals –, we reduce longer time series to a maximum length of 3 seconds. This is a compromise between sufficient variability for robust classification and quick user interaction. The final value is computed simply by averaging up to a cut-off frequency f_{max}

$$C_{xy} = \frac{1}{f_{max}} \int_{0}^{f_{max}} C_{xy}(f)\, df$$

With this heuristic, we threshold C_{xy} to create a binary decision of similarity for our authentication protocol. As explained below, our experiments have shown that, with a sampling rate of $r = 256\,\text{Hz}$ and windows of $w = 256$ samples with an overlap of 7/8 and a cut-off frequency of $f_{max} = 40\,\text{Hz}$, coherence provides good distinction between two devices being shaken by one person from two devices being shaken by two people, one each.

5.2 Quantized FFT Coefficients

Coherence is a powerful measure of similarity, but, due to its use of continuous values, does not lend itself to directly creating cryptographic key material out of its results. Keys must be bit-for-bit equal, and thus be based on discrete instead of continuous values. By retaining basic features of the coherence measure and condensing them into discrete feature vectors, we can use those for a different way of comparing two accelerometer time series. Coherence is based on FFT coefficients, so it seems logical to quantize them into discrete values.

Huynh and Schiele compared different features with different window sizes and found that pairwise adding of neighboring FFT coefficients and grouping into exponential bands performed best in recognizing activities with moderate to high intensity levels, while other features like pairwise correlation or spectral energy were worse [17]. They also reported that the highest FFT peaks could generally be found up to the tenth coefficient, which backs our own findings that coefficients above 20 Hz do not contribute significantly.

We compared four variants of FFT-based feature vectors: linearly or exponentially quantized coefficients used either directly of added pairwise. Our experiments have shown that pairwise added, exponentially quantized FFT coefficients performed best, as also suggested in [17]. When aiming for equivalence of feature vectors, there is however an additional complication: small differences of values near the boundaries of quantization bands can lead to different feature values, although the FFT coefficients are only marginally different. Our solution is to quantize each FFT vector into multiple *candidate* feature vectors with different offsets. These offsets range from 0 to the value of the smallest quantization band. The similarity criteria in this case is simply the percentage of matching candidate feature vectors out of all vectors sent to another device. Thresholding this percentage produces a binary decision for the authentication protocol. We achieved best results for distinguishing shaking by one person from shaking by two people, one device each, with $b = 6$ exponentially scaled bands for quantization, $k = 4$ candidates, and a cut-off frequency of $f_{max} = 20\,\text{Hz}$ at a sampling rate of $r = 512\,\text{Hz}$ with FFT windows of $w = 512$ samples, overlapping by 50%.

6 Authentication Protocols

The two feature vectors generated in task 4 constitute, if equivalent, a shared secret password. This shared string is not directly suitable to act as a secret key for cryptographic primitives, because it is neither of defined length (e.g. 128 bits) nor distributed uniformly. But it is possible to create a cryptographically secure secret key via interactive protocols, authenticated by the feature vectors.

The choice of features directly influences requirements on the cryptographic protocols. To compute the coherence measure, both vectors need to be available completely to both devices.[3] Therefore, the time series must be exchanged during the interactive protocol — in a way that does not reveal them to an attacker. Our first authentication protocol uses asymmetric cryptography to achieve this.

Feature vectors composed of quantized FFT coefficients, on the other hand, do not allow for additional differences — authentication should only proceed if both vectors are bit-for-bit equal. The advantage is that cryptographic key material can be created using only symmetric cryptography, which is more suitable for embedded devices.

For the formal descriptions of our protocols, we use the following notation: $c = E(K, m)$ describes the encryption of plain text m under key K with a symmetric cipher, $m = D(K, c)$ the corresponding decryption, $H(m)$ describes the hashing of message m with some secure hash, and $m|n$ the concatenation of strings m and n. The notation $M[a : b]$ is used to describe the substring of a message M starting at bit a and ending at bit b. The symbol $\oplus$ describes bit-wise XOR and $|S|$ the number of elements in a set S. If a message M is transmitted over an insecure channel, we denote the received message $\widetilde{M}$ to point out that it may have been modified in transit, by noise or attack. C refers to some publicly known constant. We use AES as a block cipher for E and D and $\text{SHA}_{\text{DBL}}\text{-256}$ as a secure hash for H, which is a double execution of the standard SHA-256 message digest to safeguard against length extension and partial-message collision attacks [18] and is defined as $\text{SHA}_{\text{DBL}}\text{-256} = \text{SHA-256}\left(\left(\text{SHA-256}\left(m\right)\right)|m\right)$.

6.1 Protocol 1: Diffie-Hellman and Interlock*

Fig. 3 shows our first authentication protocol, which is based on a standard Diffie-Hellman (DH) key agreement (introduced in their seminal article [19]) followed by an exchange of the condensed time series and comparison locally at each device.

Using DH key agreement, devices A and B generate two – supposedly – shared keys K^{Auth} and K^{Sess}, where it is impossible to infer one from the other (under the assumption that the hash function does not allow to find a pre-image). Creating two keys, one for authentication, one as session key, provides forward secrecy. Because DH is susceptible to MITM, the devices need to verify that their keys are equivalent. The unique key property of DH guarantees with a

---

[3] For security reasons, both devices should independently decide if authentication was successful, and thus both need to compute the coherence.

A	Message	**B**
	Choose DH key	
$x \in \{1 \ldots q-1\}$ $K_a^{Sess} :=$ $H(Y^x)$ $K_a^{Auth} :=$ $H(Y^x \vert C)$	$\xrightarrow{X := g^x}$ $\xleftarrow{Y := g^y}$	$y \in \{1 \ldots q-1\}$ $K_b^{Sess} :=$ $H(X^y)$ $K_b^{Auth} :=$ $H(X^y \vert C)$
	Split and encrypt signal	
$(i = 0 : n-1, \quad j = 1 : n)$: $a_i := a[128i : 128i + 127]$ $c_0 \in \{0, \ldots, 2^{128} - 1\}$ $c_j := E(K_a^{Auth}, c_{j-1} \oplus a_{j-1})$	$\xrightarrow{A_1 = c_0[0:63] \Vert \ldots \vert c_n[0:63]}$ $\xleftarrow{B_1 = d_0[0:63] \Vert \ldots \vert d_m[0:63]}$	$(i = 0 : m-1, \quad j = 1 : m)$: $b_i := b[128i : 128i + 127]$ $d_0 \in \{0, \ldots, 2^{128} - 1\}$ $d_j := E(K_b^{Auth}, d_{j-1} \oplus b_{j-1})$
	Interlock and reassemable	
$(i = 0 : m-1)$: $\tilde{d}_i := \tilde{B}_1[128i : 128i + 63] \Vert$ $\tilde{B}_2[128i : 128i + 63]$ $\tilde{b}_i := D(K_a^{Auth}, \tilde{d}_{i+1}) \oplus \tilde{d}_i$ $\tilde{b} := \tilde{b}_0 \vert \ldots \vert \tilde{b}_m$	$\xrightarrow{A_2 = c_0[64:127] \Vert \ldots \vert c_n[64:127]}$ $\xleftarrow{B_2 = d_0[64:127] \Vert \ldots \vert d_n[64:127]}$	$(i = 0 : n-1)$: $\tilde{c}_i := \tilde{A}_1[128i : 128i + 63] \Vert$ $\tilde{A}_2[128i : 128i + 63]$ $\tilde{a}_i := D(K_b^{Auth}, \tilde{c}_{i+1}) \oplus \tilde{c}_i$ $\tilde{a} := \tilde{a}_0 \vert \ldots \vert \tilde{a}_n$
	Compare	
if $a \sim \tilde{b}$ then $K := K_a^{Sess}$ else $K := null$		if $b \sim \tilde{a}$ then $K := K_b^{Sess}$ else $K := null$

Fig. 3. Protocol 1: Diffie-Hellman key agreement followed by exchange of the complete time series via interlock*

very high probability, that, if $K_a^{Auth} = K_b^{Auth}$, there can be no attacker E with $K_{e1}^{Auth} = K_a^{Auth}$ and $K_{e2}^{Auth} = K_b^{Auth}$, and subsequently, no $K_{e1}^{Sess} = K_a^{Sess}$ and $K_{e2}^{Sess} = K_b^{Sess}$.

This verification is done with an extended *interlock* protocol. Interlock [20] is not used widely, but is an efficient (in terms of message length) method to verify that two parties share the same key. By using this key as an input to a block cipher and splitting packets in halves, a MITM can only decrypt these packets after having received both halves. The interlock protocol then demands that A and B will only send their second halves after they have received the first halves from the respective other side. This has the effect that both sides must commit themselves to their values, by sending the first halves of the encrypted blocks, before they can receive, and subsequently decrypt, the other side's message. Thus, interlock can be seen as a commitment scheme (see e.g. [21] for a definition) based on block ciphers. An attacker E is now left with only two options: either to forward the original packets, or to create packets on its own. In the former case, A and B will be unable to decrypt the messages properly, because they do not share the same key. In the latter case, E must guess the contents of the messages, and encrypt them with the appropriate keys, before it has access to the actual messages. When the messages sent by A and B have an entropy of e bits. this leaves E with a single 2^{-e} chance of remaining undetected.

The original version of interlock is suitable for messages the size of the cipher block length. Because in our case the vectors of the accelerometer sensor data, condensed into a time series of magnitudes, have arbitrary length, we introduce a slightly extended protocol that we call *interlock**. In this variant, A and B encrypt their complete messages, i.e. the (zero-padded) vectors a and b with lengths of n and m blocks, respectively, with any of the well-known block cipher modes. For our motion authentication protocol, we simply use the cipher block chaining (CBC) mode with a random initialization vector (IV). The resulting cipher texts c and d with lengths of $n+1$ and $m+1$ blocks are then split into two messages by concatenating the first halves of all cipher blocks into the first messages A_1 and B_1 and the second halves of all cipher blocks into the second messages A_2 and B_2. This ensures that E can not decrypt any of the blocks, and can therefore not even learn parts of the plain text messages.

After exchanging their messages a and b, A and B verify that $a \sim b$, that is, that they are similar enough under their chosen criteria. We use coherence as described in section 5, but other suitable features can be used without changes to the protocol. Because of this possibility, we do not try to minimize the message lengths as e.g. suggested in [13]. In fact, A and B could use completely different similarity criteria, and could still authenticate using the same protocol. This is important for practical implementations, because different generations of devices will need to be compatible with each other.

The MANA III scheme described in [10] serves a similar purpose as this protocol, but using different cryptographic primitives. While we employ a block cipher, the MANA III scheme uses a MAC. Both constructions build on a mutual commitment to an authenticator string before transmitting parts of it.

6.2 Protocol 2: Candidate Key Protocol

In our second protocol, which we call the *candidate key protocol* (CKP), the shared secret key is generated from sensor data instead of by DH. As depicted in Fig. 4, feature vectors v are hashed to generate *candidate key parts* h. If the feature extraction task produces multiple "parallel" feature vectors v^i for each time window, as suggested above in section 5, then these yield multiple candidate

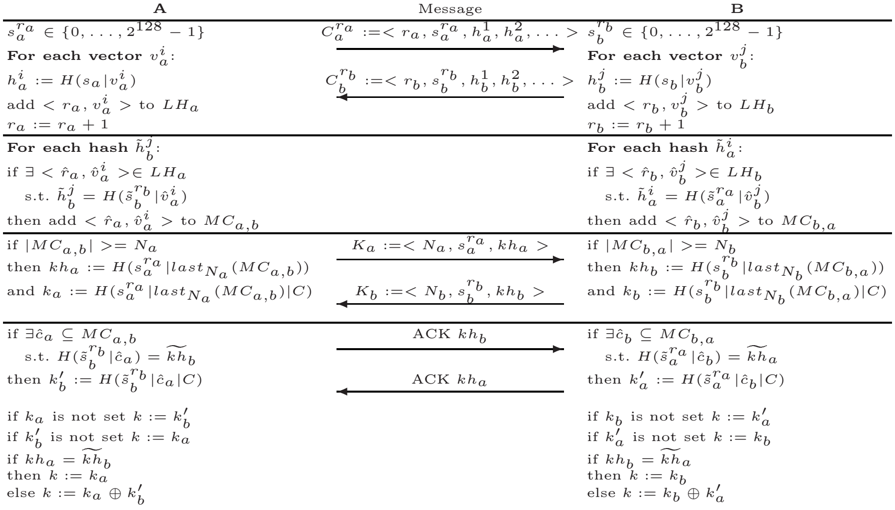


Fig. 4. Protocol 2: candidate key protocol for directly creating a secret key from common feature vector hashes

key parts h^i. The one-way hashes are a simple way to communicate that a device has generated a certain feature vector without revealing it. To make dictionary attacks harder, we use the standard method of prepending random salt values s before hashing. When B receives such candidate key parts from A, it can check its own history of recently generated feature vectors LH to check for equals. When B has generated the same feature vector, it is stored in a list of *matching key parts* MC specific to each communication partner. As soon as enough entropy has been collected in this list, B concatenates all feature vectors, appends C, hashes the resulting string, and sends a *candidate key* K to A. If no messages have been lost in transit, A should be able to generate a key with the same hash, and thus the same secret key, which it acknowledges to B. If messages have been lost, A can simply ignore a candidate key and create its own later on.

CKP is again a general protocol and can be used with any feature vectors. Here we apply it to quantized FFT coefficients, which work well for accelerometer data. A more thorough analysis of CKP itself will be provided separately.

7 Experimental Evaluation

We conducted three experiments, two to optimize parameters for the feature extraction tasks described in section 5, and one to validate our assumption of ease of use. All three experiments used four simple ADXL202JE accelerometers, two on each device, mounted at an angle of 90° so that all three dimensions could be measured with a maximum acceleration of 2 g. The accelerometers are fixed with compressed foam inside ping-pong balls (see Fig 5), and sampled at roughly 600 Hz. By choosing balls as “device” shapes and orienting the accelerometers randomly inside the balls, each data set has different orientations. The subjects were also asked to pick the devices up at the start of each sample, so that orientations change between samples. Although the accelerometers were wired to enable higher sampling rates, the attached cables were lightweight, flexible, and long enough so as not to disturb movements of subjects.

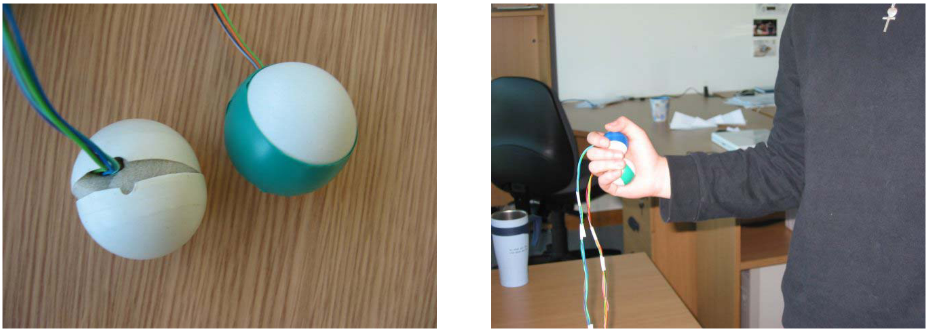


Fig. 5. Experimental setup: devices with accelerometers and subject during data collection

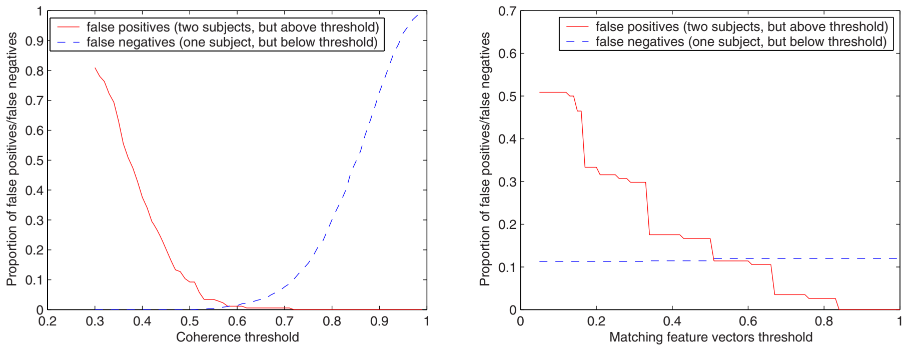
7.1 Experiment 1: Single Subjects Data Collection

The first experiment was explorative and aimed to discover how people typically shake small, lightweight objects. 51 people, 19 female aged between 20 and 55, 32 male aged between 20 and 58, of different professions, including cafeteria staff and other non-office workers, were asked to shake both ping-pong balls, explicitly without further instructions. For each subject, 30 samples of roughly 5 seconds were taken: 5 each with both balls in the left, both in the right hand, one ball in each hand, and while either standing or sitting. This extensive data set of 1530 samples shows surprisingly large differences in style, frequency, and vigor of the shaking patterns. Samples with both balls in one hand serve as our "positive" data set where authentication should be successful. The cases where one ball was shaken in each hand are "neutral": because a single person is performing the motion, authentication could, but does not have to succeed.

7.2 Experiment 2: Pairs Trying to "hack" Authentication

The second experiment served to establish our "negative" data set of cases where authentication should not be successful. It was organized as a competition with a small prize to motivate participants to try harder. The goal was for a pair of subjects to produce shaking patterns as similar as possible to each other. 8 different pairs contributed 8 complete data sets of 20 samples each and 4 incomplete sets with less samples: 5 samples each for both subjects using their left hands, both their right, one subject left, the other right, and vice versa. Each sample has roughly 15 seconds, because some time was allowed for starting the motion and synchronization. For more flexibility in moving together, the pairs were only standing but not sitting. Immediate feedback after each sample was provided to the pairs in the form of the similarity values for both protocols, so that they could adapt their shaking patterns appropriately for highest values.

Data from these two experiments was used to find parameters for detecting active segments for the temporal alignment task, and to optimize parameter combinations for the feature extraction task. The parameters for feature extraction reported in section 5 have been found by a full parameter search using this extensive data set. For coherence, we use the parameter combination that generates the maximum difference of coherence averages between all positive and negative samples. Due to the larger parameter search space with higher dimensionality, for the second protocol we use the combination that minimizes $4e_P + e_N$. e_P is the percentage of false positives, i.e. the number of successful authentications for pairs, and e_N is the percentage of false negatives, i.e. the number of authentication errors for both balls shaken in one hand. That is, false positives were weighted higher than false negatives. The values listed above in the respective sections produced optimal results on this data set. An explorative analysis of the results depending on these parameters shows that most of them are robust w.r.t. the difference in coherence averages. This suggests that even with suboptimal parameter combinations, which may be the case when using these values with different data sets, results should not deteriorate significantly.



(a) Protocol 1 (b) Protocol 2

Fig. 6. Thresholds for coherence and the number of matching FFT slices control the trade-off between false positives and false negatives w.r.t. all positive/negative samples

Figure 6 shows the trade-off between false positives and false negatives, depending on the thresholds. Unsurprisingly, the percentage of false positives decreases with increasing thresholds for both protocols. For protocol 1, shown in Fig. 6a, false negatives begin to increase noticeably at a threshold of around 0.6, while for protocol 2, shown in Fig. 6b, they remain nearly constant. The threshold, either for coherence or for the percentage of matching candidate feature vectors, can be set by the application, or possibly even by the user. From a security point of view, we obviously prefer to restrict the number of false positives to zero. With a coherence threshold of 0.72 and a threshold of 84% matching parts, we achieve false negatives rates of 10.24% and 11.96%, respectively, with no false positives. These false negatives are sufficiently low to provide user friendly interaction, as also shown by our third experiment. The feedback of a failed authentication is immediate, and users just need to shake the devices again.

There is room for improving the results for our first protocol using coherence. As explained in section 5, we only use 3 seconds for comparing the time series. If active segments are longer than this, we can choose freely which parts to use. Figure 7 shows the average coherence values for our "negative" data sets, depending on the offset of the compared time series parts. Number 1 corresponds to the first 3 seconds, number 2 to the time series between 3 and 6 seconds, etc. The graph shows that two people tend to loose synchronization the longer the common movement needs to be sustained. We could exploit this fact by skipping the first few seconds and comparing later parts, at the expense of forcing users to shake devices longer. The results given above were generated by taking the beginning of the active segments, and thus with the most difficult parts.

Data from the first experiment was also used to estimate the entropy of feature vectors used for our second protocol. Using the parameters found with the first two data sets but 256 Hz instead of 512 Hz sample rate, quantized FFT coefficient vectors were computed over all 1530 samples. This parameter combination generates feature vectors of 21 discrete values from 0 to 5. Each subject

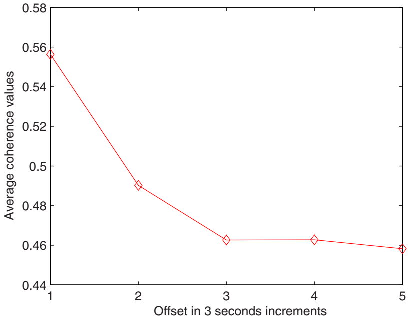


Fig. 7. Average coherence values depending on the segment offset show that people tend to synchronize their movements better at the beginning of their coordinated motions

generated on average 526.86 different feature vectors, with a minimum of 140 and a maximum of 1037. Aggregated over all subjects, there were 5595 different vectors for the left hands, 4883 for the right, and 7988 and 7770 for device 1 and 2, respectively. Overall, 12220 different feature vectors were generated during the first experiment, corresponding to an entropy of 13.58 bits per feature vector. If we assume an attacker to know which device, person, and hand are involved in a protocol run, this entropy decreases to around 7 to 10 bits, depending on the person. Overlapping feature vectors will have even less entropy, but we can still assume to generate at least 7 bits entropy per second using our second protocol.

7.3 Experiment 3: Single Subjects Live Usability Validation

The third experiment was run in "live" mode instead of data collection with batch processing, and used the same parameters. 30 subjects were asked to shake both devices in their dominant hand, with the aim of achieving successful authentication for both protocols. A simple GUI showed the status of both devices (active/quiescent) and the similarity values for both protocols, with green background if it was higher than the respective threshold and red for lower values. Subjects were asked to read a short list of tips for improving the similarity values (to align the devices roughly along the movement axis, to keep the wrist stiff, to shake quickly and vigorously, and to keep the elbow steady) and then to use interactive trial&error for achieving successful authentication. 8 of the subjects could immediately and reproducibly achieve this for both protocols starting with their first try, 8 subjects after at most 5, and 2 subjects after at most 10 tries. The remaining 12 subjects had more difficulty, but 7 could reproducibly achieve authentication after being shown once how others did it, and then within at most 3 further tries. 5 subjects only achieved authentication with either of the protocols, but not with both at the same time. This experiment shows that, even though the average rate of false negatives is low for the extensive data set from the first experiment, some people have more trouble to generate strong but similar movement patterns than others. Nonetheless, it also shows that the method is easy enough to learn within a few minutes from printed instructions and trial&error, and that it can be used intuitively after this brief learning period. The fact that a few subjects performed significantly better after being shown suggests that the printed instructions need to be improved.

8 Conclusions

We have proposed two protocols for authentication based on accelerometer data that generate secret keys between two devices when they are shaken together. Although using similar techniques for accelerometer data analysis, it is evident that the protocols achieve their aim very differently, from a security as well as from a protocol point of view. We consider the first protocol more secure, but the second to be more scalable. That is, if a large number of devices are in range of the wireless network, a device using protocol 1 may need to run Diffie-Hellman key agreement with a considerable number of other devices to find that which it is shaken together with. For the second protocol, it only needs to broadcast its candidate key parts stream, and the matching device can "tune in", i.e. synchronize, to this key stream. On the other hand, the security level of our CKP-based protocol 2 is limited to the entropy of the feature vectors, and is susceptible to offline attacks. When (pessimistically) estimating the entropy rate at around 7 bits per second, 20 seconds of shaking should be sufficient to achieve a security level of 128 bits. Users or applications may choose lower security levels.

Another potential issue in terms of security of protocol 2 is that secure hash functions, the cornerstone of our design, have been subjected to considerably less theoretical analysis than the DH construction or block ciphers which are used in protocol 1. New attacks on hash functions are being discovered [22], although the SHA-256 family of hashes, including the even more conservative SHA_{DBL}-256, is still considered secure. Additionally, protocol 1 utilizes these well-studied cryptographic primitives within a conservative design. An attacker has a one-off chance for an online attack – to guess the whole time series – and is thus significantly less likely to be successful than an offline attack on protocol 2. Although we can not currently quantify the security level against such unlikely online attacks, the security level of protocol 1 against offline attacks is 128 bits even after only 3 seconds of shaking (assuming DH to be secure). By introducing two protocols with different design, application developers can decide on this well-known trade-off between security and performance according to their requirements. Protocol 2 offers benefits for devices with limited resources, large wireless networks, and quick interaction, while we recommend using protocol 1 for higher security demands.

Feature extraction and cryptographic protocols are mostly independent of each other. Improvements in feature extraction to generate higher entropy and/or be more robust against off-center rotational effects in the movements can be used without modifying the cryptographic protocols, with the potential to significantly increase the entropy rate and thus decrease shaking time. For protocol 1, such improvements can even be distributed independently while remaining compatible to older devices. We note that our cryptographic protocols are also suitable for use with other types of sensors, while pre-processing and feature extraction tasks would most likely need to be modified.

Potential applications for our pairing protocols are manifold; coupling a mobile phone with a Bluetooth headset, establishing a transient secure connection between two smart cards for exchanging digital money, or passing access rights between key chains are prominent examples. 3D accelerometers are now being embedded into off-the-shelf mobile devices like the “Nokia 5550 Sport” and can immediately be used for authentication with our protocols. In our experiments described in section 7, we intentionally used simple, cheap accelerometers that are suitable for mass deployment.

The user interaction for authenticating devices is limited to just shaking them together for a few seconds, and is thus unobtrusive. By combining the explicit user interaction – taking two devices into one hand and shaking them as an indication that they should pair – with implicit authentication, we limit the burden placed on users. Connections are secured by default, not only as an option.

Full source code of our implementation including a demonstration application as well as our data sets are available as open source at `http://www.openuat.org`.

Acknowledgments

We gratefully acknowledge support by the Commission of the European Union under the FP6 Marie Curie Intra-European Fellowship program contract MEIF-CT-2006-042194 “CAPER”, and especially thank Kristof van Laerhoven for insightful discussions on analyzing accelerometer data and providing sensor boards.

References

1. Hoepman, J.H.: The emphemeral pairing problem. In: Proc. 8th Int. Conf. Financial Cryptography, Springer-Verlag (2004) 212–226
2. Holmquist, L.E., Mattern, F., Schiele, B., A., P., Beigl, M., Gellersen, H.W.: Smart-its friends: A technique for users to easily establish connections between smart artefacts. In: Proc. UbiComp 2001, Springer-Verlag (2001) 116–122
3. Stajano, F., Anderson, R.: The resurrecting duckling: Security issues for ad-hoc wireless networks. In: Proc. 7th Int. Workshop on Security Protocols, Springer-Verlag (1999) 172–194
4. Kindberg, T., Zhang, K., Im, S.H.: Evidently secure device associations. Technical Report HPL-2005-40, HP Laboratories Bristol (2005)
5. Kindberg, T., Zhang, K.: Validating and securing spontaneous associations between wireless devices. In: Proc. ISC’03: 6th Information Security Conf., Springer-Verlag (2003) 44–53
6. McCune, J.M., Perrig, A., Reiter, M.K.: Seeing-is-believing: Using camera phones for human-verifiable authentication. In: Proc. IEEE Symp. on Security and Privacy, IEEE CS Press (2005) 110–124
7. Goodrich, M.T., Sirivianos, M., Solis, J., Tsudik, G., Uzun, E.: Loud and clear: Human verifiable authentication based on audio. In: Proc. ICDCS 2006: 26th Conf. on Distributed Computing Systems, IEEE CS Press (2006) 10

8. Nicholson, A.J., Smith, I.E., Hughes, J., Noble, B.D.: LoKey: Leveraging the sms network in decentralized, end-to-end trust establishment. In: Proc. Pervasive 2006, Springer-Verlag (2006) 202–219
9. Čagalj, M., Čapkun, S., Hubaux, J.P.: Key agreement in peer-to-peer wireless networks. IEEE (Special Issue on Cryptography and Security) **94** (2006) 467–478
10. Gehrmann, C., Mitchell, C.J., Nyberg, K.: Manual authentication for wireless devices. RSA Cryptobytes **7** (2004) 29–37
11. Castelluccia, C., Mutaf, P.: Shake them up! In: Proc. MobiSys 2005: 3rd Int. Conf. on Mobile Systems, Applications, and Services, ACM Press (2005) 51–64
12. Hinckley, K.: Synchronous gestures for multiple persons and computers. In: Proc. UIST '03: 16th ACM Symp. on User Interface Software and Technology, ACM Press (2003) 149–158
13. Lester, J., Hannaford, B., Borriello, G.: "Are you with me?" – Using accelerometers to determine if two devices are carried by the same person. In: Proc. Pervasive 2004, Springer-Verlag (2004) 33–50
14. Batina, L., Mentens, N., Verbauwhede, I.: Side-channel issues for designing secure hardware implementations. In: Proc. IOLTS: IEEE Online Testing Symp. (2005)
15. Lukowicz, P., Junker, H., Tröster, G.: Automatic calibration of body worn acceleration sensors. In: Proc. Pervasive 2004, Springer-Verlag (2004) 176–181
16. Aylward, R., Lovell, S.D., Paradiso, J.A.: A compact, wireless, wearable sensor network for interactive dance ensembles. In: Proc. BSN 2006: Int. Workshop on Wearable and Implantable Body Sensor Networks, IEEE CS Press (2006) 65–68
17. Huynh, T., Schiele, B.: Analyzing features for activity recognition. In: Proc. Soc-EUSAI 2005. ACM Int. Conf. Proceeding Series, ACM Press (2005) 159–163
18. Ferguson, N., Schneier, B.: Practical Cryptography. Wiley Publishing (2003)
19. Diffie, W., Hellman, M.E.: New directions in cryptography. IEEE Trans. on Information Theory **IT-22** (1976) 644–654
20. Rivest, R.L., Shamir, A.: How to expose an eavesdropper. Commununications of ACM **27** (1984) 393–394
21. Vaudenay, S.: Secure communications over insecure channels based on short authenticated strings. In: Proc. CRYPTO 2005, Springer-Verlag (2005)
22. Wang, X., Yin, Y., Yu, H.: Finding collisions in the full SHA-1. In: Proc. CRYPTO 2005, Springer-Verlag (2005)

Virtual Walls: Protecting Digital Privacy in Pervasive Environments*

Apu Kapadia[1], Tristan Henderson[2], Jeffrey J. Fielding[1], and David Kotz[1]

[1] Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA
{akapadia,fielding,dfk}@cs.dartmouth.edu
[2] School of Computer Science, University of St Andrews, St Andrews, KY16 9SX, UK
tristan@cs.st-andrews.ac.uk

Abstract. As pervasive environments become more commonplace, the privacy of users is placed at increased risk. The numerous and diverse sensors in these environments can record users' contextual information, leading to users unwittingly leaving "digital footprints." Users must thus be allowed to control how their digital footprints are reported to third parties. While a significant amount of prior work has focused on location privacy, location is only one type of footprint, and we expect most users to be incapable of specifying fine-grained policies for a multitude of footprints. In this paper we present a policy language based on the metaphor of physical walls, and posit that users will find this abstraction to be an intuitive way to control access to their digital footprints. For example, users understand the privacy implications of meeting in a room enclosed by physical walls. By allowing users to deploy "virtual walls," they can control the privacy of their digital footprints much in the same way they control their privacy in the physical world. We present a policy framework and model for virtual walls with three levels of transparency that correspond to intuitive levels of privacy, and the results of a user study that indicates that our model is easy to understand and use.

1 Introduction

As sensor-rich pervasive environments become more common, users' privacy will be at increased risk [16]. Sensors can record a user's activities and personal information such as heart rate, body temperature, and even conversations. Users may unwittingly leave "digital footprints" (information about users derived from sensors) that can threaten their privacy. These footprints can be disseminated to applications, or stored for later retrieval, giving rise to useful context-aware applications. For example, applications can involve direct queries from other users ("What is Bob doing now?"), triggers ("Alert me when Bob is nearby"), or higher-level actions triggered by notifications ("Create a virtual meeting when Alice and Bob are free in their offices"). Such applications may be useful, but without adequate precautions, digital footprints may be accessed by unwanted parties. While several proposed mechanisms protect location privacy, location is just one kind of digital footprint, and these mechanisms do not directly apply to *all*

---

* This research program is a part of the Institute for Security Technology Studies and was supported by the Bureau of Justice Assistance under grant 2005-DD-BX-1091. The views and conclusions do not necessarily reflect the views of the United States Department of Justice.

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 162–179, 2007.

types of footprints. As the number and variety of sensors grows, it will be cumbersome for users to specify fine-grained policies about who can access which footprints.

We address one specific problem: the *confidentiality of digital footprints*, where we define a digital footprint as contextual information derived from raw sensor readings. By confidentiality, we mean that only authorized users should be able to access footprints as defined by the user's *privacy policy*. We believe that the term "digital footprints" is more intuitive to lay users than "context," since digital footprints evoke a sense of a digital trail that a user may leave in the virtual world. We feel that users will be more motivated to protect the privacy of their "digital footprints" rather than their "context."[1]

We propose a policy framework based on the intuitive concept of "virtual walls" that extends the notion of privacy provided by physical walls into the virtual realm. For instance, users are aware of their physical privacy in a closed room — outsiders cannot see or hear them. In a pervasive environment, however, their virtual privacy could be quite the opposite. Digital footprints from a videocamera and a microphone could expose their privacy in the virtual world, where other users *can* see and hear them by accessing their footprints. Figure 1 shows a meeting room where Alice and Bob have physical privacy, but sensors are disseminating personal footprints, such as their images and speech, to unwanted parties. Using virtual walls, users can "bolster" physical walls by specifying intuitive policies that control access to all their personal footprints in a way that is consistent with their notion of physical privacy. Virtual walls also relieve the burden of specifying separate policies for several footprints, which would be cumbersome in sensor-rich environments.

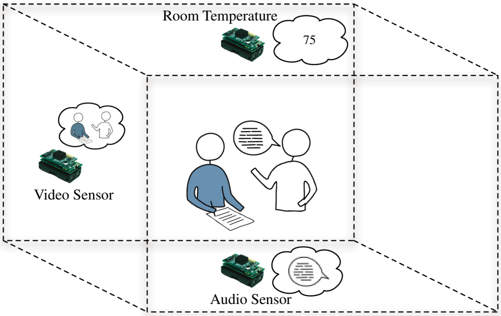


Fig. 1. Alice and Bob are aware of their physical privacy within the meeting room, but sensors are leaking personal footprints to the pervasive environment

We define three levels of transparency for virtual walls: *transparent*, *translucent*, and *opaque*, each with semantics that match what a user would expect in the physical world. For example, Alice's transparent virtual wall for a room allows users to "see" her personal digital footprints through the wall. An opaque wall restricts the visibility of all footprints within the room, including general footprints such as room temperature or

---

[1] The term "digital footprints" is not new. Its increasing use in the legal world [2,24] and the popular media [23,27] highlights growing attention to the consequences of digital tracking.

humidity. A translucent virtual wall discloses general digital footprints to outsiders (people are present, movement, etc.) while keeping the identities and personal footprints of people hidden. This is similar to physical translucence, through which outsiders cannot identify people, but can see general movement, light, occupancy, and so on.[2] Figure 2 shows a translucent virtual wall that prevents reporting of personal footprints, but keeps general footprints visible. To allow users to better control their privacy, we extend this metaphor by allowing users to create different virtual walls for different queriers. For instance, Alice could create a transparent wall around the cafeteria for her friends, but a translucent wall for her professors. These walls allow her friends to see her personal footprints, but disallow her professors from doing so. We validate the usability of our model through a user study and show that users are indeed able to understand and use our model to express privacy requirements.

This paper makes the following contributions:

A policy abstraction that is easy to understand and use. The metaphor of "virtual walls," validated by a user study, allows lay users to specify their privacy preferences in a way that is consistent with their notion of physical privacy.

Addressing privacy in sensor-rich environments. Users control access to their footprints without having to resort to fine-grained policies for each type of footprint.

A usable interface for specifying policies. We developed a prototype system, and evaluated our graphical user interface (GUI) in a user study. The study validates the ease of use of the GUI and provides valuable insights for further improvement.

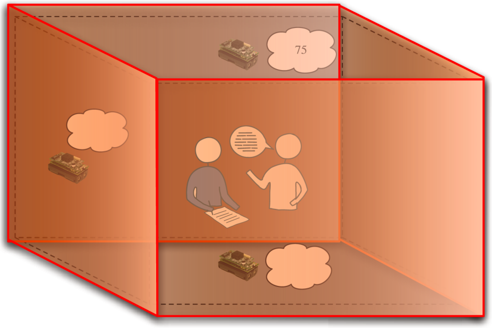


Fig. 2. Alice and Bob deploy a translucent virtual wall to prevent unwanted disclosure of personal footprints

Next, we outline our system architecture and then describe the virtual walls model in Section 3. Section 4 describes a user study that tested our model and GUI. In Section 5

---

[2] It is arguably impossible to guarantee complete privacy for users — for example, Alice's location might be hidden while she is in a meeting room, but if Charlie observes Alice enter the room, then her location privacy is compromised. Our goal, however, is to mirror Alice's expectations of privacy in the physical world, where she is already aware of such threats.

we discuss some challenges that would arise in a real deployment of virtual walls, and suggest future work. Section 6 discusses related work, and Section 7 concludes.

2 Architecture

Our system consists of a *context server*, which collects and disseminates digital footprints; clients, who instantiate walls or request footprints; and a sensing infrastructure that extracts footprints from sensor inputs. Users instantiate virtual walls by contacting the context server, which then uses the virtual walls to regulate access to footprints.

We use Solar [4] as the sensing infrastructure in our prototype system, but any such service will do. Solar is an open publish/subscribe framework for processing and distributing contextual events in a pervasive environment. For example, a footprint regarding Bob's current activity can be derived from an accelerometer and body-temperature sensors. Any higher-level inferences (made from raw sensor data) about users' activities generated by the Solar framework are delivered to the context server as footprints containing the activity (e.g., dancing or sleeping), a timestamp, and details about which sensors were used.[3] As we will see, the locations of sensors used to derive the footprint are used to describe where a footprint "originated," and footprint access is controlled in part by their origin. Footprints are treated as soft state, and never recorded to persistent storage. Users query the context server for the most recent footprints, and the footprints are returned if allowed by the virtual walls protecting those footprints. Users create virtual walls by using a GUI, which records the walls in a persistent database at the context server. Figure 3 illustrates the architecture of our system with an example.

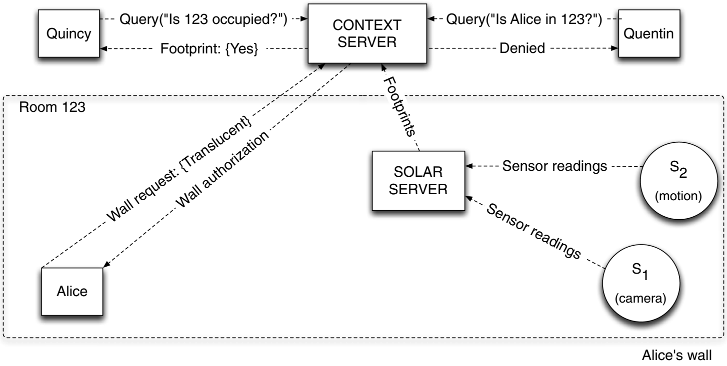


Fig. 3. Alice sets up a translucent virtual wall for Room 123. Raw sensor data are reported to the Solar framework, which generates higher-level footprints such as "Room 123 is occupied." Access to footprints is regulated by the virtual walls for the room. Quincy is granted access to general footprints within Room 123, but Quentin is denied access to Alice's personal footprints.

---

[3] Solar itself does not make any inferences from raw sensor data, but provides an architecture to do so. We assume that footprints are generated by inferring user activities from raw sensor readings [18,25].

Our proposed virtual wall system requires various security assumptions:

Trusted context server: All users trust the context server in terms of the confidentiality and integrity of their footprints. Only recent footprints are maintained, and only in non-persistent storage, which limits the long-term privacy risk.

Secure location claims: Secure location claims, for instance using a location-limited channel [22], are needed for the creation of opaque virtual walls (Section 3). The location system thus needs to have a level of granularity that is able to identify "places" such as rooms and hallways.

Sensor security: We assume that all sensors are approved by the administrators of the system. We ignore sensors that are deployed by malicious users (e.g., hidden microphones); virtual walls are intended to make pervasive environments more usable by securing *known* devices. We also assume that sensors communicate with the system over a secure channel that provides confidentiality and integrity of sensor data.

3 The Virtual Walls Model

We aim to keep the semantics of virtual walls as intuitive as possible, so that users who specify virtual walls have a clear understanding of who can access their footprints. In keeping with the metaphor of privacy afforded by physical walls, transparent virtual walls allow queriers to access any footprints, even a user's personal information (such as their heart rate or whether they are speaking); opaque walls block access to all footprints originating from within the wall; and translucent walls allow queriers to access only general information such as room temperature and the presence of motion. To add flexibility, users may create walls of varying transparencies for different queriers. For example, Alice may create a transparent virtual wall around her dorm room that applies to her friends and an opaque virtual wall that applies to her professors, thereby allowing finer control over the dissemination of her footprints. In essence, virtual walls are a means for users to specify discretionary privacy policies for their digital footprints.

3.1 The Model

To define our model we must describe *places*, *footprints*, *queries* and *virtual walls*.

Places. We refer to physical spaces such as rooms and buildings as *places*. Places have symbolic names, or *labels*, that are readily identifiable by users (e.g., "Room 251, Computer Science Building"), and users may deploy virtual walls around these places. We assume that physical areas are partitioned into non-overlapping or *atomic places* (e.g., a building is partitioned into distinct rooms), and users can define *aggregate places* that map to multiple atomic places. Virtual walls can be defined for any atomic or aggregate place. For example, Alice may specify a virtual wall around the "first floor" place. Our system maps this request to multiple virtual walls around the set of atomic places (rooms and hallways) on the "first floor." We assume a set of predefined places $\mathscr{P}$ and that regions are well-specified, i.e., each has a correct label meaningful to all users. This labeling allows users to specify virtual walls around recognizable places.

Footprints originate from places. Digital footprints derived from sensor data taken from place p are said to *originate* from place p, or places in set $P \subseteq \mathscr{P}$ if they are generated from sensor data in different places or if an individual sensor covers more than one place. For example, the footprint that describes Alice's speech activity, inferred from raw sensor readings taken in Room 251, is said to "originate in Room 251." We categorize footprints into two types: *general*, i.e., those footprints that do not reveal identifiable information about people (such as room temperature), and *personal*, that contain identifiable information (such as Alice's heart rate). We assume that footprints can be classified into one of these two categories, i.e., administrators or programmers can categorize footprints when new footprint generators are defined in the system. Personal footprints contain identifying information about a set of one or more people. For example, "Alice's heart rate" contains information about Alice, and a camera image of a room contains information about the people in that room. We call this set of people the *owners* of the footprint, and the owners of a footprint decide the visibility of that footprint. We assume a predefined set of users $\mathscr{U}$ to which an owner must belong.

Definition 1. *A* **general footprint** *is a tuple $f = \langle d, P, ts, v \rangle$, where descriptor d is a textual description of the footprint such as "Room temperature," $P \subseteq \mathscr{P}$ is the set of places where the footprint originates, ts is the timestamp when the footprint was generated, and v is an object that represents the value of the footprint. A* **personal footprint** *is defined with the same elements plus a set of owners $O \subseteq \mathscr{U}$. Only an owner of a personal footprint can regulate access to that footprint.*

Definition 1 formalizes general and personal footprints. In these definitions we do not concern ourselves with implementation details. Sophisticated implementations would specify d with several attributes (e.g., with an ontology [21]). Our prototype uses a simpler textual representation of d because efficient representation of footprints is orthogonal to the privacy problem that we are addressing. Similarly, the value v of a footprint can be implemented as different data types. Our implementation in Solar provides classes for various types of footprints, which are delivered as "events" in the Solar framework. For simplicity of presentation, we assume that there is only one owner o of a personal footprint, and explain group ownership of personal footprints in Section 3.2.

Footprints are categorized by their descriptor, origin, and owner if any. We refer to these categories as "footprint IDs." The footprint ID for a general footprint $\langle d, P, ts, v \rangle$ is the tuple $\langle d, P \rangle$. Likewise, the footprint ID for a personal footprint is $\langle d, P, o \rangle$. For example the footprint $\langle \textit{room temperature}, \{\textit{Room 241}\}, 16:42:01, 72 \rangle$ has the footprint ID $\langle \textit{room temperature}, \{\textit{Room 241}\} \rangle$. A personal footprint that infers whether Alice is moving may have the ID $\langle \textit{Movement}, \{\textit{Room 241}\}, \textit{Alice} \rangle$. In our implementation, a more recent footprint replaces an older footprint with the same ID and thus the context server only maintains the most recent versions of footprints. Personal footprints are expired from the system after a certain time period so that historical data about users are not recorded. We note that our model can be extended to record historical information.

Query model. We assume that users can query footprints from the context server using any combination of the fields defined for a footprint: place, descriptor, timestamp, owner, and value. For instance, Quentin could ask for a list of all footprints for a place (such as a conference room), and the context server would return a list of all footprints to which he has access that originate within that place. Similarly, Quentin could query Bob's footprints, such as "Is Bob speaking?" or "Is Bob moving?" Complex descriptors, as used in ontological databases, could enable more general searches. For example, "What is Bob's current activity?" could be mapped onto all footprints that measure Bob's activities such as his motion and speech footprints.

Virtual walls. Virtual walls protect the privacy of users by allowing them to control the visibility of their personal footprints and general footprints in their vicinity. In our implementation, users create virtual walls through a GUI. The context server records walls in a persistent database and uses them to enforce the user's access control policies.

Definition 2. *A* **virtual wall** $w = \langle o, p, t, A \rangle$ *belongs to owner o and protects footprints that originate in place p. We say that virtual wall w is* **around** *place p (e.g., a virtual wall around Room 251). The virtual wall w has* **transparency** $t \in \{\text{transparent}, \text{translucent}, \text{opaque}\}$. *Based on the transparency t, the wall w controls access to footprints originating in p from querying users* $A \subseteq \mathcal{U}$ *where* $\mathcal{U}$ *is the set of users in the system. We call A the* **apply-set***, since the wall w applies to users in A.*

The semantics of transparency are as follows. The transparency t of virtual wall $w = \langle o, p, t, A \rangle$ affects the reporting of digital footprints from place p to the set of users in A. For any personal footprint f_o for owner o and general footprint f_g originating from within the virtual wall w, a querier q

1. *(if t = transparent)* is allowed access to footprint f_o *only if* q is in the apply-set A (not "if," because another wall may block access),
2. *(if t = translucent)* is denied access to footprint f_o if q is in the apply-set A, or
3. *(if t = opaque)* is denied access to f_o and f_g if q is in the apply-set A.

Access to f_g is granted in the absence of an opaque wall. We will discuss the default behavior for access to f_p in the absence of virtual walls further below, but in summary, access to f_p is denied by default, unless the owner (or all the owners) of the footprint allows access with a transparent wall.

The creator of a virtual wall is the owner of that wall. For example the virtual wall $\langle \text{Alice}, \text{Room 256}, \text{translucent}, \{\text{John}, \text{Andy}, \text{Jim}\} \rangle$ is "Alice's translucent virtual wall around Room 256 that applies to her family," or in other words, Alice's family cannot access her personal footprints originating in Room 256. We assume a GUI through which users can define groups such as "Family," which are mapped to multiple users, such as {John, Andy, Jim}, by the system. Our prototype GUI contains this functionality and allows users to define groups, although we did not test this in our user study.

Virtual walls affect access to the *owner's* footprints; Alice's translucent virtual wall will protect *her* personal digital footprints, and not Bob's personal footprints (since Bob is not the owner of that wall). Our model allows different users to set up personal virtual walls around the same place to control access to their own footprints, and each user can create multiple walls for the same place. Section 3.3 discusses interaction between virtual walls created by different users, and how conflicts between walls are resolved.

Notice (in the above semantics) that transparent and translucent walls do not limit access to general footprints that originate within the wall; these are freely available to all users. This policy is based on our assumption that unidentified data do not usually threaten an individual's privacy. For instance, the temperature or occupancy of a publicly-shared place such as a cafeteria is unlikely to affect the privacy of people in that place. For "personal spaces" such as one's home or office, however, general information can be quite revealing — if Bob's office is occupied (a general footprint), then someone is in Bob's office, most probably Bob. In such cases, users can create opaque virtual walls around places where general footprints can result in a breach of privacy. Note that users are already aware of such threats in the physical world (lights on in an office indicate a user's presence), and so an opaque virtual wall is an intuitive countermeasure. Opaque walls may also be useful in public places. For example, if you are known to be the only person who works in a lab after 2am, general footprints from the lab may reveal your location. Since general footprints are not tied to any particular user, the owner of the opaque virtual wall is implicitly claiming ownership of the general footprints within that wall. As this affects other users within the virtual wall, we require unanimous consent from these users. In other words, Bob can freely create opaque walls around a place (even if he is not present in that place), but requires the consent of users present in that place. Section 3.3 discusses this issue in more detail.

In some cases a query may not have virtual walls controlling it, as it is unrealistic to expect that users will specify walls for every possible place and querier. To prevent the accidental release of personal footprints, we block access to personal footprints if there are no virtual walls that apply to a query. In effect, the semantics defined above protect such footprints by a 'default' translucent virtual wall, and users are informed as such. Despite this default behavior, we believe users find it more useful to specify translucent virtual walls explicitly. For instance, Bob may create an opaque wall around a room that applies to all users in the system. If he decides later to make a translucent virtual wall for his family, he could either remove "family" from the apply-set of the opaque wall, and rely on the default behavior, or explicitly create a translucent wall for his family. We believe that the latter is less confusing, and so we allow Bob to explicitly create translucent walls. Furthermore, we are exploring the use of more sophisticated default modes that depend on the type of place or the type of user; e.g., a "free-minded" user could have a default transparent wall, or a "paranoid" user may desire a default opaque policy. Our model also supports "mandatory" system policies, i.e., policies set by administrators that cannot be overridden by users, but we omit discussion for brevity.

3.2 Group Ownership

Some personal footprints in the system record data about a *group* of people, for instance, images captured by a camera. We call such footprints with multiple owners, *shared personal footprints*. Access to a shared personal footprint thus needs to be protected by the virtual walls of *all* the "co-owners," and access can be granted only if all these virtual walls permit the access. One could also envision systems in which these co-owners negotiate a "shared virtual wall" for their shared personal footprints. We leave the possibility of negotiation of group policies for future work. For now, we assume that the system has some way of identifying the group of owners for a shared personal footprint and applies the most restrictive wall of all the owners.

It is possible that unknown users (e.g., users who are not detected or recognized by the system) may be present in a shared personal footprint such as an image. Our system protects the privacy of those users who can be identified in the system; protecting the privacy of unidentifiable users is outside the scope of this work. Posted signs could inform such users that "cameras are present and your images may be broadcast over the network." In other words, users who are not protected by virtual walls implicitly provide informed consent by entering places with cameras. Group ownership of footprints is a powerful concept and Section 5 discusses our thoughts for future work. We now explain how conflicts arise between different virtual walls and how they are resolved.

3.3 Virtual Wall Interaction

Since multiple virtual walls can be defined for a place, and places themselves may overlap, we need to define the semantics of "conflicting" virtual walls. Two virtual walls *conflict* if there exists a querier for whom two or more different transparencies apply for access to a particular digital footprint. These conflicts can occur 1) between different walls owned by the same user (e.g., if Alice creates both a transparent and a translucent wall for her family), 2) by an opaque wall of one user that contradicts the transparent or translucent virtual walls of other users, or 3) between different walls for different owners for a group-owned personal footprint. For the first case, our framework ensures that conflicts between walls are resolved at creation time, and users are presented with feedback on the conflict. When creating a new wall that conflicts with an old wall, the user is required to choose whether the old or new transparency should be maintained for the affected users in the apply-set, and therefore conflicting walls for the same owner cannot exist within the system for the same place. For footprints with multiple origins, however, the most restrictive wall is applied. For the second case, after all users in the place have agreed to the creation of the opaque walls, the wall is added to the system. If access to a general footprint is restricted by an opaque wall, then access to that footprint is denied, overriding the transparent or translucent walls of the consenting users. For the third case, the most restrictive wall is used. Users are informed of this conflict (users are required to carry a device for creating walls and other interaction with the system) so that they are aware of the restriction placed by the wall of a co-owner of the footprint.

An opaque wall blocks access to the general footprints of all users within a place. Therefore, we require other users present within the place to collectively agree to the presence of an opaque wall. There are several possible approaches for such a negotiation based on who should be given priority in the negotiation. The different negotiation strategies can result in too many, or too few, opaque walls. As such, more work is needed to identify a reasonable strategy. For now, we assume a simple strategy based on unanimous consent. Opaque walls around a place can be removed by a new user entering that place. Entering users are asked whether they agree to the continued existence of the opaque wall. If not, the owner(s) of the opaque wall is given a small time window (e.g., 5 minutes) after which the opaque wall is removed from the system.[4] Therefore,

---

[4] We note that this "5 minute rule" appears arbitrary, and experimentation is required to identify a reasonable time window.

opaque walls are more effective, and reliably maintained, in places for which the owner has physical control — such as in an office, or in a reserved meeting room with restricted access. An outsider entering a reserved meeting room can be told to leave the room before he or she has the opportunity to disable the opaque wall. We are also exploring the possibility of assigning owners to places (e.g., Alice can be the owner of her personal office), who can then create opaque walls based on authority.

3.4 Limitations of the Model

Virtual walls control *all* footprints within the personal and general categories uniformly. Users may want finer-grained control over some footprints (such as location) and specialized mechanisms could be used in conjunction with virtual walls. We leave such a hybrid approach to future work. We have also equated the term "footprints" with context in an effort to make the model simple to understand. Our user study found that users had more difficulty with general footprints than with personal footprints (although the absolute performance numbers were high for both types of footprint). Perhaps another term may be better, but "general footprints" seems workable. Our model does not address multiple queries — Alice may want to restrict the rate of queries for her personal footprints. Lastly, in the physical world, Alice can see observers through a transparent wall and has a sense of her "exposure." In contrast, our model does not provide Alice with information about queriers. Such functionality would need to consider the privacy of queriers as well, and this is an exciting area for future research. It is unclear, however, how far one should push this analogy — at night, observers outside Alice's house can see her through her transparent window, but she may not be able to see them.

4 Evaluation

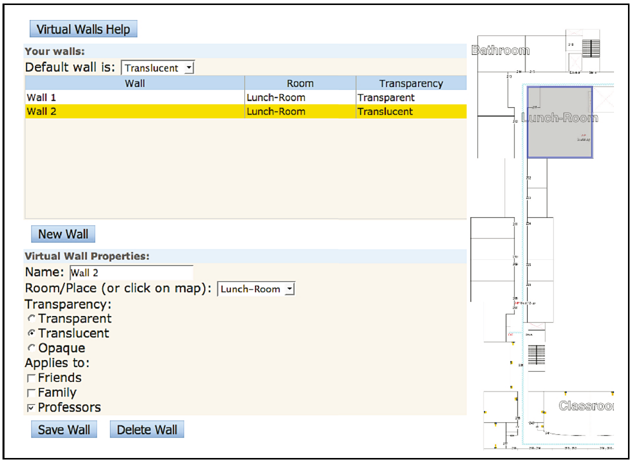
We built a prototype of our proposed system, including sensors, a context server and a GUI for creating virtual walls. We designed a user study to test the usability of both our model and the GUI; specifically, we tested the ease of understanding of the virtual walls model, the ease of use of the model, and the ease of use of the user interface.

4.1 Study Description

We recruited participants using flyers posted around campus, and advertisements on class and departmental e-mail lists, the student-run newspaper, and the popular social networking website "Facebook.com." Participation was *not* restricted to students, and was open to all adults in the community. In total we had 23 participants. We did not record any identifying characteristics such as age or gender, but only asked them whether they had ever taken a programming class, so as to classify the participants by computing experience: 9 participants had never taken a Computer Science (CS) class. The study comprised a paper booklet and an interactive component that used our GUI on a computer. The booklet contained a four-page introduction to pervasive environments and the virtual-wall system, and three sections of questions as listed below.

1. *Testing the ease of understanding of the virtual walls model.* This booklet section described three different scenarios, where some walls had already been created. In each scenario, participants were asked six questions about whether particular individuals would be able to access different types of data. For example, one scenario included a transparent wall around a dorm room that applied to family members, and asked if a parent could determine if the participant was awake in their dorm room.
2. *Testing the ease of use of the virtual walls model.* This booklet section contained three different scenarios, each with different privacy requirements. Participants were asked to construct walls (with checkboxes in the booklet) that satisfied these requirements. For example, they were asked to set up walls around a lunch room that would allow their friends to query footprints such as what they are eating.
3. *Testing the ease of use of the virtual walls user interface.* Participants followed "wizard" dialogs that introduced our GUI, and then performed tasks to create, modify, or delete walls, where each wall had a maximum of three elements in its apply-set. In each task the participant was told exactly what kind of wall was required — thus these tasks only tested the ability to use the GUI. For instance, participants were asked to set up a transparent wall around a lunch room that applied to friends.

The GUI was implemented using AJAX (Asynchronous JavaScript and XML) so that it could run in a web browser. It used a three-column interface. The first column, shown in Figure 4, contained a list of the participant's current walls, and a dialog box for creating, modifying or deleting walls. This dialog box contained an input box for naming walls, a drop-down box for selecting the room to which the wall applied, and checkboxes to determine transparency or apply-sets. The second interface column contained a map of the area where the walls were being created (in our study, this was a floor of our CS building). Existing walls were displayed on the map using different colors for



Your walls:

Default wall is: Translucent

Wall	Room	Transparency
Wall 1	Lunch-Room	Transparent
Wall 2	Lunch-Room	Translucent

Virtual Wall Properties:

Name: Wall 2

Room/Place (or click on map): Lunch-Room

Transparency:
○ Transparent
● Translucent
○ Opaque

Applies to:
☐ Friends
☐ Family
☑ Professors

Fig. 4. Portion of the virtual walls user interface

different transparencies, and clicking on particular rooms would allow a participant to create or modify walls for that room. The third interface column contained a description of the task that the participant needed to perform.

The participants took an average of 28.3 minutes to complete the entire study.

4.2 Study Results

We now present more detailed results of our study by breaking down the questions in each section into various categories. For the first two sections, we categorized questions based on the transparency of the wall in the question's scenario, and whether the footprints being accessed were personal or general. For these two sections we define a "Correct Response" as the case where a participant correctly selects or creates the wall to which the question refers. The third section of the study only considered the UI, and so we broke down the questions by the particular UI tasks. For this section a "Correct Response" is the case where a participant successfully performs the task in question; for instance, if a participant was asked to create a wall that applied to "Family" but instead created a wall that applied to "Friends," this would be an incorrect response.

Table 1 shows the correct responses in the three sections of the study. Unsurprisingly, participants who had taken a programming class ("CS participants") performed better than non-CS participants. All of the successful response rates are high and point towards a usable model. In the following we break down the results by section.

Table 1. Overall study responses

	Correct responses	
Section	*CS participants*	*non-CS participants*
1. Ease of understanding the model	99.4%	90.1%
2. Ease of use of the model	96.3%	90.5%
3. Ease of use of the user interface	97.2%	95.5%

Table 2. Successful responses by topic of question

(a) Section 1, understanding the model

Topic	*Personal Footprints*	*General Footprints*
Transparent	95.7%	91.3%
Translucent	93.5%	88.0%
Opaque	100.0%	95.7%

(b) Section 2, use of the model

Topic	*Correct responses*
Transparent	93.5%
Translucent	94.8%
Opaque	86.96%

Section 1: ease of understanding the model. Table 2(a) shows that participants had more difficulty with the concept of general footprints than personal footprints. Likewise, participants had the most difficulty with translucent walls. This was to be expected — transparent walls allow access to all footprints, and opaque walls restrict access to all footprints, leaving little room for error. Translucent walls, on the other hand, are

Table 3. Time to complete interface tasks

	Average (Standard deviation) time to complete task (seconds)		
Section	*CS participants*	*non-CS participants*	*non-CS (outlier removed)*
3a. Creating a wall	30.4 (13.7)	31.3 (19.2)	27.7 (11.9)
3b. Modifying a wall	11.6 (5.8)	15.1 (21.3)	10.9 (10.4)
3c. Deleting a wall	11.1 (4.2)	11.3 (9.6)	9.3 (4.7)
3d. Resolving a wall conflict	31.4 (11.2)	37.0 (26.2)	34.6 (23.5)

Table 4. Successful responses for interface tasks

	Correct responses	
Section	*CS participants*	*non-CS participants*
3a. Creating a wall	94.1%	92.9%
3b. Modifying a wall	100.0%	100.0%
3c. Deleting a wall	100.0%	96.4%
3d. Resolving a wall conflict	100.0%	92.9%

a combination of the two, and showed the highest error rates. In all six categories, however, the success rates were 88% or more, which point toward a usable model.

Section 2: ease of use of model. With regards to use of the model, it was only practical to categorize questions by the transparency of the wall that participants were asked to create (Table 2(b)). While opaque walls were easy to understand when presented to participants (Table 2(a)), three users did not use opaque walls correctly. A better explanation of opaque walls or a few more training examples might help users avoid such mistakes. We thus plan to improve our GUI to provide better feedback, e.g., through a dialog box that says "To block general footprints, you should create an opaque wall." After users become familiar with the model, they can choose to suppress such feedback.

Section 3: ease of use of interface. Table 3 shows the time taken to perform various tasks in the GUI. Resolving conflicts took the longest time, as this requires users to read a dialog box explaining the conflict, and pick the correct option to resolve the conflict. Creating a wall took about the same time, as this requires the largest number of commands: clicking on a room or the "new wall" button, selecting a transparency and an apply-set (restricted to a maximum of three elements), and clicking the "save wall" button. Modifying a wall involved changing just one wall element, and thus the times are similar to those for deleting a wall (which involves just one action to delete). There is a high variance in the responses for the non-CS participants, most of which can be explained by one outlier participant who appeared to have trouble with most of the tasks in the study. We have removed this outlier in the third column of Table 3, but even after doing so there is a high variance in the times taken to resolve conflicts, as there were two participants who took a very long time to do this. Even with these outliers, the longest task took about half a minute, which we believe makes our current GUI usable (this was reiterated by the participants' positive comments as described below). The next revision of our GUI, however, will concentrate on streamlining the wall-creation and conflict-resolution processes and incorporating users' suggestions for improvement.

Overall, participants were able to use the interface successfully (Table 4). There were two participants that had trouble with creating walls (for example, one user specified "Family" instead of "Professors" in the apply set), but apart from this the other mistakes were infrequent and appeared to be random.

User comments. The study offered participants the opportunity to provide feedback on the system through free-form written comments in the booklet. Many comments implied that participants would be protective of their personal information in a pervasive environment. For instance, one noted "Personally, I don't want people to be able to search the internet for what I am wearing, eating, etc," while another said "I'd refuse to have this kind of software following me." Within the constraints of our system, one participant said that "I think the default wall should be opaque in some cases," while another generalized by saying that "I think many people would choose 'opaque' for the majority of rooms/situations." The study helped us recognize that users may want default opaque walls for personal spaces. Another participant emphasized the need for a secure system: "[I]n the era of identity theft, there is heightened concern about privacy. How easy would it be to get someone's password and reconfigure their Virtual Walls?"

Some comments referenced confusion about the walls concept that was reflected in participants' answers to questions in the study. Two users were confused about locations: one said "I didn't quite udnerstand (*sic*) whether one could query about a specific person without knowing what room they are in" while another was "confused as to location and translucent walls — I understand that translucent would mean they could see people in [the] room but not which people." This indicates that the query model must be explained better to users of the system. Comments regarding the GUI itself were generally positive. Many requested further functionality, such as hotkeys, the ability to define rooms (for simplicity the GUI used in the study only had predetermined rooms), or sounds and additional colors to highlight particular events or interface components.

4.3 Limitations of the Study

Our study focused on the usability of the virtual walls *model*. We note that the GUI used in the study was designed to fit in a web browser on a desktop computer. We need to further explore scaling this GUI to smaller displays, such as on a mobile phone or PDA. We believe that our use of standard AJAX technologies should, however, facilitate the porting of our interface to such devices. It would also be useful to evaluate our model against other metaphors for usable privacy policies, but we leave this to future work.

5 Discussion and Future Work

Our user study affirms that the virtual walls model is easy to understand, and users can effectively translate privacy preferences into policies with virtual walls. For a real-world deployment, we anticipate the following challenges and opportunities for future work.

Creating walls: It may be cumbersome for users to constantly think about deploying virtual walls for every place they visit. A usable system will need to support higher level rules (in addition to default virtual walls) so that users can create virtual walls for not only a particular room, but a set of rooms based on certain conditions. For example, "Transparent wall around all work-related rooms during work hours," or "Translucent wall around current room if I am with my spouse."

Group ownership: A group of users may want to create a shared virtual wall to control footprints related to the group. For example, a group in a room may negotiate a shared translucent wall to restrict access to all personal footprints belonging to the group. Enabling groups to set up walls and negotiate their transparency and apply-set would require an extension to our model and a usable mechanism for such negotiation.

User disruption: Opaque walls, and possibly group walls in the future, require input from and feedback to users. In places with many users entering and leaving, users in those places may be disrupted by several messages. To reduce user input, users' responses could be automated by using stored preferences.

Data perturbation: It may be desirable to perturb footprints (e.g., changing granularity [9], darkening images [14] or adding "noise" [20]) when using translucent walls. This approach would be closer to the physical metaphor of translucence.

Mobile places: One can envision mobile places, such as a bus enriched with sensors. For instance, the bus may be parked inside a building and so be affected by virtual walls around that building. We would like to explore the semantics of virtual walls for mobile places, and how they interact with static places.

Deception: User studies in location privacy [15] have identified the need for deception, where users can lie about their location. Given the broad range of digital footprints that our framework is targeting, deception may be a challenging task.

6 Related Work

Several context-dissemination systems in pervasive computing support access-control mechanisms for protecting sensitive context information. Dey [7] built an experimental mechanism to control access to context in the Context Toolkit. The developer of a widget object can specify the "owner" of the information being sensed by implementing a function that computes the owner of an event. As mentioned earlier, we assume some such mechanism to infer the owner of a footprint.

While sensor-derived context information is still an active area of research [18,25], many applications based on context "sensed" from the user's computer, and systems to deal with the privacy of context information, are being studied. IBM's Grapevine service provides a user's context information to other users. A user's context is computed by monitoring her activities on her computer, and other users in the system can check to see her activity before initiating communication. Christensen et al. [5] report that while other users found it useful to query a user's context in certain situations, revealing context was a sensitive issue for most users and they ended up blocking context to all queriers. Fine-grained access control mechanisms were rarely used. It appears that in the absence of a usable policy language, users will be burdened by fine-grained policies, and end up being loathe to part with their (sensitive) context information. A policy language that balances ease of use and the granularity of access control is therefore needed, and our virtual walls system is an attempt to meet this balance.

A context-privacy system similar to ours is the "Digital Territory" project [3]. This project proposes "bubbles," which are "a temporary defined space that can be used to limit the information coming into and leaving the bubble in the digital domain." Bubbles can be shared between individuals and groups, and the flow of information in and out of the bubble can be adjusted. The bubble is similar to our virtual wall, although instead of translucency, bubbles offer a larger number of policies, which we believe will be unmanageable for complex environments. As far as we know, their system has yet to be implemented, and our translucency concept may prove useful for making the system usable. The pawS [17] system is also similar to ours — it sends privacy beacons to users as they enter pervasive environments to inform them of privacy policies. pawS concentrates on using machine-readable policies, however, while our focus is on policies that are easy for users to understand. Wickramasuriya et al. [26] examine context privacy in media spaces. They use RFID tags for localization, and then start monitoring users (through video sensors) only when policy violations occur (such as movement into a specific area). We anticipate that virtual walls will be used in environments where sensors are not just used for policy enforcement, and that users may opt for continuous monitoring due to the perceived benefits of the resulting context-aware applications.

Location is a primary piece of information for context-aware computing, and so several systems provide access-control mechanisms for location. Geopriv [6] defines a framework for securely disseminating location data by distributing location objects that are coupled with privacy rules. Geopriv, does not, however, address how privacy rules are defined. Hengartner and Steenkiste [12] support two types of authorization policies for location information: user policies specify who is allowed to access their location information, while room policies state who has the privilege to identify people in a particular room. The room policy is similar to virtual walls in that it is associated with a certain geographical area. Unlike our system, the owner of a room (not the users in that room) determines the room policy of that room. Other systems such as Confab [13] and LocServ [19] allow users to specify fine-grained policies to control their location information. For example, LocServ's authorization language expresses constraints such as time, location, and quality of service (i.e., the granularity and anonymity of location information). As mentioned earlier, we provide a policy language for expressing privacy policies about all kinds of footprints, of which location is only one example.

We believe that virtual walls will make privacy in pervasive environments more usable, in particular where large numbers of sensors and users are involved. The idea of using room or wall-like metaphors has been applied in non-security scenarios, for instance by Henderson and Card [11]. Similarly, the idea of simplifying privacy policies into easily-understandable levels has been explored elsewhere, for example, Hawkey and Inkpen's "privacy gradients" [10]. Usability in sensor-network scenarios, however, has been little studied. Barkhuus and Dey examine location-tracking services and find that users are more concerned about privacy if their location is being tracked, rather than if the device is simply aware of its own position [1]. Iachello et al. [15] develop and trial a location-aware application, and produce a set of guidelines for application designers, which we discussed in Section 5. Finally, Elliot et al. [8] conducted interviews to study how households handle communication information at home, and find that people do attach ownership to information according to the ownership of physical spaces.

7 Conclusion

In this paper we outline the need for an intuitive policy abstraction to address the privacy of users' digital footprints in sensor-rich pervasive environments. We believe that most users will be incapable of specifying complex privacy policies for information sensed about them, and so an abstraction for this purpose must be easy to understand and use. To meet these goals, we proposed and evaluated a policy framework that extends the metaphor of physical walls to pervasive environments. Virtual walls control the spread of personal and general contextual data, or "footprints," and offer privacy analogous to that afforded by physical walls. We formalized the semantics of access control using virtual walls, and evaluated our model through a user study. Our results indicate that the model is easy to understand (users can correctly identify the behavior of virtual walls), and easy to use (users can translate privacy requirements into an appropriate set of virtual walls). Moreover, comments from study participants indicate that privacy in sensor-rich environments is an important problem that might affect the deployment of such environments. Based on our results, we believe that virtual walls are a promising metaphor for specifying usable privacy policies in pervasive environments.

Acknowledgments

We would like to thank Kazuhiro Minami, Harald Vogt, the Dartmouth CMC Lab and the anonymous reviewers for their helpful comments, and Ron Peterson for suggesting the original concept. We also thank Alex Iliev, Chris Masone, Sara Sinclair, and Phoebe Wolfskill for pre-testing our user study.

References

1. L. Barkhuus and A. Dey. Location-based services for mobile telephony: a study of users' privacy concerns. In *Proceedings of the 9th IFIP TC13 International Conference on Human-Computer interaction (INTERACT 2003)*, Zürich, Switzerland, Sept. 2003.
2. M. H. Barrera and J. M. Okai. Digital correspondence: Recreating privacy paradigms. *International Journal of Communications Law and Policy*, 1(3), Summer 1999.
3. L. Beslay and H. Hakala. Digital territory: Bubbles, 2005. Draft publication. http://cybersecurity.jrc.es/docs/DigitalTerritoryBubbles.pdf.
4. G. Chen, M. Li, and D. Kotz. Design and implementation of a large-scale context fusion network. In *Proceedings of Mobiquitous 2004*, pp 246–255, Boston, MA, USA, Aug. 2004.
5. J. Christensen, J. Sussman, S. Levy, W. E. Bennett, T. V. Wolf, and W. A. Kellogg. Too much information. *ACM Queue*, 4(6):50–57, July-Aug. 2006.
6. J. R. Cuellar, J. B. Morris Jr, D. K. Mulligan, J. Peterson, and J. M. Polk. Geopriv requirements, Feb. 2004. RFC 3693.
7. A. K. Dey. *Providing Architectural Support for Building Context-Aware Applications*. PhD thesis, College of Computing, Georgia Institute of Technology, Dec. 2000.

8. K. Elliot, C. Neustaedter, and S. Greenberg. Time, ownership and awareness: The value of contextual locations in the home. In *Proceedings of UbiComp 2005*, pp 251–268, Sept. 2005.
9. M. Gruteser and D. Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *Proceedings of MobiSys 2003*, pp 31–42, San Francisco, CA, USA, May 2003.
10. K. Hawkey and K. M. Inkpen. Privacy gradients: exploring ways to manage incidental information during co-located collaboration. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, pp 1431–1434, Portland, OR, USA, Apr. 2005.
11. D. A. Henderson, Jr. and S. K. Card. Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics*, 5(3):211–243, July 1986.
12. U. Hengartner and P. Steenkiste. Protecting access to people location information. In *Proceedings of the First International Conference on Security in Pervasive Computing*, pp 25–38, Boppard, Germany, Mar. 2003.
13. J. I. Hong and J. A. Landay. An architecture for privacy-sensitive ubiquitous computing. In *Proceedings of MobiSys 2004*, pp 177–189, Boston, MA, USA, June 2004.
14. S. E. Hudson and I. Smith. Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In *Proceedings of the 6th ACM Conference on Computer Supported Cooperative Work*, pp 248–257, Boston, MA, USA, Nov. 1996.
15. G. Iachello, I. Smith, S. Consolvo, M. Chen, and G. D. Abowd. Developing privacy guidelines for social location disclosure applications and services. In *Proceedings of the 2005 Symposium on Usable Privacy and Security*, Pittsburgh, PA, USA, July 2005.
16. M. Langheinrich. Privacy by design - principles of privacy-aware ubiquitous systems. In *Proceedings of UbiComp 2001*, pp 273–291, Atlanta, GA, USA, Sept. 2001.
17. M. Langheinrich. A privacy awareness system for ubiquitous computing environments. In *Proceedings of UbiComp 2002*, pp 237–245, Göteborg, Sweden, Sept. 2002.
18. J. Lester, T. Choudhury, and G. Borriello. A practical approach to recognizing physical activities. In *Proceedings of Pervasive 2006*, pp 1–16, Dublin, Ireland, May 2006.
19. G. Myles, A. Friday, and N. Davies. Preserving privacy in environments with location-based applications. *IEEE Pervasive Computing*, 2(1):56–64, Jan.-Mar. 2003.
20. B. A. Price, K. Adam, and B. Nuseibeh. Keeping ubiquitous computing to yourself: A practical model for user control of privacy. *International Journal of Human-Computer Studies*, 63(1-2):228–253, July 2005.
21. A. Ranganathan, J. Al-Muhtadi, and R. H. Campbell. Reasoning about uncertain contexts in pervasive computing environments. *IEEE Pervasive Computing*, 3(2):62–70, Apr.-June 2004.
22. N. Sastry, U. Shankar, and D. Wagner. Secure verification of location claims. In *Proceedings of the ACM Workshop on Wireless Security*, pp 1–10, San Diego, CA, USA, Sept. 2003.
23. B. Schneier. Your vanishing privacy. *The Star Tribune*, page 1AA, Mar. 05, 2006.
24. P. Sommer. Digital footprints: Assessing computer evidence. *Criminal Law Review*, pp 61–78, Dec. 1998.
25. E. M. Tapia, T. Choudhury, and M. Philipose. Building reliable activity models using hierarchical shrinkage and mined ontology. In *Proceedings of Pervasive 2006*, pp 17–32, Dublin, Ireland, May 2006.
26. J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian. Privacy protecting data collection in media spaces. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, pp 48–55, New York, NY, USA, Oct. 2004.
27. E. Wieffering. Protecting your digital footprints. *The Star Tribune*, page 1D, Nov. 07, 1999.

Trust, Privacy and Relationships in 'Pervasive Education': Families' Views on Homework and Technologies

Katie Fraser[1], Tom Rodden[2], and Claire O'Malley[1]

[1] Learning Sciences Research Institute
[2] Mixed Reality Laboratory,
University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham
`{kcf, tar}@cs.nott.ac.uk, com@psychology.nottingham.ac.uk`

Abstract. Extensive educational research discusses the potential for information and communication technologies in supporting homework, but most has focused on providing content. The research in this paper focuses instead on the issues around managing homework and balancing home and school through the capabilities of ubiquitous technologies. As part of our requirements capture we presented three families with demonstrators of ubiquitous computing systems. Our technologies provoked reactions to situated and embedded information capture and access, and locational information capture through mobile devices. The subtlety and complexity of roles and relationships of different family members raised issues around trust and privacy in relation to children's homework practices. We consider how these drove acceptance of the technologies, and how the contrasts between family and educational relationships produced different requirements for technologies managing information transfer inside and outside the home. Overall, we highlight how respect for these concerns can inform the design of pervasive technologies, particularly within the domestic and educational contexts bridged.

1 Introduction

The challenge of integrating pervasive technologies into the domestic environment has been the focus of a wide body of research since the late 1990s [1-3]. Issues surrounding the adoption of these technologies, their social acceptance, and the privacy and security of data collected in the home have both established and echoed general concerns in ubiquitous computing [4-6]. This focus on the domestic context means the majority of this work concerns technologies that capture information to complement and service domestic life – domestic capture for domestic application. These applications have included the development of technologies to link distributed family members, the development of assistive technologies to support "aging in place", and tools to help with domestic organization and coordination [7-11].

In this paper we will explore the boundary between the domestic and other aspects of life. While many of the existing scenarios of future smart homes have highlighted home and work links we wish to consider an educational perspective. In particular, we

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 180 – 197, 2007.

will focus on the ways in which future pervasive technologies may be exploited to support children’s homework, school activities set by the school for children, and generally expected to be completed at home, out of school hours. This exploration allows us to consider how the lessons of the domestic reach out. How are the established social issues surrounding domestic ubiquitous computing [5] affected when we consider activities that span beyond the home?

Our research into technologies in homework has identified the transfer of information between home and school as a concern for both families and educationalists. The project has involved children ranging from 8-16 in a variety of different family arrangements – in the current paper we focus on a core study with three families with children aged 12-16. Current rhetoric within the educational world claims an ideal use of mobile ICT technologies is to support home-school links [12] and so this is an obvious area where the uptake of pervasive technologies could be imminent, making this a critical time for the pervasive community to examine such claims. Could such ideas actually happen in practice? Are families interested in blurring the boundary between home and school and do they feel ubiquitous computing techniques are an appropriate way to support homework? This research will attempt to address this agenda, by studying pervasive computing within the home and children’s homework, and families’ reactions to it.

2 Ubiquitous Computing and Education

While there has been little direct research into ubiquitous computing and education, the technical properties of pervasive technologies [6] are complementary to current drives in education to produce richer, personalized learning experiences. Many authors within the area of educational technologies have called for the use of ubiquitous devices in the future of learning [12-14] suggesting that these can allow the seamless integration of both computing, and the education it supports, into everyday life [15-19]. Mobile technologies, with their ability to run context-aware applications, allow the delivery of personalized learning experiences away from school [20]. In both mobile and embedded technologies, the automated capture and access of information about location and activity allow further context-sensitive delivery of learning materials, and the information collected offers both teachers and students the chance to reflect upon learning experiences [14].

In terms of intentional learning at least (as opposed to informal or incidental learning), the majority of children’s out-of-school educational activity currently takes place within the home. Homework is interesting from a research perspective because it offers a different lens through which to see (and capture) domestic activity compared with previous studies. What is more, the application of UbiComp in educational domains is still little explored.

In addition to capitalizing on the benefits of ubiquitous computing, education will also need to grapple with the challenges it faces. With the sharing of domestic information outside the home (i.e., in school), the obvious technical issues facing ubiquitous computing include privacy and security of information [21]. We are most interested in the social implications and negotiations needed around capturing and sharing information about the domestic with the outside world. Technology for capturing contextual information about children in the domestic – monitoring their activities and location – is generally used by parents of quite young children – e.g., using technologies such as baby monitors to check safety. From the school end, although teachers have been traditionally viewed as in loco parentis[1], we are interested in investigating whether teachers are seen as having the same rights of access to contextual information as parents. This contrast between home and school use will be a driving concern in this paper.

Researchers in education seem content that the benefits of educational improvement will be sufficient for family members to accept and adopt technologies in the home, with most concerns surrounding families' finances and the socioeconomic drivers of uptake [22, 23]. This study will follow from a tradition of work looking at both domestic and educational uses of technology [5, 24-27] inspiring critical consideration of the effects of technologies on such domains of application. We will attempt to bridge the gap between the theoretically focused work of proponents of such technologies, and the more practical studies that do exist, but tend to focus on the academic success of the implementation of a specific device [28, 29]. However, we also wish to identify when and where privacy issues might occur, when sharing information is seen as mutually beneficial [30], and what technological solutions might aid or exacerbate this. In the following section we will therefore report on the reactions of families to a range of ubiquitous computing technologies.

3 The Study

This study reported in this paper was informed by several other elements of a three-year research project into technologies in homework, consulting multiple stakeholders from the school and home contexts. Our focus on introducing pervasive technologies into the home-school relationship and the technologies explored within the study arose as a result of a survey of Heads of ICT at 34 local schools and interviews conducted with 6 Heads of ICT from this pool. These interviews identified networks linking home and school as the future of homework activity. In order to assess how this trend might impact homework we looked at the use of traditional homework technologies through video diaries with 8 families and discussion groups with 180 children. Our findings suggested that negotiation of parental access to homework activities with and through homework technologies was often necessary for children, and we wished to see whether negotiation of access would be necessary and manageable in future pervasive computing contexts.

In order to investigate the negotiation of access in more depth, we returned to a school from our study of Heads of ICT, where use of technology to link home and school had already begun. We asked them for feedback on the systems they used, and for general details of their use of technologies in the home-school area. Ten families agreed to speak to us initially and they identified three areas in which technologies were currently used. These were *accessing information through the Internet and school intranet*, *transferring information between home and school*, and *coordinating*

---

[1] *In loco parentis* allows the school to deal with children in its care as a parent or guardian would.

home and school trips. The potential links between these issues and pervasive homework technologies drove the selection of three demonstrators of future homework technology use. In order to explore potential use of pervasive technologies to support homework we set up in a tour of our research laboratory for three families, and asked them for feedback and comments on the systems, the ways they might affect home-school links and how these technologies might fit within their homes.

4 The Key Demonstrators

Given our concern with integrating future technologies into the home, we sought to make the demos provocative, emphasizing the range and degree of information that could be gathered about families’ lives using pervasive technologies to explore their acceptability. Each demonstrator is described below with short usage examples, the educational motivators and pervasive technologies involved.

Demonstrator 1 - Situated Information Access

Our interviews with the families showed that their primary use of digital technology in homework was to access information, mostly through search engines, but that they were unable to efficiently judge the quality of this information. One major benefit of pervasive technologies mentioned by educationalists is its ability to connect users to a variety of specialists. We therefore took ways of improving situated information access through specialist knowledge as the theme for our first demo.

Mixed Reality Boundary. A situated display allows children to navigate around a bazaar of teachers, each offering subject specialties, and clustered together so that the answer of Geography teacher 1 can be compared to that of Geography teacher 2. As seen in Figure 1, the Mixed Reality Boundary is a situated display that allows the child (pictured top left) to chat face to face with teachers, and navigate around a virtual area, exploring multiple viewpoints and learning from teachers with a variety of expertise. The technology presents a view of an “always on” communication future that allowed us to explore families’ reactions to a two-way exchange of information.

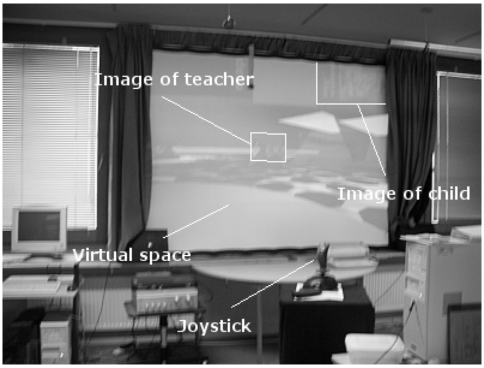


Fig. 1. The Mixed Reality Boundary. This video conferencing system could be placed on the wall of a home, and used to select and communicate with experts.

Table 1 summarizes the key educational motivations and the key pervasive computing technologies within this arrangement.

Table 1. Educational motivation behind the situated information access demonstrator

Demonstrator	**Educational motivator**	**Pervasive technology**
Mixed Reality Boundary (MRB)	Provides easy access to teachers, who would necessarily provide higher quality material than search engines, and allows children to easily compare multiple viewpoints.	Situated display technologies with video conferencing links.

Demonstrator 2 - Embedded Information Capture and Access

Our interviews identified families' discontent with the current school system linking home and school through technology. In particular they felt this was awkward and slow. Another major benefit of pervasive computing expounded upon in the educational literature is its ability to seamlessly transfer contextual information between different settings. We therefore took the transfer of information through pervasive technologies in a 'lounge' within our lab space as our second demonstrator, which illustrated a number of pervasive computing arrangements and use scenarios.

Intelligent Text Books. A child opens his schoolbag and places a copy of Macbeth on the table. This opens a webpage on his television displaying additional study resources, which end in a quiz for the teacher to assess. The text books were set up to display webpages with appropriate educational content when placed on the table, as seen in Figure 2.

Seat Photo Capture. A child sits down to do their homework, and a camera records the information. The picture is sent to the teacher with details of when and where the

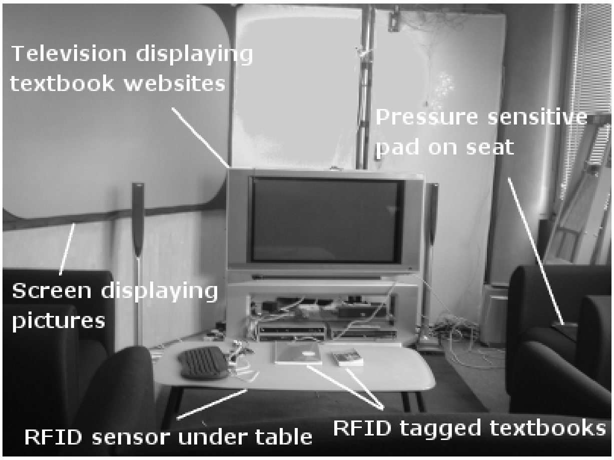


Fig. 2. The Lounge area with screen, pressure sensor and tagged homework resources on table

homework task began, so she can check work is being completed promptly. Here, a pressure sensitive pad was placed on the chair, so it registered when a family member sat down and took a photograph of them, as seen in Figure 2.

The Homework Cupboard. A child comes home after a day at school, and opens the cupboard where his textbooks are stored. The display screen on the adjacent cupboard lights up with his homework task. A light-reactive sensor was placed on the cupboard seen in Figure 3 so it triggered a homework diary on a display on the next cupboard it was opened.

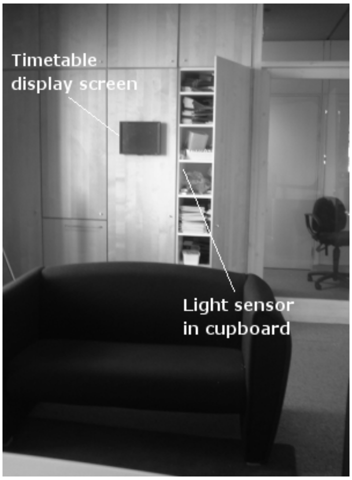


Fig. 3. The Lounge area homework cupboard and cupboard display

These demonstrators allowed us to investigate how children and parents felt about essentially ‘negotiation-free’ transfer of information built into their domestic environments. We wished to establish which, if any, information was deemed suitable for handling automatically in this way. Table 2 summarizes the key educational motivations and the key pervasive computing technologies within this arrangement.

Table 2. Educational motivation behind the embedded pervasive technologies

Demonstrator	Educational motivator	Pervasive technology
Intelligent text books	Pervasive technologies offer the ability to receive record and deliver information based on the child’s actions. RFID tags can exchange information at the trigger of an everyday activity, such as placing homework books on a table. Uses might include sending reminders or recording children’s progress.	RFID tags and sensor embedded within physical artefacts to access digital information.

Table 2. (*Continued*)

Seat photo capture	As with the textbooks, here information exchange could be triggered by a child sitting down to their desk. The camera shows how information can also be recorded for later.	Pressure sensor and cameras used to support activity based triggering.
The homework cupboard	As with the previous two examples, opening a cupboard or a bag to start a homework task might trigger events.	Light sensor used to track and record activity
Output screens	The output screens offer the opportunity to use children's actions as a cue to display information, such relevant webpages.	Appropriation of screens in domestic settings

Demonstrator 3 - Locational Information Capture Through Mobile Devices

Our final demo took inspiration from the families' use of mobile phones to communicate and coordinate the journey home from school. It linked together this parental wish to track information with the educational literature's ideas on the use of mobile devices for learning, which have include tracking devices which alert the child (or teacher) to educational opportunities around them.

Locational Tracking. A child carries their mobile phone with them, and a tracking system sends reminders and alerts dependent on their location. Their parent notices they are on their way home early, which will allow a change of plans for the evening, making it possible for the child to get their homework done and visit their aunt. The child sees a piece of litter that needs to be recorded for a project, and takes a photograph, with the locational information available to map litter locations at school.

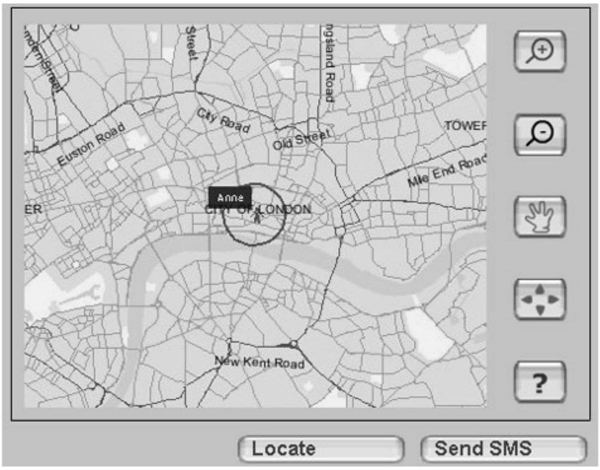


Fig. 4. The Location-based service

This demonstrator showed systems for tracking children and recording information about their location at specific times through a mobile network, with results such as that seen in Figure 4.

Table 3 summarizes the key educational motivations and the key pervasive computing technologies within this arrangement.

Table 3. Educational motivation behind the locational capture demonstrator

Demonstrator	**Educational motivator**	**Pervasive technology**
Mobile phone tracking	When children can be located through a mobile system it allows them to be notified when near a library, or for location-related school projects.	Mobile tracking devices.

5 The Families

The families involved in this study were recruited from an independently funded boy's high school, where use of technologies was high in the school and in families' lives in general. We chose this group to ensure a good background in technology use between home and school, and familiarity with the basic concepts behind the demonstrator technologies. The families represent a range of family arrangements and children in different stages of their high school education. This set of families demonstrated diversity in terms of stages of schooling but also the relationships within the different families.

Family A consists of a mother, father and only son, aged 12. The father works full time and the mother part time. They own a single computer, placed in a peripheral conservatory, which is used mainly for accessing the Internet and playing games. Technology does not appear to play a central role in the family's life, and television and mobile phones are the only other technologies that are used as routine. The PC was assembled by a close family member, and the father stated it was purchased for the son's education, although he uses it often and the mother occasionally.

Family B consists of a mother and father, a son, aged 14, and a daughter, aged 12. Both parents work full time. Family B has integrated technology more fully into their everyday lives. The main family PC is set up in a central position in the lounge. The son has an extensive work and entertainment centre in his bedroom, with PC, digital television and Playstation2. He has been particularly been encouraged by his parents to use technologies to overcome difficulties in writing stemming from dyslexia. Again, family members all have mobile phones.

Family C is a single mother, a son, aged 14, and a daughter, aged 16. The mother works full time. The family has a television, and a PC, which is well equipped with printer and scanner, but shared within the family, and slow to load. The PC is used extensively and creatively by all the family members, for a variety of home and school projects. It is set up in a corner of the mother's bedroom, and used individually. Each family member has his or her own mobile phone.

6 Lessons Learned

Each of the three families started the technology tour at a different demo, and then rotated around each demonstration area in turn. Each family was debriefed after their tour, using a semi-structured interview encouraging them to discuss how they could see these technologies being used generally within a domestic context, and specifically within homework. We also wanted to contrast how these information types were seen for home use and school use, so asked families to discuss how both parents and teachers might benefit from them. In this section we wish to present the key lessons to emerge in terms of three key themes

- The link between the pervasive information capture and interpersonal relationships.
- The impact of emerging rights and responsibilities within family structures.
- Managing access to those outside the family.

Key comments from the families are presented within this paper as verbatim quotes from interview logs, contextualized within conversations, in order to illustrate these themes in depth. Initial reactions to the technology were positive. Families welcomed the use of new technologies, particularly within an educational context, as can be seen in their attitude to the embedded sensor technologies:

> *"Good, because the book's there and then you don't get in detentions or whatever, because you know, you know that you've brought the right books in and everything" [son, Family A].*

We also noted that the mobile location technologies were received with slightly more scepticism than others, but families still acknowledged their practicality:

> *[father] "It's good, but, er, they can turn the phone off, so then you can't find where they are, so if they go somewhere where they're not meant to be then they can just turn the phone off and then it won't be finding them"*
> *[interviewer] "Yeah, so you can trick your parents into not finding you"*
> *[father] "You wouldn't do that, would you? No, we thought it was a good idea, didn't we?" [Family A].*

These reactions suggest that the face value of such technologies is clear to families, and ties in with the educational rhetoric that practical applications of pervasive computing technologies will avoid concerns. Further conversation, however, elicited deeper concerns. Our analysis drew out the three themes we use to consider how families felt pervasive technologies might influence relationships and negotiation. The first two explicitly examine family relationships between individuals and as a whole. The last looks at relationships between the family and the outside world.

6.1 Information Capture and Interpersonal Relationships

Situated information capture involving seamless and non-negotiated capture of information was seen as acceptable within the context of certain relationships in the home. Ubiquitous devices could monitor the 'vulnerable', in the context of relationships established to ensure safety. In keeping with previous research, the very old or the very young were seen as candidates for this kind of care:

> *“any person that you feel might be vulnerable, you know, living on your own” [father, Family A]*

Within the families, we focused on the relationships between our parents and children. The acceptability of capturing information about day-to-day life, particularly within homework, seemed linked to the emerging roles of children as independent adults as they grew up within the family settings. As we look across our families a pattern of acceptance, resistance, and finally rejection of parental access to children’s lives seemed evident as children grew older. The only son in Family A, at age 12, seemed to be quite accepting of information capture in his day-to-day life:

> *[father] “I think it’s getting used to doing that instead of just picking up the phone and dialling… punching his number in”*
> *[interviewer] “Would you take long to get used to it? How would you find it?”*
> *[son] “I wouldn’t feel like I was being tracked [on] really” [Family A].*

In Family B the elder son, aged 14 talked about resisting information capture:

> *[mother] “they’ll not be able to say to you ‘well, no, I did loads’ when they didn’t.”*
> *[interviewer] “And how would you feel about that kind of thing then?”*
> *[son] “I like being able to lie about it!” [Family B].*

In Family C, however, where the eldest child was a 16-year-old daughter, both daughter and mother rejected the capture of information about her life, suggesting that pervasive technologies would not only be useless, but also inappropriate:

> *“you know if she was out on a Saturday night, it probably is better that I don’t know!” [mother, Family C].*

These differences suggested that the parent-child dynamic changed as time went by, something that was also explicitly mentioned in our interviews:

> *“obviously when he gets older, when he’s out with his girlfriend and stuff, he doesn’t want dad: ‘where are you?’” [father, Family A].*

Relationships and Trust Change Privacy of Information. The definition of personal or private information differs according to relationships, as has been shown in other research into social networks [31]. We found that carers had the strongest right to capture information, whether caring for the generally vulnerable, or for their own children. This suggests that the parent-child relationship, where parents are seen as the ultimate ‘carer’, should support the automatic transmission of information more readily than the teacher-child relationship. However, even within-family monitoring is not a simple concept. As children became older, the role of a parent as carer becomes more blurred, and more information is viewed as personal, private, or generally inappropriate for capture. The relationship has to be constantly renegotiated and evolves. The effect of this emerging role of teenagers is enhanced in a context like the management of homework, which is notoriously sensitive [32]. The changing roles of children as they become teenagers are well established. Here, however, the elements of trust implicit in carer relationships are applied to the capture of information where inflexible technologies might make it difficult to allow this relationship to grow with the children.

6.2 Rights, Responsibilities and Family

Within the family, rights [33] and responsibilities [34] create a complex social environment for sharing information. Trust between parent and child facilitates the sharing of such information. For example, families saw the potential of tracking, as with current mobile use, for coordination – two-way sharing of locational data, rather than one-way monitoring. This is contrary to previous research which suggested that mobiles are primarily a one-way monitoring, even if a two-way negotiation tool [35]:

> *"it would probably minimise the phone calls of 'where are you? how long are you gonna be?' them sort of things, erm, and like for us, to know when we've left work, or I'm still at work" [mother, Family B].*

However, these were non-homework related examples. Although the trust relationship could lead to families trading information freely, demanding information about a child's homework activities could suggest a lack of trust. Both parents and children were concerned that use of captured data might violate the rights of children, and felt that reliance on information capture subverted the idea that a child should and could take responsibilities for their actions, especially in a homework context:

> *"I think it's really mostly the child's responsibility, that they have to take responsibility really, and it shouldn't always be up to the parents to make them do it, and they should accept the consequences" [son, Family C].*

Reactions to such a violation differed from family to family. For example, the 14-year-old boy in Family B protested far more than the 14-year-old boy in Family C, with conversation in Family B focusing on the boy wishing to maintain privacy, and in Family C more on the boy's responsibilities. What is more, rights were not always earned by a display of increased maturity, as the daughter from Family C shows:

> *"isn't it breaking a bond of trust, if you can't trust them to do that, and... I think it's, it's just part of growing up that you don't, you don't do it [homework]... I just hate the fact that it would be completely controlled" [daughter, Family C].*

Trust Works in Mysterious Ways. While trust levels in the family were higher than those outside, this did not always lead to acceptance of ubiquitous information capture and disclosure. With children coming into adolescence, as with our group, both locational and activity recording technologies were seen as the right to privacy and the need for the child to take responsibility for their actions. Similarly, the adoption of responsibility by the child could be compromised by implying a lack of trust in this way. The interviews certainly suggested that information capture and disclosure between parent and child generally needed to decrease as time went by, and mobile and personalized technologies do offer the ability to configure information capture and disclosure in this way. However, we must ask how such configuration would be implemented, with the effect of age overshadowed by individual family relationships.

Technologies Make Relationships Explicit. Familial trust around information access and capture is certainly heavily situated. Trust is necessary for the sharing of information, but demanding the sharing of information can violate trust. There is an arrangement between parent and child that demands responsibility, and conveys rights, but when children avoid responsibility or parents ignore rights, the arrangement is flexible and seems to acknowledge that children are still in training as adults. Homework is cited as a core example of this training, where children were learning to take responsibility for their own actions. The complexities of responsibility have been explored in sociological literature [36], but what do they mean for technological design? Such complexities were presented to our interviewers as unproblematic facts of everyday life. Trust is not just negotiated in a highly complex and situated manner within the family but rights and responsibilities are also more intimate, and may be instinctively rather than explicitly negotiated.

However, introducing configurable tools to manage information sharing requires explicit rules, so here designers face a large challenge. Suchman [37, 38] has argued that technological attempts to make interactions explicit – accepting or denying access to information being such an example – can highlight both the tool and the natural social process, causing breakdowns in relationships and actions. Making technologies configurable may make these interactions even more explicit. For the transmission of homework information within the family, this might be addressed by allowing space within the design for social solutions to the problem, allowing the signal sent through the technology to be ambiguous, giving the user the space to create socially acceptable stories around these signals [31, 39, 40]. On top of this, the constantly evolving nature of relationships can be accommodated in such processes by the constant renegotiation of relationships.

6.3 Managing Access to Outside the Family

We have discussed the complexity of information sharing within the family. Lastly families considered the exchange of information with the outside world. Here trust was lower. Families were sceptical that information gathered by environmental sensors or mobile devices would remain under their control, and even though the demos we showed them were mostly configured to transfer information into rather than out of the home they felt that the information was inevitably vulnerable:

> *"you've got a database so people can watch what you're doing" [son, Family B].*

However, it was not just the dangers of unnamed malicious parties families had privacy concerns about. Reactions to the Mixed Reality Boundary (MRB) showed that even when teachers were involved, allowing access to the home put a completely different slant on the acceptability of technology:

> *[daughter] "Erm, it didn't feel intrusive, because it was just… if it was in homes or something, I think it would"*
>
> *[interviewer] "Yeah, so we're talking about this TV screen being able to do the same sort of thing, so we could actually go…"*
> *[daughter] "In your home?!"*
> *[interviewer] "In your home, or, or in your classroom, or…"*

> *[daughter] "I think it's a fun concept"*
> *[interviewer] "Right"*
> *[daughter] "But I wouldn't like it in my own home!" [Family C].*

The only way in which domestic information would be shared was with strong family control of disclosure, as with this response to sharing homework tasks:

> *"I think the essential thing is switching it on and off" [mother, Family C].*

Lastly, from the families' discussion of sensors it seemed that educational uses were only acceptable when providing strong incentives, i.e. diagnosing difficulties:

> *[mother] "if there's a record of what textbooks have you looked at … if you've then got a poor mark in a test that you've got proof that you've really revised hard for, then you know that there's some underlying problem." [Family B].*

The contrast between views on information sharing within and outside the family was particularly strong with mobile devices. As discussed before, families saw coordination as important for parents and children's social interactions, but were reluctant to draw this into homework. What is more, no family was open to sharing this information with schools, as with previous media coverage [41].

The Importance of Information Type. Negative reactions to sharing domestic information increased as this information became more personal and situated. What made some information more sensitive than others? Currently available search technologies were viewed positively – these tend to collect aggregated, low-fidelity information about activities. However, in our more provocative demos, we exposed families to domestically located, detailed, temporally stamped information, and strong negative reactions were seen. This viewing of some high-fidelity data within the domestic as highly personal replicates findings in other studies of information capture in the domestic environment [11]. The Mixed Reality Boundary, which transmitted visual, temporal, and some locational information was seen as our most controversial technology, a common theme for video media spaces [42]. However, its use within a library or classroom was totally accepted, confirming that privacy is not just about the type information sent, but the place of the home. Similarly, sharing of personal locational information was not even considered outside the family. We suggest that the particularly strong reaction against the use of this technology outside the home is because locational information is always considered personal and private, based on extreme reactions to locational information even in locations like schools [41].

Situated Control of Personal Information is a Necessity. A database was seen as a threatening record, with fear about the ability to control captured information high. Even strong educational scenarios in the sensor and information finding demos, while acknowledged as useful, did not overcome these concerns. The family was the only group who were trusted to both receive and use personal information without filtering. However, this was in a limited range of circumstances – such as with children trusted to track parents as much as vice versa. The model transfers poorly into more one-sided homework sharing. When it came to controlling the sharing of homework information through technologies, only the family was trusted to control disclosure of domestically situated information. Technologically based solutions to disclosure we discussed with families such as security filters were rejected. Simple controls were demanded, and seemed to be the best ways to assure families of safety. To achieve this, designers need to ensure visibility of processes – for example, using awareness tools to show families how the information gathered in their home is transmitted and to allow them to manipulate a good working model of the network and its devices.

7 Conclusions

This study has explored the acceptability of ubiquitous computing technologies to support homework within a group of families, and has been undertaken as part of a long-term engagement with a range of families. Within this paper we have presented reactions to three distinct technological scenarios. These highlight a number of distinct conclusions of broad interest to the pervasive computing community and those interested in the application of advanced technologies within education.

The Domestic and 'Ubiquitous' Education

In seeking to deploy pervasive technology in domestic settings to establish stronger links between the home and educational settings of our children we are infiltrating an environment in which a complex series of values are already present. One of the core challenges that has already been identified in this move to domestic settings is understanding the social implications of pervasive technologies [5]. We have described how the sharing of information gathered inside this environment with external agents such as teachers provides an expanded set of issues surrounding privacy and relationships, be they interpersonal, or within and without the family. Alongside calls for an increased use of ubiquitous technologies in education, this has established the transfer of educational information within and outside the domestic as an important area for research.

The differences between domestic and educational contexts and the effect this has on the reception of ubiquitous computing have been clear. We have shown that trust is constantly expected between parent and child in the family: such implicit trust arrangements are unlikely to exist in schools, where it is usual for children to be monitored through assessment and attendance records. Although we do not mean to suggest that trust is absent between teacher and child in school, it is often negotiated at the interpersonal level, rather than accepted school-wide. These core differences suggest that domestic and educational contexts will clash in the messages they send to children, and designing pervasive technologies to adapt to this difference is a complex challenge.

Ubiquitous Computing and Education

References to roles, rights and responsibilities in this article have emphasised that there are many social restraints on the immediate adoption of ubiquitous computing in education. Some of the difficulties in using technologies that share information about domestic contexts with educationalists might be tempered by including social norms in the process that mirrors those seen in the home.

For example, trust and coordination seemed to be the basis on which family members shared information. By framing ubiquitous computing technologies as coordinating between teacher and child, trust might be established on a more equal footing, and the acceptability of such devices in home-school relations increased. There is a choice to be made here between replicating the relationship between children and parents – allowing a great deal of information to be shared, but demanding some degree of reciprocity – and settling for highly controlled information exchange, which current models of schooling support.

Children's increased control of education is the principle behind introducing ubiquitous technologies into education. However, increasing the autonomy of children may not be always about allowing them to choose the learning content and support they access, but also about allowing them to negotiate when and how to access learning resources. When using technologies as a way of making education more ubiquitous, we need to consider both the control of information sharing, and the control of educational infiltration. This is particularly true in domestic contexts, where research has become more focused on the ludic [43] and the difficulties of transferring work (and presumably educational) principles and technologies into the home.

Guidelines for Design

Basic guidelines for ubiquitous computing design can also be gathered from this work. Although much of our work has taken a cautionary tone we believe these represent a positive contribution to the development of technologies in this area.

We have discussed that information sharing within the family is a highly sensitive issue, and when insufficiently negotiated between relevant parties, it can undermine trust, responsibilities and rights. However, we wish to establish that in the family, privacy and trust go beyond simple information control. While configurable devices are one way to address this problem, they rely solely on controlling the information captured, and not on supporting established relationships. The complicated interactions of family members appear impossible to model or support with technological controls, making technological solutions to privacy and control an inadequate solution to the problem. Allowing families to negotiate their own solutions while children adjust to new rights and responsibilities – as they have, unaided, for many years, may avoid situations where making such interactions explicit could cause issues. As designers our best strategy for transmitting homework information seems to be to allow within-family signals to be ambiguous and interpretable.

The contrast comes when homework information is shared with the school. Visibility and control of technologies were core demands of families when sharing information with the outside world. Families were apprehensive about the recording of data about their everyday lives, with an obvious solution to this being explicit highlighting of the channels through which this data travels and mechanisms for controlling this. We have discussed that this reflects current models of trust between child and teacher, rather than a necessary condition of their future relationship.

Our last comment must be on the importance of context in framing the sensitivity of and manipulation of information. Avoiding sensitive types of information is a way to sidestep issues, with our findings suggesting that the capture of temporal, visual and locational information might best be avoided where possible. However, we can also repurpose, rather than reduce this information, with the positive effects of trust in the family suggesting wider opportunities for increasing the acceptability of pervasive technologies. Social steps such as these, or more local workarounds where we are limited in our technical solutions, such as increased control of devices or the space for social negotiation of device use, are applicable across domestic, educational, and many more contexts. The challenge for pervasive computing is the design of technologies that allow flexible social negotiation and are able to span the more fluid world of the home and the structured setting inherent within our educational environments.

References

[1] Crabtree, A. and Rodden, T. (2003). Finding a place for UbiComp in the home. In Proceedings of UBICOMP, Seattle, WA, USA: Springer-Verlag.
[2] O'Brien, J. and Rodden, T. (1997). Interactive systems in domestic environments. In Proceedings of DIS, Amsterdam, The Netherlands: ACM Press.
[3] Mateas, M., et al. (1996). Engineering ethnography in the home. In Proceedings of CHI: ACM Press.
[4] Hong, J.I., et al. (2004). Privacy risk models for designing privacy-sensitive ubiquitous computing systems. In Proceedings of DIS, Cambridge, MA, USA: ACM Press.
[5] Edwards, W.K. and Grinter, R.E. (2001). At home with ubiquitous computing: Seven challenges. In Proceedings of UBICOMP Conference: Springer.
[6] Abowd, G.D. and Mynatt, E.D. (2000). Charting past, present, and future research in ubiquitous computing. ACM Transactions on CHI, 7(1): p. 29-58.
[7] Mynatt, E.D., et al. (2001). Digital family portraits: Supporting peace of mind for extended family members. In Proceedings of CHI, Seattle, WA, USA: ACM Press.
[8] Dalsgaard, T., et al. (2006). Mediated intimacy in families: Understanding the relation between children and parents. In Proceedings of International Conference for Interaction Design and Children, Tampere, Finland: ACM Press.
[9] Vetere, F., et al. (2005). Mediating intimacy: designing technologies to support strong-tie relationships. In Proceedings of CHI: ACM Press.
[10] Mynatt, E.D., Irfan, E., and Rogers, W. (2000). Increasing the opportunities for aging in place. In Proceedings of ACM Conference on Universal Usability, Arlington, Virginia, US: ACM Press.
[11] Beckmann, C., Consolvo, S., and LaMarca, A. (2004). Some assembly required: Supporting end-user sensor installation in domestic ubiquitous computing environments. In Proceedings of UBICOMP, Nottingham, UK: Springer.
[12] Green, H., et al. (2005). Personalisation and digital technologies. Nesta Futurelab. p. 1-29.
[13] Taylor, J., et al. (2004). An E-learning Research Agenda. EPSRC, ESRC and E-Science. p. 1-9.
[14] Ainsworth, S., et al. (2005). Cyberinfrastructure for Education and Learning for the Future: A Vision and Research Agenda. A. Bernat, J. Smith, and D. Rothschild, Editors, Computing Research Association. p. 1-40.
[15] MacKinnon, K.A., Yoon, S., and Andrews, G. (2002). Using "Thinking Tags" to improve understanding in science: A genetics simulation. In Proceedings of CSCL, Boulder, CO, UK.
[16] Mazalek, A., Davenport, G., and Ishii, H. (2002). Tangible Viewpoints: A physical approach to multimedia stories. In Proceedings of ACM Multimedia Conference, Juan-les-Pins, France: ACM Press.

[17] McNerney, T. (2004). From Turtles to Tangible Programming Bricks: Explorations in physical language design. Personal and Ubiquitous Computing, 8(5): p. 326-337.

[18] O'Malley, C. and Stanton Fraser, D. (2005). Learning with Tangible Technologies. In Literature Review Report, Nesta Futurelab: Bristol, UK.

[19] Price, S., et al. (2003). Using 'tangibles' to promote novel forms of playful learning. Interacting with Computers, 15: p. 169-185.

[20] Sharples, M. (2000). The design of personal mobile technologies for lifelong learning. Computers and Education, 34: p. 177-193.

[21] Proctor, T. (2005). Data security threats in the home environment. In Proceedings of IFIP WG 9.3 Home-Orientated IT (HOIT) Conference, York, UK: IFIP.

[22] Lewin, C., Mavers, D., and Somekh, B. (2003). Broadening access to the curriculum through using technology to link home and school. The Curriculum Journal, 14(1): p. 23-53.

[23] Valentine, G., Holloway, S., and Bingham, N. (2002). The digital generation?: Children, ICT and the everyday nature of social exclusion. Antipode, 34: p. 296-315.

[24] Palen, L. and Dourish, P. (2003). Unpacking 'Privacy' for a Networked World. In Proceedings of CHI, Ft. Lauderdale, Florida, USA: ACM Press.

[25] Selwyn, N. (2000). The National Grid for Learning: Panacea or Panopticon? British Journal of Sociology of Education, 21(2): p. 243-255.

[26] Land, R. and Bayne, S. (2002). Screen or monitor? In Improving student learning using learning technology, C. Rust, Editor. OCSLD: Oxford, UK. p. 125-138.

[27] Greenfield, A. (2006). Everyware: The dawning age of ubiquitous computing. Berkeley, CA, USA: New Riders.

[28] Luckin, R., et al. (2005). Using mobile technology to create flexible learning contexts. Journal of Interactive Media in Education, 22: p. 1-21.

[29] Hughes, M., et al. (2003). Exchanging knowledge between home and school to enhance children's learning in literacy and numeracy. In Proceedings of ERNAPE Conference, Poland.

[30] Brin, D.G. (1998). The transparent society: Will technology force us to choose between privacy and freedom? Reading, MA, USA: Perseus Books.

[31] Iachello, G., et al. (2005). Developing privacy guidelines for social location disclosure applications and services. In Proceedings of Symposium On Usable Privacy and Security (SOUPS), Pittsburgh, PA, USA: ACM Press.

[32] Solomon, Y., Warin, J., and Lewis, C. (2002). Helping with homework? Homework as a site of tension for parents and teenagers. British Educational Research Journal, 28: p. 603-622.

[33] Davies, S.G. (1997). Re-engineering the right to privacy: How privacy has been transformed from a right to a commodity. Technology and Privacy: The New Landscape, Ed. P. Agre, E. and M. Rotenberg. 1997, The MIT Press: Cambridge, MA, USA.

[34] Such, E. and Walker, R. (2004). Being responsible and responsible beings: Children's understanding of responsibility. Children and Society, 18: p. 231-242.

[35] Williams, S. and Williams, L. (2005). Space invaders: The negotiation of teenage boundaries through the mobile phone. The Sociological Review, 53(2): p. 314.

[36] Solomon, Y., et al. (2002). Intimate talk between parents and their children: Democratic Openness or Covert Control. Sociology, 36(4): p. 965-983.

[37] Suchman, L. (1994). Do categories have politics? Computer Supported Cooperative Work, 2(3): p. 177-190.

[38] Suchman, L. (1987). Plans and situated actions: The problem of human machine communication. Learning in doing: Social, cognitive and computational perspectives, Ed. R. Pea and J. Seely Brown. Cambridge, UK: University of Cambridge.

[39] Boehner, K. and Hancock, J.T. (2006). Advancing Ambiguity. In Proceedings of CHI, Montreal, Quebec, Canada: ACM Press.

[40] Aoki, P.M. and Woodruff, A. (2005). Making space for stories: Ambiguity in the design of personal communication systems. In Proceedings of CHI, Portland, Oregon, USA: ACM Press.

[41] 'Student ID badges raise privacy issues' (2005). Fox News. http://www.foxnews.com/story/0,2933,149814,00.html. Access 14 November 2005.

[42] Boyle, M. and Greenberg, S. (2005). The language of privacy: Learning from video media space analysis and design. ACM Transactions on Computer-Human Interaction, 12(2): p. 328-370.

[43] Gaver, B., Dunne, T., and Pacenti, E. (1999). Cultural Probes. Interactions, 6(1): p. 21-29.

Operating Appliances with Mobile Phones – Strengths and Limits of a Universal Interaction Device

Christof Roduner, Marc Langheinrich,
Christian Floerkemeier, and Beat Schwarzentrub

Institute for Pervasive Computing, ETH Zurich, 8092 Zurich, Switzerland
{`roduner, langhein, floerkem`}`@inf.ethz.ch`, `be@student.ethz.ch`

Abstract. Mobile phones are increasingly becoming ubiquitous computational devices that are almost always available, individually adaptable, and nearly universally connectable (using both wide area and short range communication capabilities). Until *Star Trek*-like speech interfaces are fully developed, mobile phones seem thus poised to become our main devices for interacting with intelligent spaces and smart appliances, such as buying train passes, operating vending machines, or controlling smart homes (e.g., TVs, stereos, and dishwashers, as well as heating and light). But how much can a mobile phone simplify our everyday interactions, before it itself becomes a usability burden? What are the capabilities and limitations of using mobile phones to control smart *appliances*, i.e., operating things like ATMs or coffee makers that typically do not benefit from remote control? This paper presents a user study investigating the use of a prototypical, mobile phone based interaction system to operate a range of appliances in a number of different task settings. Our results show that mobile devices can greatly simplify appliance operation in exceptional situations, but that the idea of a universal interaction device is less suited for general, everyday appliance control.

1 Introduction

Personal mobile devices, such as mobile phones and PDAs, represent an important building block in many ubicomp systems discussed in the community [1,2,3,4]. Their widespread use and their characteristics as a general purpose computing platform make them appear as ideal devices for implementing many interactive services in intelligent environments. Scenarios involving personal mobile devices range from attaching digital annotations to physical objects [5,6,7], sharing public displays [8,9], and interacting with appliances of all sorts [10,11].

Using mobile phones and PDAs to query and control smart environments and artifacts is attractive due to four main aspects of today's devices:

- *Wireless Communication:* Apart from the continuously expanding wide area coverage, mobile operators are also increasingly offering digital (i.e., packet switched) communication services such as GPRS, EDGE, or UMTS, which

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 198–215, 2007.

can provide fast, reliable, and economic device communication from almost anywhere in the world, both indoors and outdoors. Moreover, short range communication protocols, such as infrared (IR) and Bluetooth, allow local ad-hoc communication between similar devices.

- *Tag Detection:* The recent addition of *Near Field Communication* (NFC) technology not only improves intra-phone communication (i.e., simplifying the often complicated Bluetooth setup process) but also allows mobile devices to detect and read out passive (NFC-compatible) RFID-Tags. Moreover, camera phones can use 2D barcodes to allow even printed paper to "send" information to a mobile device.
- *Computational Resources:* Mobile phones and PDAs have become powerful computing devices, often featuring processors with hundreds of megahertz and considerable RAM and Flash memory. Given their energy demands for sustained wide area communication provision, their powerful batteries can often easily support substantial calculations and short range communications without significantly affecting the primary usage (e.g., telephone or organizer) of the device. Users are also accustomed to recharging their devices periodically, thus providing a virtually unlimited energy supply for locally installed applications.
- *Programmable Screen and Keyboard:* Many devices already feature large color displays and programmable soft keys, 5-way navigation buttons, click wheels, or even touchscreens, allowing system designers to quickly build a wide range of attractive and versatile interfaces. Built-in microphones and speakers, together with dedicated MP3 processors, can additionally support the use of speech commands and audio cues.

Together, these four aspects can provide two important novel provisions to the control of appliances, namely

- *Information Provision:* The mobile device can provide additional information regarding the state of the appliance, either by providing a display to an otherwise displayless appliance, or by extending an already existing, but smaller embedded display.
- *User Interface Provision:* The mobile device's programmable screen and keyboard can be used to extend or even personalize the appliance's user interface, especially in situations where space on the appliance is limited.

The use of handheld devices for controlling the environment has already a long tradition, based on the (usually infrared-based) remote controls that provide access from afar to audio/video equipment such as TVs and stereos, but also lights, shades, garage doors, or even cars. Given the many remotely controllable appliances found in today's households, however, it is becoming less and less practical to have a separate remote control for each of them. Also, users increasingly need to carry out tasks that involve more than a single appliance, e.g., switching on the DVD player while also turning on the TV at the same time. Last but not least, many of today's remote controls are overloaded with functionality that users hardly ever need, resulting in large, unwieldy devices and hard-to-use interfaces.

A number of research projects (e.g., [12,13,14]), as well as commercial products (e.g., Philips Pronto[1]), have grown out of these needs. They often use a PDA to dynamically download a user interface from the appliance. Nichols et al. and Zimmermann et al. developed these ideas further and proposed to use a PDA as a *personal universal controller* [15], or *universal remote console* [16], respectively, which in turn led to the standardization of the *universal remote console framework* within the INCITS V2 technical committee [17]. Among the appliances that are typically considered controllable by such universal remote controllers are video recorders, DVD players, TVs, video projectors, stereos, answering machines, light switches, home security systems, ATMs, elevators, copy machines, cars, vending machines, heating control systems, microwaves, ovens, and washing machines.

Researchers have recently begun to look at the suitability of different possible interaction techniques (such as scanning a barcode, pointing with a laser beam, or touching an RFID tag) for such scenarios [18,19]. However, there is surprisingly little work addressing the question of which appliances are actually suitable for this new paradigm of interaction, and under which circumstances they are so. Koskela et al. [20] have studied the use of mobile phones, PCs, and media terminals in a household over six months. However, the handheld device could only be used to control lights and curtains in their setting. Rode et al. [21] conducted an ethnographic study to find out which household appliances users choose to program. Their research gives, however, no indication for which appliances a personal mobile device may be an appropriate interaction tool. Moreover, they do not consider spontaneous interaction with an appliance, but focus on the programming of actions for the future and the creation of macros to facilitate repeated tasks.

User studies investigating the performance of personal mobile devices for the spontaneous interaction with appliances were carried out by Nichols et al. [11,22] as part of the evaluation of their personal universal controller [15]. They studied the use of a PocketPC to control a stereo and a telephone/digital answering machine. In particular, they compared the performance of 12 subjects when accessing the appliance either using the PDA interfaces, or the interface on the actual appliance. The authors found that, compared to the user interface on the PocketPC, interaction based on the physical appliance interface took twice as long, entailed twice as many errors, and required five times more external help.

While these results seem very encouraging to the vision of using mobile phones as universal interaction devices in the future, they might strike one as somewhat counterintuitive: Why would a softkey-based interface of a generic PDA be more efficient for playing back voice messages of a *real* answering machine than the machine's own buttons? Why wouldn't the direct interaction with the physical machine help users with understanding and operating the device, by making use of the machine's *perceived affordances* [23]?

---

[1] See `www.pronto.philips.com`

Obviously, using a PDA or mobile phone as an interaction device will be greatly beneficial when this presents the only way to access an otherwise invisible device, e.g., for information services such as voicemail systems or online media libraries. Similarly, a universal interaction device might be the only means for users to discover the invisible information services available in a smart room or attached to a smart object. And obviously, as the success of the TV remote control has shown, handhelds should be well suited to control appliances where interaction at a distance is desirable, such as a heating control system. However, it is less clear whether personal mobile devices are beneficial for interacting with physical appliances that require the user's presence to operate, such as ATMs, elevators, or microwave ovens.

With this in mind, we set out to conduct a user study exploring the benefits and limits of using a mobile phone to operate physical appliances, i.e., devices that would typically not benefit from being remotely controllable. Our aim was to identify in particular the conditions under which devices like coffee makers, printers, or microwave ovens would benefit from being operated not directly, but through a mobile phone, or, conversely, when it would be a hindrance, rather than an advantage, to have such a separate interaction device.

The remainder of this paper presents our user study in detail (Section 2), starting with the experimental design and participants, describing our apparatus and materials, and outlining our procedure. Section 3 contains the results of our study, both analytically and anecdotically. We close with a discussion and conclusions.

2 User Study

Our study tried to assess the benefits and limits of handheld devices in appliance operations by asking study participants to use a range of appliances in a variety of situations, both traditionally using the appliances' physical interface, and with a specifically developed universal interaction device. We then obtained quantitative data by measuring the time it took a test subject to complete a specific task, as well as qualitative data by observing and recording the users' actions, thoughts (through think-aloud techniques [24]) and opinions (through a post-test questionnaire). This section presents the hypotheses, tasks, and procedure of our study, including a description of our prototypical universal interaction device, the AID (short for "Appliance Interaction Device").

2.1 Hypotheses

We began our project with a set of three hypotheses, which together would either support or weaken our intuitive notion that the use of universal interaction devices has limits. Specifically, we hypothesized as follows:

- For controlling an appliance in exceptional situations, interaction based on a mobile phone would be faster than interaction based on the traditional user interface.

Table 1. *Appliance tasks in the user study.* Participants were asked to complete 18 tasks distributed among the four available appliances. Not all appliances had suitable tasks in all four categories.

	Control	**Problem solving**	**Everyday**	**Repeated control**
Dishwasher	adjust water hardness, activate child safety lock	fix error "F2", white film on dishes	start program	—
Coffee maker	adjust water hardness, set switch-on time	—	brew coffee	adj. water hardness
Printer	change paper type, print cleaning page	fix faded print	cancel print job	change paper type
Radio	set clock, store preset station	—	change channel	set clock

- Looking up context-dependent information on the handling of an appliance would be faster using a mobile phone than using traditional means (e.g., user manuals).
- To carry out everyday tasks, the use of an appliance's traditional user interface would be faster than mobile phone-based interaction.

2.2 Appliances and Tasks

We used four typical appliances for which we found a number of use cases where mobile phone based interaction might be beneficial: a dishwasher (V-ZUG Adora S 55), a coffee maker (Jura Impressa S70), a printer (HP LaserJet 4050tn) and a radio (Sangean ATS-505). For each appliance, we defined a number of tasks for participants to work through – once using the appliance's native controls, once using our AID device. We grouped these tasks (18 in total, listed in Table 1) into the following four categories:

- *Control tasks* involve the adjustment of a special device setting (e.g., setting the water hardness for the coffee maker) or the invocation of an unusual operation (e.g., create a printer cleaning page). Theses tasks reflect the use of a mobile phone for user interface provision.
- *Problem solving tasks* confront users with an abnormal situation (e.g., a malfunctioning dishwasher displaying an error code) that must be dealt with. These tasks correspond to the use of a mobile phone for information provision.[2]
- *Everyday tasks* are control tasks that are most typical for an appliance (e.g., brewing a coffee) and are performed very often.
- *Repeated control tasks* are control tasks that a user has performed very recently and is still very familiar with.

Using our prototype presented in [25], a small pilot study was run in advance to identify tasks suitable for comparison. For example, the coffee maker's manual

---

[2] See Section 1, page 199.

contained instructions on how to brew a latte macchiato, which required the user to adjust various settings in no less than 10 steps. Obviously, brewing a latte macchiato could simply be offered on the mobile device as a single menu item. As this would have drastically favored the AID device, we took care to select only tasks that would require a comparable degree of interaction when executed directly on the appliance and on the mobile interaction device. Similarly, we omitted tasks that were so poorly supported by the appliance manufacturer that they proved very difficult and lengthy when tested in our pilot study, e.g., changing the coffee maker's water filter. As these tasks could be improved easily (e.g., through better documentation), we did not consider them for our main study.

2.3 Participants

We tested a total of 23 participants,[3] 10 (43%) of which were male and 13 (57%) of which were female. Most of them were undergraduate or graduate students recruited through mailing lists from different universities. There were 10 participants with a background in sciences or engineering, 10 participants from the humanities or social sciences, and 3 participants with no academic background. All of them spoke German as their native language and owned a mobile phone. Except for two subjects, all participants used their mobile on a daily basis. The average age of participants was 29.8 years, ranging from 21 to 50 (SD=6.1) years. None of them had any relevant previous experience with the appliances used in our experiment. Participants were compensated for their time with a USB memory drive or a cinema voucher, according to their preferences.

2.4 Apparatus

In order to evaluate the usefulness of a universal interaction device, we had to provide our test subjects with a corresponding mobile unit that would let them properly experience the use of such a remote control unit. Our AID prototype system supports mobile phone based interaction for all of the tasks outlined above. Our system is implemented as a Java MIDlet that we deployed on a Nokia 6630 mobile phone. This mobile phone features a color display with a resolution of 176 x 208 pixel. Unlike the devices used in other evaluations (see Section 1 above), the Nokia 6630 does not offer pen-based input capabilities, but features only a simple keypad.

Instead of actually coupling the AID with our four appliances, we simulated both appliance identification, as well as the transmission of appliance status, by having the user press the phone's main navigation button. As users performed the tasks in separate steps, the experimenter had time to use an undocumented button combination on our AID to quickly select the proper context-dependant

---

[3] Our pilot study included 9 participants, none of which participated in our main study later.

reactions for the upcoming task, thus giving users the illusion of having our AID actually detect and read out the appliance.[4] Obviously, this setup made it impossible to really *control* the appliance in any way through the AID – a shortcoming that we pointed out to participants, indicating that we were only interested in seeing the right buttons being pressed, not an actual appliance being controlled.

(a) Radio main menu (b) Adj. water hardness (c) Dishwasher user manual (d) Printer troubleshooter

Fig. 1. *AID prototype implementation.* The screenshots show examples of the AID user interface for each of the four appliances used in the study. Appliance functions could either be selected by traversing the main menu (a) or by having the AID automatically (i.e., simulated, see Section 2.4) detect the appliance's state (d).

Figure 1 shows four example screenshots of the AID during the study, one for each of the four appliances. Invoking our AID device on an idle appliance brings up the appliance's main menu, as shown in Figure 1(a). For each appliance, this main menu would offer all tasks that are available through the appliance's physical interface (in a hierarchical menu). The user interface for a typical task is shown in Figure 1(b). We also included a "troubleshooting" menu entry for each task, which would contain the contents found in the corresponding section of the appliance's user manual. Figure 1(c) shows such a list of common problems that might occur at the dishwasher. Finally, we provided several step-by-step assistants that help users with physical appliance manipulations. The assistant supporting the task of clearing a paper jam at the laser printer is illustrated in Figure 1(d). For each step, the system highlights the part of the printer the user must operate next. When the user pushes the right arrow on the phone's keypad, the assistant advances to the next step. Assistants and troubleshooting

---

[4] The device used in our pilot study [25] was an NFC-enabled Nokia 3220 that actually performed a true wireless identification of each appliance. However, as both the computational resources and the screen estate of the Nokia 3220 were limited in comparison with the Nokia 6630, and because no discernable difference was noticeable between the real and simulated NFC action, we decided on the setup described above.

Fig. 2. *Participants interacting with appliances.* The images above show three of our participants carrying out tasks: using the AID device to operate the coffee maker; troubleshooting the printer using the printed manual; and using the AID device to operate the radio.

tips can either be accessed manually through the menu, or they are displayed automatically when the appliance is in a state where such help is needed.[5]

2.5 Procedure

The experiment began for each participant with a brief introduction to the concept of the AID, our user study, and its goals. Participants were then asked to fill out a profile questionnaire that allowed us to gather information on their background (age, education, previous experience with devices used in the experiment, etc.). We explained the basic concepts of the AID (i.e., the user interface provision and the information provision) and demonstrated it interactively using example tasks that did not reoccur in the course of the study. We also told participants explicitly not to think of the AID as a remote control, and that they would only be able to use it on an appliance after they had touched it, followed by pressing the phone's main navigation button. We finally handed them the AID device and guided them through a number of simple preparation tasks in order to familiarize them with the phone's controls and user interface.

The beginning of the study was conducted in a semi-public lounge in our university, as it offered a dishwasher that we could use. There were only few distractions here, and for the rest of the study we moved to a quiet office where we had set-up the coffee maker, the laser printer, and the radio. The introduction and initial explanations described above were also conducted in this office. All devices where ready to use and equipped with the appropriate handbook in German language.

For each of the four appliances, we handed the test subjects small cards with their assignment printed on. They were asked to work through each task twice.

---

[5] As pointed out previously, this appliance state detection would be setup secretly by the experimenter prior to giving out a particular task to the user (as no communication between the appliance and our AID device takes place).

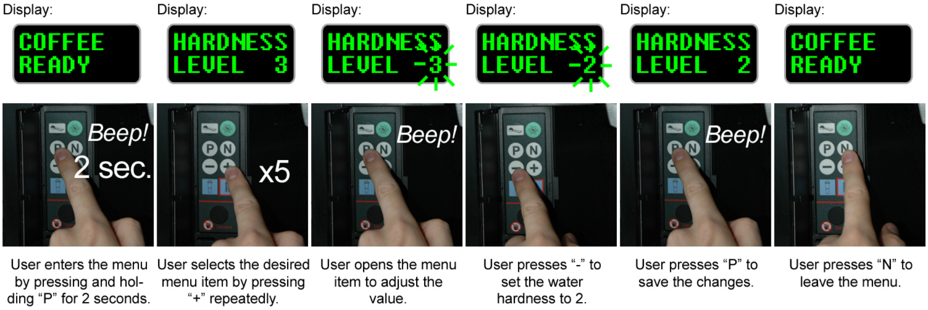


Fig. 3. *Example of traditional task solving.* At one point, study participants were asked to change the level of water hardness in the coffee maker. The above pictured steps usually required an extensive study of the printed manual.

One time, users should use the traditional method to solve the task, i.e., they should interact directly with the device using the physical interface. The other time, they should use the AID device. Users were explicitly told that they could, but would not have to use the user manual when completing a task in the traditional way. In order to minimize potential learning effects, we used counterbalancing, i.e., participants were divided into two subgroups that worked through the cards in different orders. Group A was asked to complete every task first with the traditional method, and then again with help of our AID device. Group B was asked to begin with the AID device and then use the traditional interface afterwards. To get comparable results, the order of the tasks was the same for all participants.[6] Learning effects were compensated for by letting the users perform the first task of each appliance again at the end of the study.[7] We measured the time needed for each task and then asked users whether it was easy or difficult for them to solve the task, as well as which method they liked better. For tasks that required users to find a solution to a problem and telling us about it, time stopped with their first utterance. Figure 2 shows some of our participants carrying out tasks, Figures 3 and 4 show an example of an entire task (adjusting the water hardness in the coffee maker) being done using the traditional method and using the AID device, respectively.

The final part of our study asked our participants to complete seven different tasks in a row (see Figure 6 on page 210 for a list), which they had to perform as fast as possible. They were, however, free to chose any method to operate the appliances, i.e., they could pick up the AID, consult the printed manual, or directly operate the physical appliance interface to accomplish the assignment. We then recorded their individual choices for each of the seven tasks. In order to

---

[6] The actual task order can be found in Table 1: for each device, the control, problem solving, and everyday tasks were performed. Finally, the column "repeated control" was tested from top to bottom.

[7] For practical reasons, we did not move back to the lounge area again to test the dishwasher a second time.

«Water Hardness» was saved successfully.

User touches the appliance with the mobile phone.

The menu shows up automatically. User selects "Settings".

User selects "Water Hardness".

User selects the desired value and saves the changes with "OK".

Setting is changed on the appliance.

Fig. 4. *Example of task solving using the AID.* The water hardness can be changed more easily using the hierarchical menu of the AID, typically without consulting any manual.

get participants to truthfully choose the methods they thought would be most effective, we offered a portable MP3-player to the user with the fastest time.

The study ended for each participant with a final questionnaire that collected their opinion and suggestions on the AID device. A single session typically lasted about 80-120 minutes.

3 Results

We collected both qualitative and quantitative results. Qualitative results used both the explicit answers from a post-test questionnaire administered after all tasks were completed (reported in Section 3.2), as well as notes and recordings taken during the tasks that captured the participants' thoughts through a think-aloud technique, which they were instructed to employ (see Section 3.3). The quantitative results simply measured the time between the begin and end of a task, though for tasks involving an answer rather than asking the user to operate an appliance (i.e., the problem solving tasks, see Table 1 on page 202), the time until the user started to reply was measured. This data is reported in the following section.

3.1 Quantitative Results

Figure 5 and Table 2 show the average task completion times we measured in each condition. As predicted, the mean time of task completion decreased for control tasks and problem solving tasks, whereas it increased for everyday tasks.

We further examined the collected data using analysis of variance (ANOVA). We ran a two-way ANOVA with factors interaction method and task type, which confirmed a significant main effect of interaction method ($p < .001$, $F_{1,736} = 100.23$). Also, a significant main effect of task type ($p < .001$, $F_{3,736} = 661.18$) and a significant interaction effect between task type and interaction method ($p < .001$, $F_{3,736} = 218.68$) was found. Focusing on this interaction effect, we continued by

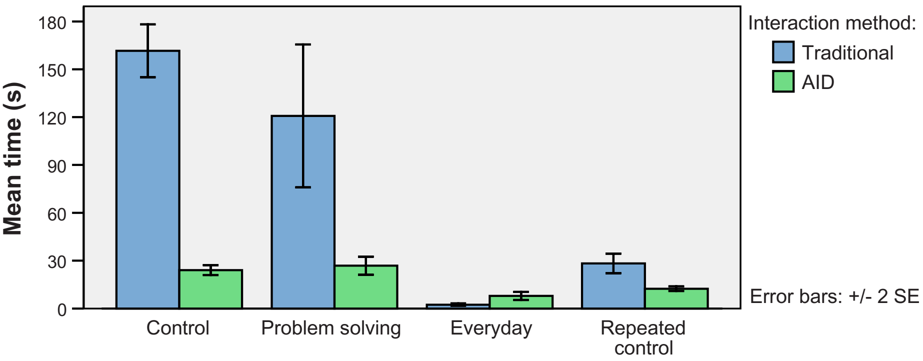


Fig. 5. *Mean time of task completion.* While the use of the AID device significantly cut down the execution time for exceptional control and problem solving tasks, it was two to four times slower for everyday tasks.

analyzing the effects of interaction method for each of the four task types. For each task type, an ANOVA showed that there was a significant difference between interaction method (control tasks: $p < .001, F_{1,366} = 600.1$; problem solving tasks: $p < .001, F_{1,64} = 34.823$; everyday tasks: $p < .001, F_{1,182} = 198.302$; repeated control tasks: $p < .001, F_{1,124} = 44.851$). We therefore find our hypotheses confirmed that mobile phone based interaction significantly reduces completion times for control and problem solving tasks, while significantly slowing down everyday tasks. Interestingly, controlling an appliance with the mobile phone was significantly faster even after participants had familiarized themselves with the task and could be considered experienced users of the respective appliance.

We also studied the effects of other factors, namely age, gender, and experience with advanced phone features. We did not find an interaction effect with interaction method for any of these factors. We therefore conclude that, for the 23 participants in our study, the use of the AID was beneficial irrespective of their gender, age, or previous mobile phone experience.

An individual analysis of our 18 tasks showed a consistent pattern. For each of the everyday tasks, traditional interaction proved faster, while users were faster using the AID in all other cases. However, there was a single exception from this picture, namely the repeated control tasks performed on the printer. While interaction with the AID took slightly less time ($M_{AID} = 16.9s, M_{trad} = 21.6s$), this difference fell short of significance at the .05 level ($p = .177, F_{1,40} = 1.885$). We presume that this is due to the the well-designed user interface of the printer used in our study. Several users made the informal comment that they liked the layout of the printer's control panel and that it was relatively easy to navigate in its menu because buttons were labeled in a helpful and familiar way.

Apart from the repeated control tasks, we found the results of the problem solving tasks particularly interesting. While in one task the appliance supplied context (i.e., an error code) that the AID used to automatically show the relevant instructions, it did not do so in the two other repeated control tasks. We merely

Table 2. *Mean time (in seconds) and standard deviation.* See Figure 5 for a graphical representation of this data.

	Control	**Problem solving**	**Everyday**	**Repeated control**
Traditional	162 (112)	121 (131)	2 (4)	28 (24)
AID	24 (21)	27 (16)	8 (12)	12 (6)

gave participants some unspecific information about a problem that they had to find manually using either the printed user guide or the AID. Even though we made sure the AID covered all the topics we had found in the printed documentation, participants were significantly faster when they used the AID (printer: $M_{trad} = 253s, M_{AID} = 25s$; dishwasher: $M_{trad} = 91s, M_{AID} = 29s$). This is surprising especially for the dishwasher as its user manual is very compact and, as we find, well made. However, the mobile phone's menu hierarchies that allow for the structuring of content seem to prove beneficial for this task.

3.2 Qualitative Results

Figure 6 summarizes the results of the final contest, in which participants had to solve a list of tasks in the shortest possible time, but for which they could freely choose the interaction method. Overall, the AID was used in 69% of the control and problem solving tasks, even when participants had previously used the traditional user interface in similar situations just as effectively as the AID. Most of the participants stated that already slightly different tasks would make them insecure, fearing that they would not know where to find it in the appliance's menu structure. They would therefore opt for the AID as it "*gives me a better overview than the printer's two-line display and allows me to complete the task faster*", as one participant explained. On the other hand, participants hardly used the AID to store or select a station in the radio's preset memory, stating that they didn't even think of using the AID as it was "*just more natural*" to interact with the radio directly.

In the post-test questionnaire, we asked users to rate a number of statements on a five-point Likert scale. The statements and participant responses are shown in Figure 7. We also asked users to rate the value of the AID in some given situations. The results of this question are depicted in Figure 8.

Finally, every participant was asked to answer three concluding questions. At the outset, we asked them: "*Do you use appliances for which you would like to have the AID? If so, please name these appliances.*" 22 participants (96%) replied positively and listed both devices from our study, as well as microwave ovens, TVs, DVD players, car radios, heating systems, and "*appliances that you change often, such as rental cars*". Only 1 participant (4%) would not want our AID. We also asked: "*If you owned an AID for an appliance, would it be ok if you could access rarely used functions* only *through the AID? Why / why not?*" 18 (78%) of our users agreed, 4 (17%) disagreed, and 1 participant was not sure. Some participants expressed concern that they would no longer be able to interact with their appliances in case their mobile phone or PDA was unavailable.

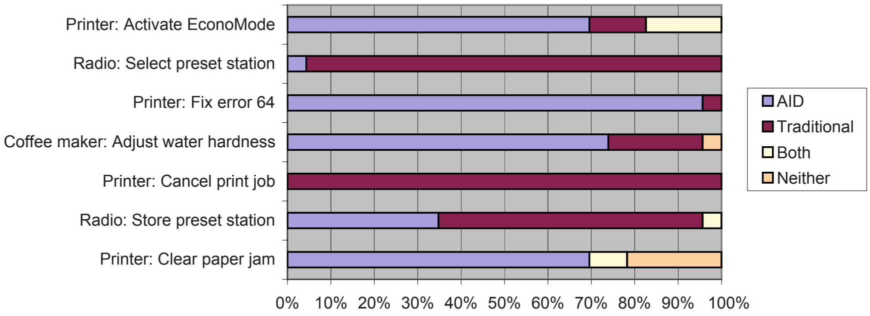


Fig. 6. *Interaction method usage during contest.* For the final contest, participants were free to use any method that they felt most comfortable with for each taks. Out of the seven tasks, four were previously completed tasks, while three ("Activate EconoMode", "Fix error 64", "Clear paper jam") were new tasks.

We ended our questionnaire with the following question: "*Could you imagine accessing* all *functions offered by an appliance only through the* AID*? Why / why not?*" While some participants could imagine giving up the traditional interface (4 users, i.e., 17%), most replied that a software-only user interface was not an option for them (17 participants, i.e., 74%). 2 participants were unsure. Most users answered that they would rather not use the AID for simple tasks. However, two respondents stated that they might be willing to accept an appliance without a traditional user interface if its price were lowered in turn. Two participants added that they could think of appliances that would benefit from a full replacement of the user interface, as this would make ugly control panels redundant and improve visual appearance.

3.3 User Feedback

In the course of the study, there were a number of issues that were brought up by the participants. In this section, we review these informal comments.

The biggest concern that users expressed was that of increasing dependence on technology, and in particular the mobile phone, through the use of an AID device. Most often, participants wondered how they would use their appliances if their phone was misplaced, lost, stolen, malfunctioning, without network coverage, or had simply run out of battery. Two participants also mentioned that they did not want to always carry their phone with them, just to be ready to use their appliances. Other concerns were more diffuse: "*I don't fully trust the mobile phone and would like to have the buttons on the device. Just in case...*" Someone else perceived the AID as yet another burden in everyday life: "*We've got so many devices around us already, so I don't think we need another one just for rarely used functions.*"

What most participants pointed out as the AID's biggest advantage over traditional interfaces was its menu structure. It was described as "*easy to use*", "*well structured*", "*clearly laid out*" and "*intuitive*". Many participants felt that

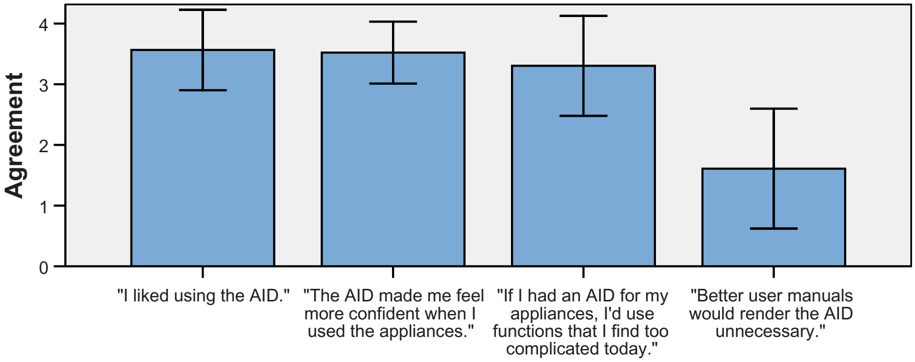


Fig. 7. *Subjective Likert scale responses.* Participants rated the usefulness of the AID device after they finished all tasks (0 = strongly disagree, 4 = strongly agree) – Error bars: +/- 1 SD.

this was mainly due to the larger display size, especially when compared to a small appliance LCD. Several participants explained that they would hardly get lost in the menu, thus enabling them to easily find even previously unknown functions. One participant said: "*If I had an AID, I could forget about how I did something and still be able to do it a few months later without any problems.*" Another participant explained how the AID's menu system allowed her to interactively explore the full range of functions: "*I always like to gain an overview of all available functions of my appliances. The AID would be ideal for that.*"

While they liked our prototype implementation of an AID, some participants expressed doubt if an actual appliance manufacturer would be able to come up with an equally user-friendly piece of software for their own products. One participant suggested that "*every manufacturer should include a menu item called 'troubleshooting', just like all Windows applications have a 'Help' menu.*" Many users especially liked the immediate reaction of the AID to special situations, e.g., by displaying helpful instructions instead of just a status code. As one participant put it: "*It is also extremely convenient that I'm immediately told what the cause of an error is.*" However, there were also cases where context-sensitive behavior confused people more than it helped, as they expected the normal menu to show up.

Most participants expressed their frustration with user manuals. As one user said: "*[Most handbooks contain] too much information that I don't actually want to read.*" They therefore found it very helpful to have the essential problem- or task-oriented information available on the AID and were not bothered by its relatively small display size. We asked some participants if a well-made quick-reference guide could make this content obsolete on the AID. Most of them agreed that this was not the case: "*I don't want to go find a manual that I anyway may have thrown away.*" Two participants also said that the AID could be improved by adding search capabilities, because "*then I can search for keywords with CTRL+F just like when I read a manual in PDF format.*"

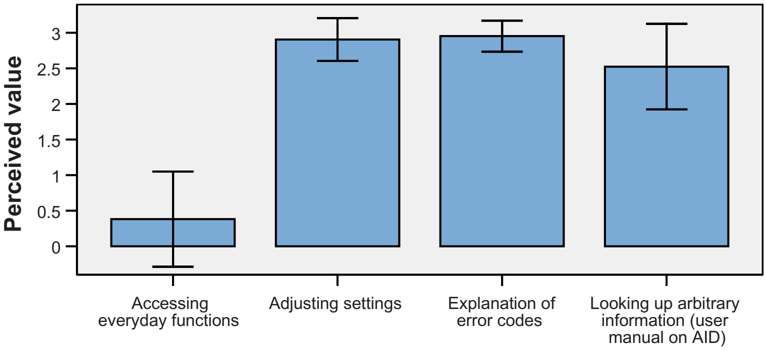


Fig. 8. *Concluding questions.* Participants were asked to rate the perceived value of the AID device in various situations (0 = no value, 3 = great value) – Error bars: +/- 1 SD.

Overall, while having some concerns on the ever-increasing dependence on technology, participants generally liked the AID and gave very positive feedback on its use. In one participant's words: "*In the beginning I was sceptical, but now I'm very excited.*"

4 Discussion

At first sight, our measurements seem to disagree with the results of the user study by Nichols et al. [22], which showed that their participants took twice as long interacting with the actual physical interface even for those everyday tasks that required only "one or two button presses on the actual appliance". Nichols et al. attribute this superiority of the handheld approach to the "poor labeling, insufficient feedback, and overloading of some buttons with multiple functions" of the appliances under test. With respect to the AT&T phone they used for some of their experiments, Nichols et al. mention that "this phone has several buttons that can be pressed and held to activate another function. There is no text on the telephone to indicate this". Nichols et al. specifically addressed these drawbacks of the physical interface in the handheld implementation by providing an intuitive virtual interface, which explains the superiority of the handheld approach. In our experiments, the physical user interface on the four appliances is very intuitive and convenient for all four different everyday tasks. In the paper by Nichols et al. there is also little information on the actual tasks that were performed. The result is that we cannot compare the "one or two button"-tasks mentioned by Nichols et al. with the everyday "one button"-tasks we studied.

In our opinion, the value of the mobile phone to complete exceptional tasks stems from the shortcomings of the physical user interfaces and the corresponding manuals. User interface designers commonly have to deal with the conflicting constraints of cost, size, and usability. On the one hand, expensive high-resolution displays with many buttons help to build a user interface that is convenient and intuitive even for uncommon tasks. On the other hand, the cost and size restrictions typically limit the number of buttons available and the size and resolution of the displays. For most appliances, the result is that interface designers are forced to realize uncommon tasks by a combination of buttons. The corresponding instructions are then listed in a complementary manual. This requires a significant effort from the user, however. The manual might not even be close-by, there is no direct link to the corresponding section in the manual, and the limited expressiveness of a manual makes the association of the printed instructions with the actual physical interface of the appliance non-trivial. The AID addresses all of these shortcomings, since it represents a cost-effective way to equip any appliance with a high-resolution display and a multitude of buttons. It thus enables user interface designers to address the conflicting constraints of usability, cost, and size associated with user interface design by leveraging the external mobile phone with its significant display and input capabilities.

The AID provides not only a high resolution display and input capabilities, it also permits the personalization of the user interface due to the personal nature of the mobile phone. While designers usually have to develop an interface for the average user, the AID concept allows them to build a software interface that automatically adapts the language, adjusts the features available according to the capabilities of the user, and lists the history of recently performed tasks. The user interface designer would also benefit from the availability of well-supported software development platforms, such as J2ME or Symbian. The long range communication capabilities of the mobile phone might also facilitate software maintenance, since the user interfaces can be updated remotely.

In our study, the appropriate mobile phone software was preloaded on the mobile phone and we simulated the appliance identification with a mobile phone. The rationale for simulating the identification was the limited display and computing capabilities of today's NFC-enabled mobile phones and the fact that our pilot study indicated that there was no discernable difference between real and simulated NFC action. As the NFC technology becomes available in mobile phones with high-resolution displays, future user studies could incorporate the appliance identification and possibly also the downloading of the software. While we do not believe that this will impact the findings of our study, it would make the overall application scenario even more realistic.

5 Conclusion and Outlook

The goal of our study was to assess the benefits and limits of using mobile devices, such as PDAs or mobile phones, for appliance control. While the idea of a *universal remote control* is an appealing one, given the technical capabilities and prevalence of such devices, we questioned the sheer limitless uses that today's designers and researchers often envision for them. Hypothesizing that universal appliance controllers might be superior to traditional, physical appliance interfaces in *exceptional* situations only, but not for carrying out *everyday* tasks, we had our 23 test subjects perform a series of 18 tasks distributed among four appliances. By collecting quantitative measurements, we could confirm that our users were significantly faster when having to solve exceptional tasks with our AID (our prototypical universal appliance controller), but slower when performing everyday tasks. Our qualitative methods further confirmed that users would often prefer using the AID, but still liked the "natural interaction" with a device if a simple, straightforward task was to be solved and the tasks required them to be in the vicinity of the appliance anyway.

These findings seem to suggest that hybrid approaches that combine traditional, haptic user interfaces with extended user interfaces on a mobile device, offer the best of both worlds. Users could continue to directly interact with appliances, which is both faster and more convenient in most everyday situations. However, in special situations where users would have to remember complex and clumsy sequences of pushing buttons or manipulating the appliance, it is much more intuitive to use a mobile device with its powerful and versatile user interface for interaction. The results thus suggest that the proliferation of mobile phones with high resolution displays and short range communication capabilities will enable appliance manufacturers to overcome the conflicting constraints on cost, size, and usability of the user interface, by leveraging the user's mobile phone.

References

1. Kindberg, T., Barton, J., Morgan, J., Becker, G., Caswell, D., Debaty, P., Gopal, G., Frid, M., Krishnan, V., Morris, H., Schettino, J., Serra, B., Spasojevic, M.: People, Places, Things: Web Presence for the Real World. Mob. Netw. Appl. **7**(5) (2002) 365–376
2. Griswold, W.G., Shanahan, P., Brown, S.W., Boyer, R., Ratto, M., Shapiro, R.B., Truong, T.M.: ActiveCampus: Experiments in Community-Oriented Ubiquitous Computing. Computer **37**(10) (2004) 73–81
3. Intille, S.S.: Designing a Home of the Future. IEEE Pervasive Computing **1**(2) (2002) 76–82
4. Davies, N., Gellersen, H.W.: Beyond Prototypes: Challenges in Deploying Ubiquitous Systems. IEEE Pervasive Computing **1**(1) (2002) 26–35
5. Carter, S., Churchill, E., Denoue, L., Helfman, J., Nelson, L.: Digital Graffiti: Public Annotation of Multimedia Content. In: CHI '04: CHI '04 Extended Abstracts on Human Factors in Computing Systems. (2004) 1207–1210
6. Smith, M.A., Davenport, D., Hwa, H., Turner, T.: Object AURAs: A Mobile Retail and Product Annotation System. In: EC '04: Proc. of the 5th ACM Conf. on Electronic Commerce, ACM Press (2004) 240–241
7. Rohs, M., Roduner, C.: Camera Phones with Pen Input as Annotation Devices. In: Pervasive 2005 Workshop on Pervasive Mobile Interaction Devices (PERMID). (2005) 23–26
8. Myers, B.A., Stiel, H., Gargiulo, R.: Collaboration Using Multiple PDAs Connected to a PC. In: CSCW '98: Proc. of the 1998 ACM Conf. on Computer Supported Cooperative Work. (1998) 285–294
9. Ballagas, R., Rohs, M., Sheridan, J.G., Borchers, J.: BYOD: Bring Your Own Device. In: UbiComp 2004 Workshop on Ubiquitous Display Environments. (2004)

10. Beigl, M.: Point & Click - Interaction in Smart Environments. In: HUC'99: First Int. Symposium on Handheld and Ubiquitous Computing, Springer-Verlag (1999) 311–313
11. Myers, B.A., Nichols, J., Wobbrock, J.O., Miller, R.C.: Taking Handheld Devices to the Next Level. Computer **37**(12) (2004) 36–43
12. Hodes, T.D., Katz, R.H., Servan-Schreiber, E., Rowe, L.: Composable Ad-Hoc Mobile Services for Universal Interaction. In: MobiCom '97: Proc. of the 3rd Annual ACM/IEEE Int. Conf. on Mob. Comp. and Networking, ACM Press (1997) 1–12
13. Ponnekanti, S., Lee, B., Fox, A., Hanrahan, P., Winograd, T.: ICrafter: A Service Framework for Ubiquitous Computing Environments. In: UbiComp '01: Proc. of the 3rd Int. Conf. on Ubiquitous Computing, Springer-Verlag (2001) 56–75
14. Olsen, D.R., Jefferies, S., Nielsen, T., Moyes, W., Fredrickson, P.: Cross-modal interaction using XWeb. In: UIST '00: Proc. of the 13th Annual ACM Symposium on User Interface Software and Technology, ACM Press (2000) 191–200
15. Nichols, J., Myers, B.A., Higgins, M., Hughes, J., Harris, T.K., Rosenfeld, R., Pignol, M.: Generating Remote Control Interfaces for Complex Appliances. In: UIST '02: Proc. of the 15th Annual ACM Symposium on User Interface Software and Technology, ACM Press (2002) 161–170
16. Zimmermann, G., Vanderheiden, G., Gilman, A.: Prototype Implementations for a Universal Remote Console Specification. In: CHI '02: CHI '02 Extended Abstracts on Human Factors in Computing Systems, ACM Press (2002) 510–511
17. LaPlant, B., Trewin, S., Zimmermann, G., Vanderheiden, G.: The Universal Remote Console: A Universal Access Bus for Pervasive Computing. IEEE Pervasive Computing **3**(1) (2004) 76–80
18. Rukzio, E., Leichtenstern, K., Callaghan, V., Holleis, P., Schmidt, A., Chin, J.: An Experimental Comparison of Physical Mobile Interaction Techniques: Touching, Pointing and Scanning. In: UbiComp '06: Proc. of the 8th Int. Conf. on Ubiquitous Computing, Springer-Verlag (2006)
19. Ailisto, H., Pohjanheimo, L., Välkkynen, P., Strömmer, E., Tuomisto, T., Korhonen, I.: Bridging the physical and virtual worlds by local connectivity-based physical selection. Personal Ubiquitous Comput. **10**(6) (2006) 333–344
20. Koskela, T., Väänänen-Vainio-Mattila, K.: Evolution towards smart home environments: empirical evaluation of three user interfaces. Personal Ubiquitous Comput. **8**(3-4) (2004) 234–240
21. Rode, J.A., Toye, E.F., Blackwell, A.F.: The fuzzy felt ethnography-understanding the programming patterns of domestic appliances. Personal Ubiquitous Computing **8**(3-4) (2004) 161–176
22. Nichols, J., Myers, B.A.: Studying the Use of Handhelds to Control Smart Appliances. In: ICDCSW '03: Proc. of the 23rd Int. Conf. on Distributed Computing Systems, IEEE Computer Society (2003) 274–279
23. Norman, D.: The Design of Everyday Things. 2nd edn. Basic Books (2002)
24. Lewis, C., Rieman, J.: Task-Centered User Interface Design – A Practical Introduction. University of Colorado (1993)
25. Roduner, C.: The Mobile Phone as Universal Interaction Device – Are There Limits? In: MobileHCI '06 Workshop on Mobile Interaction with the Real World. (2006)

Portable, But Not Mobile: A Study of Wireless Laptops in the Home

Allison Woodruff[1], Ken Anderson[2], Scott D. Mainwaring[2],
and Ryan Aipperspach[1,3]

[1] Intel Research Berkeley, 2150 Shattuck #1300, Berkeley, CA 94704 USA
`woodruff@acm.org`
[2] Intel Research, People and Practices Research Lab, 7144 SW 29th Ave
Portland, OR 97219
`{ken.anderson, scott.mainwaring}@intel.com`
[3] Berkeley Institute of Design, Computer Science Division,
University of California, Berkeley, Berkeley, CA 94720 USA
`ryanaip@cs.berkeley.edu`

Abstract. We report a qualitative study of the use of physical space and wireless laptops by ten United States households. Although wireless laptops purportedly offer the opportunity and affordances to go "anywhere in the home," laptops were generally used in a small set of particular places rather than moving fluidly through the home: wireless laptops were portable, but not mobile *per se* We present factors that influence laptop movement in the home. We also present a model of people's use of space in the home, identifying a small set of favored places of long-term use and a larger set of kinetic places used for specific tasks. We discuss how the factors we have identified generally promote use of laptops in favored places and generally discourage use of laptops in kinetic places. We discuss how our findings are relevant to the design of technologies for the home.

1 Introduction

A number of significant studies have examined the use of personal computers in the home [9,10,14,17,28,29]. Two increasingly popular technologies have the potential to dramatically change patterns that have been reported previously, particularly when taken in combination: wireless home networks and laptops. Forrester estimates that 3.8 million United States homes had both home networks and laptops as of December 2004, forecasting 30 million such homes by 2010 [25], and Strategy Analytics estimates that 19 million United States and European homes have wireless networks as of 2006 [19].

The wireless laptop offers continuous connectivity in a portable device, and while it certainly falls short of the original ubicomp vision [30], it is an important step towards it. The wireless laptop is touted as having great potential to change the nature of computing in the home. Rhetoric suggesting that people can compute "on a whim at any time and in any room" is extremely common [25]. Nonetheless, little is known about specific day-to-day practices with wireless laptops.

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 216 – 233, 2007.

The much vaunted opportunities for the wireless laptop (and more broadly for mobile, ubiquitous, and pervasive computing) are strongly tied to its potential to be used in a variety of physical contexts. Physical space plays a large role in people's day-to-day lives, influencing for example sensory, ergonomic, cognitive, and social experience. We are therefore interested in examining the physical contexts in which people choose to spend time and use technology in the home. We believe that by studying current practices around choice of physical context, we can gain insights (1) for the design of future technologies and (2) for the design of physical environments (from furniture to architecture) amenable to emergent behaviors and technology use.

More specifically, we are interested in questions such as: Where specifically do people spend time in the home, and why? What are the properties of different places in the home? In which of these places do people use computing technology, and why? Through this investigation, we hope to learn more about which characteristics of space and devices are or are not favorable to computing technology in the home.

Households with wireless laptops are an excellent population to inform such questions, as household members have the opportunity to compute "anywhere" in the home and have well-established practices that can be examined (in contrast with participants in short-term interventions or laboratory studies). In this paper, we report a qualitative study of the use of physical space and wireless laptops by ten United States households. To our knowledge, this is the first study of the day-to-day use of wireless laptops in the home. Specifically, we discuss factors that influence where wireless laptops are and are not used in the home. We characterize participants' use of space in the home and discuss the settings in which laptop use is and is not prevalent in the home. Wireless laptops are not generally used anytime, anywhere, but are instead used mainly in four kinds of favored places in the home. We present a taxonomy of these primary sites of laptop use in the home. We believe our findings can be used to reason about how future devices may be used in the home.

The remainder of the paper is organized as follows. We first review related work and then discuss our method. We then turn to findings, followed by discussion of the findings and design implications. Finally, we conclude and discuss future work.

2 Related Work

The home has long been considered a compelling domain for ubiquitous computing, and many exciting technologies have been proposed for this venue [13]. The home is however plainly a complex domain for which it is difficult to design, and the promise of many technologies is as yet unfulfilled. Even when domestic technologies succeed, the reasons for their success are not always well understood. While important initial work has been done in the domestic arena, researchers have lamented the paucity of research on the home environment as compared to the extensive body of literature on the work environment, and have called for more research to deepen our understanding of home life and the use of technology in the home, e.g., [4,6].

One of the main areas of inquiry has been the use of desktop computers and the Internet in the home [9,10,14,17,28,29]. These studies have tended to focus on the social context of computing [10]; Venkatesh and his colleagues explicitly model the home as two main components – the social space and the technological space [28,29].

However, some work has also considered physical space, typically at the level of rooms or general areas. Mateas *et al.* conducted a study of the use of desktop computers in the home, explicitly considering not only social context, but spatial and temporal context as well [17]. They identified general areas (although not specific places) that represented "behavioral clusters" in the home, such as "Work Space" and "Private Space." Frohlich *et al.* also investigated the location of desktop computers in the home [9]. In all these studies, the desktop computer was statically positioned, frequently in the home office if the family had one. Frohlich and Kraut [10] observe that placement of the computer has complex social implications and can create social tension, and they raise the possibility of portable machines that can be carried between different rooms.

The suggestion of portable devices resonates with O'Brien *et al.*'s study of home life and the use of audio-visual technology and a prototype set-top box [21]. This work focuses primarily on the social context for the use of technology, although this naturally leads to discussion of issues such as coordination and ownership of space, particularly the way in which noise-emitting technologies such as the television can mark out space. They discuss the importance of the portability of devices such as small televisions and stereos to distributing functionality throughout the home.

McClard and Somers present a rare study of the use of portable tablets [18], reporting a seven-week intervention in which they provided wireless networks and tablet computers to 13 families. Their work strongly emphasizes that the tablet can be used "anywhere." (On a related note, Forrester reports survey data indicating that laptops are used to some degree in every room in the home, although most often in the office [25].) Although participants did not consider the tablet to have equivalent functionality to the PC, they liked the fact that it was portable and could be used in "comfortable" positions and locations such as a couch or bed. Despite the significance of issues such as fatigue and comfort in the home, and their impact on the use of technology, seating position and posture are rarely considered elsewhere in the literature. In one of the only other studies of wireless computing, Grinter *et al.* [11] explored the collaborative administration of home networks, some of which were wireless. Patel *et al.* [23] studied the use of a different wireless technology, cell phones, examining the proximity of the phone to its owner, both in and out of the home.

Issues such as communication, coordination, and organizing systems have also been explored in the home [5,7,27]. Crabtree *et al.* [5] present a study of domestic routines for managing communications such as physical and electronic mail coming in and out of the home. Their study shares our interest in a fine-grained analysis of locations in the home, but focuses on locations of communication media and the role they play in issues such as the coordination of action among household members. For example, Crabtree *et al.* introduce a notion of ecological habitats to describe where communication media dwell; we have a complementary focus on the locations in which people spend time in the home.

To evaluate the use of technology in the home, we need to understand not only where technology is used in the home, but more broadly how people use space in the home. The existing literature from the social sciences is surprisingly sparse regarding daily use of space, particularly regarding specific patterns of use of places and objects; generally researchers have focused on broader issues such as identity, territoriality, gender, and power, and discussions are often at the level of rooms rather than specific places within the room (see e.g., [12,16]). In a notable exception, Oswald reports favored places for elders [22]. These favored places are motivated by a tendency towards environmental centralization, particularly when people suffer from a loss of mobility. Oswald however does not discuss in detail the characteristics of these places, and the population he is studying appears to use them somewhat differently than the more general population in our study.

Other relevant work explores the use of space and technology in the workplace environment. For example, Becker and Steele [1] discuss the concept of an *organizational ecology* that views space layout, technology infrastructure, and business processes as a unified whole. They use this concept to present arguments for non-territorial offices that support collaboration and dynamic allocation of space. Similarly, Brown and O'Hara [3] discuss how mobile workers dynamically make use of space. They explore how "place changes work" (e.g., availability of power or network connectivity dictating the tasks undertaken) and how "work changes place" (e.g., places like cafes or homes being turned into work places). While this work considers the use of portable technology in work-from-home settings, it does not explore technology and space use in the home in general. Luff and Heath [15] focus on the use of portable artifacts in workplace collaboration. In particular, they suggest that the *micro-mobility* of objects, such as the way in which multiple people orient to a paper medical record or circulate a paper work log, can be as important to the success of mobile technologies as the simple fact that they are *portable*. We return to this point later.

In this paper, we present what is to our knowledge the first study of the day-to-day use of wireless laptops in the home. Further, we present what is to our knowledge the first study to contextualize findings about technology use in a specific discussion of how participants' occupy space in the home.

3 Method

Data was collected as part of two related studies within a broader project on the use of space and technology (in particular wireless laptops) in the home. The lead author visited all homes in both studies and there was high overlap in interview content and procedure.

3.1 Participants

Participants consisted of the 28 occupants of ten households, as well as the six occupants of two pilot households. Two households were recruited from a local university community and the remaining eight were recruited via an online classified service. Participants were compensated. All households had wireless networks and at least one laptop computer. Households had a range of laptop and desktop computers (both PCs and Macs), with an average of slightly more than one computer per person.

Households were chosen to represent a diverse range of household composition and life stage. All households had multiple inhabitants; some households had children of various ages, while other households consisted entirely of adults, e.g., married couples who did not have children or "empty nesters" whose children had grown up and left home. Participants were from a range of ethnic and cultural backgrounds and had varying occupations, such as kindergarten teacher, health care analyst, tech support person, furniture salesperson, or student. Although technical knowledge varied, participants tended to be somewhat technically oriented, which is not surprising given that they were in households that were leading adopters of wireless home networks.

Eight of the households were located in the San Francisco metropolitan area, and two of the households were located in the Portland metropolitan area. The homes were in a range of neighborhoods, including for example an ethnic residential neighborhood and a quickly developing urban district. The homes in the study consisted of a range of housing types, from large "McMansions" to one-bedroom apartments. Our recruiting favored various types of open plan layouts because we were particularly interested in the use and coordination of more "flexible" and shared space, although some of the houses had traditional layouts.

3.2 Procedure and Analysis

Data was collected during 2005 and 2006. Each household was studied in some depth through multiple home visits, which included home tours, semi-structured interviews, and a variety of mapping activities (e.g., annotating floor plans, using felt maps to do "walk-throughs" of recent days [17], and interpreting visualizations of location data collected by sensors installed in some homes). While we covered a wide range of topics, our interactions with the participants particularly emphasized their use of space and technology. In all but one home, all household members were present at all interviews. The primary interviews (usually two interviews per family, for a total of approximately three to four hours per family) were typically video-taped and transcribed. We had a number of additional informal interactions with several of the families, and we typically took notes on these. Additionally, in some households, textual and photo diaries were kept by participants, while in other households ultra-wideband sensors were used to track the locations of the participants and laptops and application use was logged on laptop and desktop PCs. In this paper we focus on qualitative findings, although we have begun to explore quantitative findings as well [1]. Time-lapse photography was also collected in some of the public areas in some of the homes. Households typically participated for approximately one to three weeks, although one household participated for approximately four months, and we conducted brief follow-ups with some households after several months had elapsed.

We reviewed the transcripts, notes, videotapes, and visual artifacts (annotated floor plans, photos, visualizations, etc.), triangulating across the different sources. Through this process, we identified themes [24] that we judged as representative of the participating households. In the next section, we report the results of our analysis.

4 Findings

In this section, we present the findings that emerged most strongly from our analysis. Specifically, in Section 4.1, we present factors that shape habits regarding location of laptop use. In Section 4.2, we discuss the places people spend time in the home, introducing the notion of favored and kinetic places. We discuss how the places in which people spend time interact with the factors described in Section 4.1, resulting in particular patterns of laptop use in the home. Favored places were of particular interest because they were the primary site of laptop use. Accordingly, in Section 4.3, we describe favored places in more depth and present a taxonomy of them.

4.1 Factors Influencing Location of Laptop Use

In our analysis, we identified several factors that seemed to most powerfully affect location of laptop use in the homes we studied. The factors represent attracting or repelling forces that impacted laptop movement and use. Naturally there was individual variation, so while these were the major themes,[1] not all of these issues pertained to all participants in all homes.

The laptop functions as part of an assemblage. Laptops are drawn to established locations with appropriate infrastructure.

Laptops typically required infrastructure and space to enable general and sustained use. One manifestation of this was a general aversion to having the laptop unplugged, e.g., because of poor battery life or because participants wanted to make sure the battery would be charged "when needed." Needs varied somewhat by person and circumstance, but laptop assemblages frequently included some subset of the following items: a mouse; a horizontal surface on which to spread papers or books, use a mouse, and place the laptop itself; a power adapter; peripheral devices such as printers, monitors, keyboards, or speakers, and wires and other attachments to these devices; pens and pencils; and/or a phone. Jack describes actually entwining himself in such an assemblage. He says he sometimes sits on the couch with the laptop on his lap, a power cord running to his left, and wires to speakers running to his right; this means he can not get up from the couch without moving the laptop and disconnecting wires or trying to lift them over his head. In practice, a laptop is not simply a device, but is indeed a configuration of devices or infrastructure to support laptop use.

> Jack[2]: So we have a laptop and wireless but we're definitely wired too.

The laptop is more likely to be used when it is at hand. The laptop is not always conveniently situated or booted.

A number of situations arise in the home in which the laptop would be useful for quick tasks. For example, one may wish to "google" for the answer to a question that arises in conversation or "fill time" by briefly checking email. However, the laptop may not be at hand in those situations, and the costs of retrieving and/or booting it often dominate the benefits it would provide. Therefore, opportunities for lightweight or even longer-term computing tasks may go unfulfilled.

---

[1] Note that while we had considered that varying signal strength of the wireless network might play a role in where laptops were used on a day-to-day basis, this did not emerge as a significant theme, largely because laptops were routinely used in a relatively small number of places due to other factors. In the rare cases where signal strength was mentioned as an issue, participants reported that it was possible to modify their home networks (e.g., by adding a wireless repeater) to accommodate specific places if necessary.

[2] Participants' names have been changed to protect anonymity.

Laptops are in some senses surprisingly inconvenient to move from room to room or even from place to place within a room. A difference of just a few feet can make the difference in whether the laptop is used or not, and weight and wires are both inertial forces. Laptops often need to be detached from power or peripherals or closed before being moved – and closing them may automatically put them in standby mode or shut them down, which can lead to further delays. Further, laptops are moderately heavy and difficult to carry in one hand, and they are even more difficult to carry when one wants to bring along dangling infrastructure such as the mouse, mouse pad, or the power adapter. Additionally, if one is sitting (or lying) down and is tired, getting up to go get a laptop can be an unappealing notion indeed. Even when laptops are conveniently located, an additional obstacle is that they may not be booted and the boot cycles are not short. Participants told stories about walking to other rooms to use a computer that was already booted elsewhere in the home, rather than using the unbooted laptop right next to them.

The laptop is "fragile." People tend to avoid putting laptops in "dangerous" locations like the kitchen or the bathroom.

Many participants expressed concern that their laptop could be damaged by hazards such as cooking spatters, young children, or hard knocks. In practice, this meant they were concerned about putting their laptops in dangerous locations such as the kitchen, the bathroom, or other locations where they might be bumped, dropped, or spilled upon. Tom expressed some of these concerns, and discussed his perception that a consumer electronics device is more appropriate than a laptop for "unsafe" venues:

> Tom: The thing that bugs me sometimes is either [my teenage son] took it [the laptop] or [my teenage daughter] took it. They're in the kitchen with it, they got it open, it's sitting on there, and they're listening to their iTune collection… and I'm going, 'Don't you guys have iPods? Didn't I get you iPods with your own music on it?' You know. I don't really want this laptop sitting here because I said, cracked screen, knocking it, stuff like that… giving them $150 thing to walk around I… feel a lot more comfortable than letting them run around with a $1,000 thing.

In some cases, participants overcame their reservations and used their laptops in these more unsafe locations or situations. The important thing to note is that there is a cost-benefit tradeoff between the risk to the laptop and the value of using it in a given situation. Anne, the only participant who reported using the laptop in the bathroom, recognized that it was not always "safe" to use it in there, but she appeared to derive benefit from it and was careful to position the laptop "dead center in the middle of the counter." Different household members sometimes had different perceptions of the risk to a laptop, or at least different cost-benefit analyses. Anne's boyfriend Mark revealed that he borrowed her laptop for a friend to use while they were outside playing basketball.

> Mark: I was going to say, I didn't tell you that John came over and we were playing basketball. We had the garage open and I brought the laptop in the garage just because he wanted to check something. So I was like, "No problem."
>
> Molly: Of course you would leave that out, so she wouldn't kill you.
>
> Anne: We'll discuss matters later.

> Mark: So there's a little counter in the garage and I just put it on there for my friend to use the Internet.
>
> Molly: He shares that now, so she won't kill him. Last thing we need is the basketball to hit the laptop.
>
> Mark: We were just outside playing basketball and he was like, "I've got to check something for my school." And I was like, "Okay." Then I just brought the laptop outside so we didn't have to go back in. I wanted to keep playing and he wanted to check it really fast. So it does come in handy to have the wireless laptop… I left it there until we were done. The basketball wasn't anywhere near it so it wasn't going to break it.

This is exactly the kind of opportunistic use of the laptop that seems ideal, but such events appear to be the exception rather than the rule, even when one considers that sometimes they may not be discussed with interviewers because this might reveal them to other members of the household. The fact that sharing this particular event was so problematic reveals the importance of the issue to the participants, as well as the fact that such opportunistic use is not routine.

People sometimes want a "break" from the laptop. People sometimes put mental or physical distance between themselves and their laptop.

Participants had a range of strategies for dealing with work/home boundaries [20] and online/offline boundaries. Participants often said that it was "sad," "sick," or "pathetic" how much they worked or how much time they spent online. Some participants had few boundaries and would, for example, keep a BlackBerry on the night stand and check it in the middle of the night. Other participants would establish rules about times or spaces in which laptops or laptop use were permitted. Some areas were treated as technology-free sanctuaries, peaceful shelters in the home.

> Katherine [on her husband Sam's use of email late at night]: I just can't understand why anyone would be emailing or responding to email at 11:00... it's hard to imagine that someone's expectations are you're going to be online, reading an email that late.
>
> Anne: [F]or some reason I don't feel comfortable using [the laptop in my bedroom]. I need to be usually down here [in the living room] or somewhere else but not really in my room, because you know I kind of spend a lot of time there, my personal time, it's like my personal haven, so the last thing I want to do is work in there. So phone calls and things like that it's usually down here or outside on location… I usually like to keep my work separate from my personal life.

Another strategy for respite from online stimulation was to have the laptop in a clearly inactive state or put away. The act of closing the laptop or shutting it down was significant. Similarly, a laptop might remain in a bag when someone returned home, to allow them time to "unwind" before facing more email.

The laptop is drawn to activity. Laptops are sometimes positioned (or repositioned) so they can be interleaved with and integrated in ongoing activity in the home.

While many factors tend to limit the movement of laptops, they are in fact sometimes moved (or are strategically kept) near activity or other people in the home. Brad and Jacqueline talk about Brad using his laptop in bed at night so he can be near her:

> Brad: It's pretty easy work, so I do it in bed. At night. It's like, oh, I wanna do a bit more…
>
> Jacqueline: Yeah, as I've gone to sleep… I can't sleep by myself, so Brad has to go to bed and use his laptop in bed.

Participants reported interleaving computing with a wide range of other activities from relaxing to socializing, e.g., surfing the Internet while supervising a napping child. Participants also made it clear that computing can be fully integrated in some activities. For example, computing can play an important role in socializing – people can look at photos or choose music together, online news can prompt conversation, or people can chat online about a TV show as they watch it.

4.2 Where People and Laptops Spend Time in the Home

In the previous subsection we described factors influencing laptop use. Next, we discuss the locations in which people spent time in the home. We also discuss patterns of laptop use in these locations, and how the factors help explain the patterns we saw.

From the rhetoric associated with wireless and laptops, one might assume that the primary benefit of wireless laptops is that they let people compute anywhere in the home.[3] In practice, laptops appear to be in common use in a relatively small number of places in the home; specifically they are commonly used in the four types of favored places we describe below. Their use in other parts of the home is much rarer, although provocative and informative when it does occur. Laptops appear to be *portable* (physically able to move around), but not *mobile* (able to transition quickly and smoothly between a variety of contexts). As suggested by the factors influencing laptop use described above, the *micro-mobility* [15] of laptops is limited. In particular, laptops go far fewer places than paper-based information artifacts in the home [5,7], similar to findings about the mobility of paper versus electronic artifacts in the workplace [15].

4.2.1 Framework of Spatial Occupation of the Home

In the households we studied, each household member typically had two or three *favored places* where they spent the majority of their time during waking hours (this offers an interesting comparison to the more numerous and diverse places for communication media reported in [5,7]). Examples ranged from a particular spot on the sofa to a seat at the desk in a home office. Note that these places are quite specific – they are not simply general areas or parts of rooms. One might imagine that people living in larger houses would have a significantly larger number of favored places, but our findings suggest that the number of favored places per person remains relatively constant. While larger houses may afford more choice of location for favored places, apparently people still establish a relatively low number, due to factors such as habit

---

[3] One question that arises is the distinction between laptops with and without wireless capabilities. For example, how many of these behaviors are enabled simply by having a laptop, as opposed to a laptop on a wireless network? While it is difficult to say without making a direct comparison, we do observe that many of the participants' tasks required Internet access (e.g., web surfing and searching, email, IM, online bulletin boards), and Internet-related tasks were often interleaved with stand-alone tasks such as word-processing. Accordingly, we believe wireless is indeed a key enabler of laptop movement in the home.

and the work required to develop and maintain them. Participants spent extended periods of time in their favored places, conducting activities such as television watching, computing, socializing, or reading. Favored places naturally develop over time into habitats – places with resources conveniently arranged at hand. Objects such as drinks, books, remotes, and power adapters for laptops tend to accumulate in favored places. Placement of items can be ritualized and the environment can evolve over time to support convenient use – the spot for coffee, the cushion broken in, or the laptop adapter wedged in the sofa cushion.

In addition to favored places, participants naturally used other places in the home during waking hours. These places were generally associated with shorter duration, focused activities that involved physical manipulations, e.g., a mirror in the corner of the bedroom for doing one's hair, a door that a participant tucked her feet under to do sit-ups, or a kitchen counter that a married couple used to make sandwiches for lunch or to prepare the evening meal. Because these places were typically characterized by motion and physical manipulation, we refer to them as *kinetic places*.

4.2.2 Laptop Use in Kinetic Places

There were a number of resourceful uses of laptops in kinetic places in the home. We discussed above how Mark took the laptop outside for his friend to check email when they were playing basketball. Laptops were on occasion used for recipes in the kitchen or to check email in the bathroom. A particularly creative use was by Mareesa and Carlo – if their baby daughter had fallen asleep in the car on the way home, Carlo sometimes brought the laptop out to Mareesa so she could sit in the car in the driveway and use the Internet while supervising the child's nap. Mareesa sometimes also brought the laptop to the child's bedroom so she could be with her while she napped.

While we could emphasize the sensational and resourceful nature of these events, on the whole we felt they were much rarer than we would have expected, and that fewer participants took advantages of these opportunities than we would have expected. Further, location logging, application logging and detailed discussion revealed these events to be even more uncommon than initial discussions with participants might lead one to believe. We believe that these events are marked and tend to be disproportionately emphasized by participants because they are creative and memorable. For example, one family that talked enthusiastically about using the laptop for recipes in the kitchen turned out to use it for that purpose only once during the multiple-week study period, while another family that reported such use did not use it for that purpose at all during the study period.

These events are informative not only because participants are highly interested in them but also because stories about them tend to be accompanied by revealing descriptions of why using the laptop in a given location is troublesome. In fact, most of the factors described in Section 4.1 seem to discourage the use of laptop in kinetic places in the home. Although laptops may require less infrastructure for short-term tasks than for long-term tasks, lack of items in the assemblage was still often an issue. Participants were often reluctant to use laptops without power or a mouse. Further, lack of an adequate place to put a laptop was a barrier – laptops are plainly difficult to use while standing up and are even more difficult to operate while walking or doing other physical activities. Kinetic places in the home are many and varied, so laptops were often not at hand, and it was rarely worth the trouble to detach from power and peripherals and move the laptop. Concerns about breakage also made it less likely that the laptop would be brought to more active areas in the home. Desire for a "break" from the laptop also sometimes limited its movement in the home.

4.2.3 Laptop Use in Favored Places

In general, participants had a small number of favored places where they spent extended periods of time during waking hours in the home, and laptops were used in most of these places. In fact, we can see that most of the factors described in Section 4.1 presage the use of laptops in these favored places. We discuss the relevance of these factors in turn. First, laptops are part of an *assemblage* and they require a place for that assemblage. Favored places include habitats that have horizontal surfaces for books, projects, or food and often other desirable infrastructure such as power outlets to facilitate the use of other objects such as lamps. It is natural to extend these locations to incorporate items such as speakers, mice, or power adapters, and in fact the assemblage is often visible even when the laptop is not currently there.

Second, laptops are more likely to be used when they are *at hand*. The laptop is more likely to be at hand in favored places, simply because people spend a large amount of time in a small number of favored places and the laptop tends to be in one of those places (which is of course a self-reinforcing phenomenon). Further, visits to favored places tend to be longer than visits to kinetic places, and costs associated with settling in with a laptop, such as connecting the power supply and organizing supporting materials, are more acceptable in longer-stay than in shorter-stay spots. Additionally, laptops may be kept at hand but not used continuously, particularly in the comfortable places described below. For example, a laptop may be set aside and then picked up during commercials or when conversation prompts use. This type of intermittent use matches well with the use of other objects in favored places – many objects in favored places are kept at hand but not used continuously. Laptops fit well with the scale and treatment of these other objects: laptops can be stacked and tidied with magazines or papers, remotes can be put on top of them, or laptops can be removed entirely if the room is cleared of clutter in anticipation of visitors.

Third, laptops are perceived as "*fragile*." Favored places tend to be in "safe" locations in the home – locations where objects like books, furniture, and the human body itself are less likely to be damaged. Laptops fit comfortably into these environments. In fact, the laptop has some relationship to the human body in terms of nurturance needs. The laptop needs some of the same things as the human body on a somewhat similar time scale: it needs fuel on a multi-hour time cycle; it does not perform well in direct sunlight (the laptop screen is not very usable in bright light, and people have an aversion to sitting in direct sunlight, for example moving out of their favored places or adjusting the blinds if direct sunlight hits); and the body and the laptop both seem attracted to soft seating – given concerns about bumping the laptop, it is in some senses easier to place it rather carelessly on a soft surface than to place it carefully on a hard surface so that it does not bump, and we saw many laptops "sitting" on a couch or an ottoman rather than a coffee table. Therefore, it is somewhat natural that the laptop would be compatible with existing places for sitting.

Fourth, recall that people sometimes wanted a "*break*" from their laptops. In some senses, putting laptops in favored places seems to work against this, since for example they enable people to frequently check email from the couch. Laptops do however have some nice affordances for boundaries, e.g., it is highly significant that laptops can be closed, turned off, or put in their bag by the front door.

Fifth, recall that laptops are *drawn to activity*. As we discussed above, some of the situations in which laptops can be useful do not occur in favored places – but many do. Additionally, the habitat in favored places often includes a television, books, and other resources compatible with laptop activity.

4.2.4 Change on a Multi-week and Multi-month Time Scale

Although the number of places in which participants routinely used laptops was fairly limited, we would like to emphasize that the participants' ability to choose these places was significant. This choice was exercised with some frequency. There was naturally some variation from day to day. But even more importantly, we learned that the use of space and laptops within space was highly contingent on routines that change on a multiple-week or even multi-month time scale. Use of favored places and laptops was highly sensitive to current routines and projects. Seemingly small changes such as a difference in a child's nap schedule, a change in the weather, or a new project assignment at work could greatly perturb the system, meaning for example that laptop use ceased entirely in one location and began in another.

4.3 Types of Favored Places

In the previous subsections, we presented factors that influenced location of laptop use, and we discussed how these factors interacted with where people spent time and used laptops in the home. In this subsection, we focus in more detail on the primary sites of laptop use in our participants' homes: favored places. (While space constraints prevent a more detailed discussion of kinetic places, we note that the kinetic places were significantly more variable and idiosyncratic than the favored places.) Specifically, we discuss the characteristics and affordances of favored places. While stereotypical notions such as "dad's favorite chair" do appear to have some basis in reality, we learned that there are not simply favored places, but importantly, there are different types of favored places. We did a clustering exercise of the favored places of our participants that revealed four major clusters. These four clusters can be usefully organized into a two-by-two matrix (see Figure 1), whose axes emerged as themes in our analysis. The first axis is comfortable versus ergonomic, and the second axis is open versus closed (in United States homes, this axis is strongly correlated with public and private spaces in the home). We discuss each axis in turn.

Comfortable places are (not surprisingly) most strongly characterized by comfortable seating – usually a sofa, but sometimes a soft chair or even a bed. Comfortable places generally support a wide range of seating positions and minor variations in location. Laptops are useful in these places, since (unlike desktops) laptops can easily be used in multiple positions. Posture can be chosen based on fatigue and task, and variation in physical position can be refreshing. Comfortable places typically have a low or small table nearby which contains resources such as a drink, a book, or a laptop computer. Comfortable places are often associated with unstructured time and support a wide range of (often interleaved) activities such as television watching, talking on the phone, socializing, reading, or computing. Computing in these locations is likely to involve more "relaxed" tasks such as web surfing, IM, or email (note that a recent Pew Internet Report discusses a dramatic increase in the number of people using the internet "for fun," going online "for no particular reason" [8]).[4] Most comfortable places were associated with use of a laptop, although a small number were associated with only a television.

Fig. 1. The four types of favored places. These are the locations in which laptops were most commonly used in the home, arranged as to whether they are *Comfortable* or *Ergonomic* and are in the *Open* or *Closed* areas of the home. *Comfortable-Open*: Sam likes to watch TV and work on his laptop. His wife Katherine works on her laptop at the dining table nearby. *Comfortable-Closed*: Gaby finds a comfortable position to use the laptop on the bed. *Ergonomic-Open*: Jack and Margaret's dining table doubles as a desk. While Margaret works at the desk, Jack works at the couch nearby. Sometimes they switch places. *Ergonomic-Closed*: Kumar likes to work in the home office for privacy while his young children play elsewhere.

By contrast with comfortable places, ergonomic places are characterized by upright chairs positioned in front of tables or desks. The positions of both people and laptops are more static in ergonomic places than in comfortable places. Ergonomic places are often associated with focused computing tasks and/or tasks that involve using a mouse or spreading out papers or books on a horizontal surface. Ergonomic places seemed to be associated with a higher level of mental activation or alertness for some participants. All the ergonomic places we observed were associated with the use of either a laptop or a desktop (and all the desktop computers we saw were in ergonomic places).

---

[4] One might expect clear patterns would emerge overall regarding which tasks occur in which locations, for example, that task would dictate location. However, numerous confounding factors such as fatigue influenced the location at which a given task would occur, and participants often appeared to move fluidly among tasks while in a given location. As a result, the relationship between task and location appears to be less straightforward than one might imagine. We believe this complex interaction is a rich area for further investigation.

Table 1. The four types of favored places used during waking hours

Type	***Comfortable-Open***	***Comfortable-Closed***	***Ergonomic-Open***	***Ergonomic-Closed***
Canonical example	Family room couch	Bed with view of TV	Informal dining table	Desk in home office
Occurrence	Very common	Less common	Common	Common
Seating	Couch / soft chair	Bed / soft chair	Upright chair	Upright chair
Seating positions	Multiple	Multiple	Single	Single
Horizontal surfaces	Coffee / side table	Nightstand	Table at desk height	Desk
Task	Focused or not focused	Focused or not focused	Somewhat focused, primarily computing	Focused, primarily computing
Location in home	Central	Not central	Central	Not central
Openness	Open, spacious	Confined	Open, spacious	Confined
Visual properties	Good view and light	Lesser view or light	Good view and light	Lesser view or light
Media	View and control of entertainment center	Smaller TV	Often has view of entertainment center	Usually no TV
Facing	Outward facing	Outward facing	Outward facing	Facing wall
Clutter	Useful to tidy	Useful to tidy	Useful to tidy	May stay cluttered
Computing Device	Laptop	Laptop	Laptop	Laptop / destop

Open places are centrally located in the public, central spaces in the home, such as the family room or the dining room in an open plan home, or even the occasional spot outside. They often take advantage of the best views and light the home has to offer. Open places offer proximity to social activity in the home, and seating in these areas generally faces the room and/or other people. These places typically have a good view of the entertainment center (including items such as the TV, gaming station, and stereo). Laptops are useful in these areas because, unlike desktops, laptops lend themselves naturally to positions facing outward – toward the room, companions, the view, or the TV. Further, these areas are usually in the “front-stage” of the home, so it is often desirable to tidy them. Laptops are particularly desirable in these areas relative to desktops, because laptops can be easily tidied or removed entirely. For example, laptops can be taken out of the living room and hidden away in another room during a party, or they can be cleared off the table for dinner.

By contrast with open places, closed places are in less central areas of the home – typically home offices or bedrooms. These areas are more confined and offer more privacy than the open areas. Here privacy often means audio isolation – one might go to one’s room to watch TV or talk on the phone without disturbing others, or conversely one might go to one’s room to read or compute without being disturbed by TV watching or social activity taking place in the open areas of the home. Additionally, because they were more confined, ergonomic-closed places offered the advantage that projects in them could be left spread out, while configurations of objects often needed to be tidied up in ergonomic-open places.

Table 1 summarizes the characteristics of the four types of favored places. While the patterns are strong, naturally there is some variation so the table captures canonical types rather than exact rules. Figure 1 shows examples of each type of place.

Unlike Oswald, who observed that elders had a single favored place [22], we found that individuals have an ecology of favored places suited to different needs – each of our participants typically had a set of favored places of different types. For example, one common pattern was to have a comfortable-open place paired with an ergonomic-open place, while another was to have a comfortable-open place paired with an ergonomic-closed place. Some participants also mentioned moving to different places because they found it "refreshing" or because it helped keep them awake.

> Anne: Yeah, I'll use my laptop in here [at the kitchen table] sometimes if I want to be like sitting up or just kind of have a change of scenery. I'll sit in here, kind of face out if it's a nice day.

Favored places had clear owners, although use was of course coordinated and sometimes even shared. More common than sharing was complementary use of favored places, for example a husband's place on the couch and a wife's place at a table nearby, both used together in the evening.

5 Discussion

Participants had a set of favored places where they spent time at home – places that offered different physical, social, and sensory experiences, and the laptop was accommodated in all of these. Laptops moved with some regularity among favored places, and also occasionally to other locations in the home, or to external locations such as work. Because of this flexibility, the laptop (especially when wireless) brings computing into the home in a way the desktop does not. Laptops are positioned and re-positioned in key locations, from the most public to the most private, interleaved and integrated with ongoing activity – in a given day, a laptop may be used in the living room at the hub of social activity and then in bed before falling asleep.

In some senses one could argue that the wireless laptop is a triumph – successfully used in a variety of situations in the home. However, we believe this would be an overly simple characterization. There are significant frustrations and lost opportunities, both at favored places and at kinetic places in the home. Accordingly, we now turn to a discussion of design opportunities in these places, highlighting how the consideration of different classes of place in the home could inform the design of new objects more tightly integrated with different aspects of the domestic ecology.

While laptops were satisfactory in all four types of favored places, they were notably optimal for none of them. In its current form, the laptop is a compromise object. For example, it is not an ideal object for comfortable places, and frustrations often arose regarding body and laptop position. Keyboards can be awkward to use when lying on one's side, the screen can be difficult to see from an oblique angle, and laptops emit heat and can be too hot to place in one's lap.

Accordingly, we believe that it is an important challenge to develop technology and infrastructure that is customizable to each of the different types of favored places. We believe that our findings about the nature of favored places and the way in which they are used can serve as useful constraints for evaluating different models that might be proposed for such customization (and that they have more general application in the ongoing debate between device convergence and appliance design). Some of the key observations are: (1) people typically had a set of two or three favored places; (2) favored places offer different experiences and have different characteristics; (3) comfortable places involve varied posture; (4) a given place may be used for a wide range of tasks from lightweight surfing to "serious" work; and (5) the location or use of favored places may vary due to changes in routine.

As an example of how the constraints can be used to reason about different classes of design solutions, our finding about changes in routine suggests that it is unrealistic to build-in technology or carefully instrument the home at a set of pre-designated places. Portable technologies seem more appropriate than built-in technology for favored places. Further, we would suggest that the most compelling designs involve objects that are not simply *portable* but are *reconfigurable* as well. For example, devices with keyboards and screens that expand from a smaller object would fit well with people's existing patterns. Note that design of reconfigurable objects in the surrounding habitat is also important. Figure 2 shows two examples of participants who creatively and dynamically reconfigure their comfortable places to add ergonomic elements. Many effective design solutions may lie not only in the design of the technology itself, but also in the design of furniture to better support the technology.

Although the laptop is becoming part of the fabric of daily life in favored places, barriers frequently prevent the laptop from being used in compelling circumstances in kinetic places. Kinetic places have very different characteristics than favored places – they lend themselves well to robust devices that are free of attachments and possibly include hands-free or single-handed interfaces. For example, built-in displays, voice UIs, or smaller appliance devices may be appropriate in these locations.

As a final observation, our findings in some senses argue against the notion of fully ubiquitous access in the home. People sometimes wanted distance between themselves and technology, for example creating technology-free zones or lamenting the intrusion of technology into certain spaces. It is an interesting design challenge to try to resolve the apparent paradox between the notion of having computing "everywhere" and maintaining boundaries.

Fig. 2. Left: Part of Tony's coffee table "pops up" to become a tray for the laptop. Right: Gaby works at nesting ottomans. She puts the laptop and mouse on top of one ottoman, pushes out a second with her feet, and props her legs on a third.

6 Conclusions and Future Work

We have presented the results of a study of people's use of space and wireless laptops in the home. We have examined the relationship between where people spend time and their use of computing devices. Participants each had a small set of favored places in the home. Wireless laptops were routinely used in almost all of those places, and we have identified factors that promote their use in these locations. Wireless laptops were used much less routinely in other areas of the home, and we have identified barriers to their use in these places. We have discussed the relevance of these findings to new technologies and form factors.

There are many excellent opportunities for future work, including issues such as the complex interaction between task and location, moment-by-moment analysis of what occasions relocation from one place to another, the impact of fatigue on the use of space and devices, collaborative versus individual use of devices, the use of ensembles of devices [26] with different form factors (e.g., cell phones, BlackBerries, iPods, laptops) in different locations in the home, and extending the findings to encompass multi-cultural issues.

Acknowledgements

We thank the participants, without whom this work would not have been possible. We are grateful for discussions with Paul Aoki, Kurt Brown, John Canny, and Ben Hooker. We thank Paul Aoki and Sara Bly for comments on drafts of this paper.

References

1. Aipperspach, R.J., Rattenbury, T.L., Woodruff, A., Anderson, K., Canny, J.F., Aoki, P.: Ethno-Mining: Integrating Words and Numbers from the Ground Up. Technical Report No. EECS-2006-124, , EECS Department, University of California, Berkeley (2005)
2. Becker, F., Steele, F.: Workplace by Design: Mapping the High-Performance Workscape. Jossey-Bass, San Francisco (1995)
3. Brown, B., O'Hara, K.: Place as a Practical Concern of Mobile Workers. Environment and Planning A, 35(9) (2003) 1565-1587
4. Brush, A.J., Palen, L., Swan, L., Taylor, A.S.: Designs for Home Life. SIG at CHI 2005
5. Crabtree, A., Rodden, T.: Domestic Routines and Design for the Home. JCSCW 13(2) (2004) 191-220
6. Edwards, W.K., Grinter, R.E.: At Home with Ubiquitous Computing: Seven Challenges. Proc. UBICOMP 2001 (2001) 256-272
7. Elliot, K., Neustaedter, C., Greenberg, S.: Time, Ownership and Awareness: The Value of Contextual Locations in the Home. Proc. UBICOMP 2005 (2005) 251-268.
8. Fallows, D.: Growing Numbers Surf the Web Just for Fun. Pew Internet Reports (2006)
9. Frohlich, D.M., Dray, S., Silverman, A.: Breaking Up is Hard to Do: Family Perspectives on the Future of the Home PC. Int. J. Human-Computer Studies 54(5) (2001) 701-724
10. Frohlich, D., Kraut, R.: The Social Context of Home Computing. In: Harper, R. (ed.): Inside the Smart Home. Springer, London (2003) 127-162

11. Grinter, R.E., Edwards, W.K., Newman, M., Ducheneaut, N.: The Work to Make the Home Network Work. Proc. ECSCW 2005 (2005)
12. Hanson, J.: Decoding Homes and Houses. Cambridge University Press, Cambridge (1998)
13. Harper, R. (ed.): Inside the Smart Home. Springer, London (2003)
14. Kraut, R., Scherlis, W., Mukhopadhyay, T., Manning, J., Kiesler, S.: The HomeNet Field Trial of Residential Internet Services. CACM, 39(12) (1996) 55-65
15. Luff, P., Heath, C.: Mobility in Collaboration. Proc. CSCW 1998 (1998) 305-314
16. Mallett, S.: Understanding Home: A Critical Review of the Literature. The Sociological Review 52(1) (2004) 62-89
17. Mateas, M., *et al.*: Engineering Ethnography in the Home. Proc. CHI ’96 (1996) 283-284
18. McClard, A., Somers, P.: Unleashed: Web Tablet Integration Into the Home. Proc. CHI 2000 (2000) 1-8
19. Mercer, D.: Home Network Adoption: WiFi Emerges as Mass Market Phenomenon. Strategy Analytics (2006)
20. Nippert-Eng, C.E.: Home and Work: Negotiating Boundaries Through Everyday Life. University of Chicago Press, Chicago (1996)
21. O’Brien, J., Rodden, T., Rouncefield, M., Hughes, J.: At Home with the Technology: An Ethnographic Study of a Set-Top-Box Trial. ACM TOCHI 6(3) (1999) 282-308
22. Oswald, F., Wahl, H.-W., Naumann, D., Mollenkopf, H., Hieber, A.: The Role of the Home Environment in Middle and Late Adulthood. In H.-W. Wahl *et al.* (eds.): The Many Faces of Health, Competence and Well-being in Old Age: Integrating Epidemiological, Psychological and Social Perspectives. Springer, Heidelberg (2006)
23. Patel, S.N., Kientz, J.A., Hayes, G.R., Bhat, S., Abowd, G.D.: Farther Than You May Think: An Empirical Investigation of the Proximity of Users to Their Mobile Phones. Proc. UBICOMP 2001 (2001) 123-140
24. Ryan, G., Bernard, H.R.: Techniques to Identify Themes. Field Methods 15(1) (2003) 85-109
25. Schadler, T.: Laptops And Home Networks Transform Behavior. Forrester (2005)
26. Schilit, B.N., Sengupta, U.: Device Ensembles. IEEE Computer 37(12) (2004) 56-64
27. Taylor, A.S., Swan, L.: Artful Systems in the Home. Proc. CHI 2005 (2005) 641-650
28. Venkatesh, A.: Computers and New Information Technologies for the Home. CACM 39(12) (1996) 47-54
29. Vitalari, N.P., Venkatesh, A., Gronhaug, K.: Computing in the Home: Shifts in the Time Allocation Patterns of Households. CACM 28(5) (1985) 512-522
30. Weiser, M.: The Computer for the 21st Century. Scientific American 265(3) (1991) 94-104

Managing Communication Availability and Interruptions: A Study of Mobile Communication in an Oncology Department

Jeremiah Scholl, Per Hasvold, Eva Henriksen, and Gunnar Ellingsen

Norwegian Centre for Telemedicine
{jeremiah.scholl, per.hasvold, eva.henriksen}@telemed.no
University of Tromsø
gunnar.ellingsen@unn.no

Abstract. Wireless phones and text messaging are tremendously popular in many areas of society. However, they are still relatively unused in hospitals where pagers are a pervasive communication device that is notoriously difficult to replace. This paper studies pager and wireless phone use at the oncology department at University Hospital of North Norway. Participatory observation and interviews with physicians are used to provide qualitative analysis about the use, drawbacks and benefits of both technologies. A number of important issues are addressed that should aid designers of next generation mobile communication systems for hospitals. In particular, the data points towards specific features that will be crucial for the overall usability and acceptance of an integrated device that supports paging, voice and text services. Of particular importance will be features that allow users to manage their communication availability and avoid interruptions.

1 Introduction

Hospitals are complex organizations that are increasingly pressured to provide their services as efficiently as possible. Coordination and collaboration are key ingredients in the work of nurses and physicians in hospital environments. Healthcare is also a knowledge intensive activity where consulting colleagues is often a necessity [6]. This amounts to numerous interruptions for persons in key roles or who are knowledgeable. The balance between getting immediate access to resources (such as senior physicians) and overloading or causing interruptions in moments where this does not fit in with the activities of the resource has similarities with classical problems regarding collaboration and sharing of resources, such as of disparity in work and benefit, "prisoner's dilemma" and "the tragedy of the commons" [13].

Knowledge, experience and the power to make decisions (with possible consequences to a patient following a decision on treatment) is a limited resource in many enterprises. Access to such human resources is necessary but at the same time there is a real danger of overloading such resources. If these resources are themselves part of processes involving activities that require concentration or unavailability, we can see the outline of potential conflicts of interest. Like the commons, the knowledgeable

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 234 – 250, 2007.

and experienced are resources that may be shared. If overused however, they will not be able to keep updated and remain a resource to their colleagues.

In hospitals this has contributed to a workplace that suffers from a poor set of communication infrastructure and practices [8]. One suggested strategy for improving hospital communication is the adoption of mobile technology [8]. Some initial studies have shown a variety of potential benefits for mobile technology such as wireless phones and text messaging when deployed in a hospital setting [10], [17], [25]. However, despite seeing pervasive use in many other areas of society these technologies see limited use in many hospital environments.

A mobile communication device that *is* pervasively deployed in hospitals is pagers. Pagers offer a cheap and reliable way of contacting staff but suffer from a variety of limitations due to their simplicity. The most obvious limitation is that they require staff to locate a telephone in order to respond to a page. They also create a large amount of unnecessary interruptions [5], [14] and communication overhead since for example, the person that has placed a page is not always near the telephone when the page is returned [25].

This paper focuses on socio-technical aspects of mobile communication in hospitals. A workplace study is presented that includes participatory observation sessions and interviews with physicians at the Oncology department at University Hospital of North Norway (UNN). The study was conducted to serve as input to the design process by revealing the current situation regarding communication on mobile devices in the department. The next phase of the project will focus on the design and implementation of tailored communication devices for the department based on the study and other information.

Qualitative data regarding the current state of communication using pagers and phones is presented in the paper, as well as some relevant attitudes by staff members towards the introduction of new communication technology such as text messaging. These are critical issues because even when technically sound, over half of medical informatics systems fail because of user and staff resistance [2]. While giving each staff member an integrated device that supports paging, text messaging and phones seems to be a straight forward solution, the data presented in this paper suggests that systems based on such a device will need to have specific features that help users control their communication availability in order to avoid overloading the individual physician with too many interruptions.

The rest of this paper is organized as follows. In the next section we present background studies. This is followed the participant observation and interviews at the Oncology department in section 3. Section 4 gives a discussion of implications on design and also presents future work.

2 Background Studies

Coiera and Tombs [8] identified a number of problems with communication in hospitals. A key observation was a tendency of hospital staff to prefer interruptive communication methods. Other studies have also reported unnecessary interruptions to be a specific problem related to hospital paging systems [5], [14]. On a cognitive level this “selfish” interruptive behavior has been attributed to a highly pressured work environment that leaves workers operating with working memory at full capacity [20]. This results in a prioritization of tasks that will reduce mental burden. Interruptive communication methods are then preferred since common asynchronous methods of communication such as voicemail and e-mail do not routinely offer explicit acknowledgement that a message has been received. A variety of approaches were recommended by Coiera and Tombs [8] for improving communication. These included support, asynchronous communication with acknowledgement, and mobility.

A variety of strategies have been explored for supporting mobile hospital workers. *Mobile information systems* are expected to be widely deployed in the future and may impact communication practices by bypassing the need for human-human communication in certain situations. For example, sending informative alerts directly to pagers [21], [23] may cut out the need for human-human interaction regarding the availability of lab data. However, since information systems cannot entirely replace human-to-human communication [7] pervasive *mobile communication* tools will also be important for supporting mobile hospital work.

2.1 Mobile Communication in Hospitals

The most intuitive approach for mobile communication is to provide users with wireless phones. Several barriers however make the pervasive use of standard cellular phones impractical in hospitals. One technical barrier is interference with medical equipment. Recent studies however have recommended that the potential benefits of the technology in some situations may outweigh the risk created by interference with medical equipment [15], [19]. Other potential problems include increased noise levels, and interruptions during consultations if the public are allowed to access cellular networks in hospitals [19].

Wireless phones designed specifically for in-hospital use should be able to avoid these problems. However, pervasive socio-technical barriers may still be an issue limiting their adoption. Since wireless phones make people more available, they may result in interruptions because of conversations that would not occur otherwise [20]. This issue cannot be ignored in light of the “selfish” communication behavior attributed to hospital workers.

A limited number of studies however have reported positive results when providing nurses with wireless voice services using various technologies [17], [24], [25]. Spurck et al. [25] reported general satisfaction by nurses and doctors as well as efficiency gains and fewer interruptions to patient care when providing a surgical nursing team with wireless phones. Minnick et al. [17] provided nurses with wearable radio transceivers connected to a base station. Hospital staff reported a variety of benefits from the system including quicker updates to patient information, easier location of nursing staff, and a perceived reduction of noise levels. However, concerns were also expressed for the potential to make nurses “fatally available”, with the authors suggesting for “…continued work in designing practice patterns that minimize nurse interruptions.” A more recent study has also reported generally positive results when providing nurses with wireless headsets as a hands-free interface to the phone system [24].

Asynchronous mobile communication has been explored in hospitals using technologies such as mobile text messaging (2-way paging, email etc.) and virtual whiteboards. Mobile text messaging for hospitals has been investigated with generally positive results, [1], [10] and was shown to be preferred over other asynchronous media such as email [10]. Some limits have been reported however including problems related to character limits from small displays, and reservations from physicians about being forced to carry an additional device. This can lead to doctors routinely forgetting or refusing to carry the device with them [1].

Mendonça et al. [16] describe a strategy for mobile information and communication based on PDAs. Their system combines access to patient data with a virtual whiteboard. The whiteboard allows team members assigned to specific patients to identify each other, and also post and track the progress of routine tasks.

Some additional studies have also examined applying context-awareness to mobile communication in hospitals [18], [3]. These include an architecture for routing messages and data to people based on their location, role, time, and artifact location and state [18], and also a system that supports messaging, voice services, and contextual information about users such as location and user configured status (busy, available etc.) [3].

3 The Study

Medical informatics research usually focuses on measuring the end results of systems on factors such as mortality rates and cost of care. This is critical for guaranteeing that systems help fulfill the principle goal of medical research, which is to improve patient care. This methodology however does not generally provide information on *why* particular systems fail to live up to expectations or *how* they can be improved. Pratt et al. [22] suggest an improved approach for medical informatics would be to adopt techniques such as observation, participatory observation, semi-structured interviews, and analysis using grounded theory from CSCW. These techniques are used in order to gain a deep socio-technical understanding of how systems affect users and how they can be improved.

The fact that wireless phones and text messaging have reached near pervasive use in many areas of society but still have difficulty replacing hospital pagers suggests that this sort of analysis would prove useful for providing better understanding of mobile communication systems in hospitals. The goal of this study was to identify work practices and attitudes that may affect the adoption and overall usability of these systems.

3.1 Methodology

The study was carried out in three stages, about 1.5 years apart, at the Oncology department at UNN. UNN is a university hospital, making it the hospital with the highest level of specialists in its region of the country. The hospital serves a population of approximately 460000 people in wide geographic area that contains around 420 General Practitioners (GP), and 10 other somatic hospitals. In this respect the physicians at the department are responsible for medical advice to colleagues at other hospitals in the region as well as GPs in the primary health care system. The patients of the department are long term patients that receive some of their treatment at the regional hospital, and also at their local hospital or at home through the local GP. The specialists at this department therefore serve, not only the patients that are at the inpatient ward, but also patients who are receiving treatment at other locations. In addition, the specialists provide medical advice and guidelines for cancer treatment to the medical staff in the other hospitals as well as being part of the oncology community nationally and internationally.

The methodology for the study was selected in order to get an understanding of actual activities and use of communication devices without having the interviews and observations biased by technologies we had introduced at the department. The first stage consisted of participatory observations including open interviews and informal discussions. The stage was intended to provide an overall picture of the work practice and communication situation in the department for a separate project investigating the introduction of Instant Messaging for inter-hospital communication. Thus, it uncovered a number of issues related to mobile communication that could be used in the study. The second stage was semi-structured interviews of a selected group of physicians at various levels of the hierarchy within the department. The third stage consisted of a second round of participatory observation sessions. This stage was conducted in order to gain further insight on issues uncovered during the first two stages and also to verify that the situation at the department had not changed significantly since the first stage of the study was conducted. This was necessary because the study was conducted over a 1.5 year period due to difficulty coordinating with the department.

The participatory observation sessions also included open, unstructured, and mostly ad-hoc interviews related to a specific situation or observation of activities involving decisions and judgment not immediately understood by the observers. These were not recorded due to the constant change of location and who would be co-present. This made it impractical to inform passers by that we were recording, and would have created a risk that patient-identifiable information inadvertently got into the tapes. Instead notes relevant to the study were taken.

The observers took the role of apprentices [4] and participated in actual work activities as performed by the physicians and nurses they followed. This was agreed upon before the observations and issues of privacy and how to conduct in certain situations were discussed with the head of the department as well as in the research group. Experience from similar observations in previous projects was utilized in order to keep this within ethically accepted behavior. The observers for example, had previously signed contracts regarding the privacy of patient information.

Participatory Observation. Four observers followed the work in the oncology department during half day sessions on four separate days for a total of 43 man hours of observations. Whenever the situation would allow it, the observers asked questions to the physicians and nurses related to the situation or context they were in, and to clarify aspects of the use of communication devices they were observed to be using. The observations were set up by the head of the department. The day-shift (from 08:00 to 15:30) was selected for the observation sessions since other shifts are staffed by a limited number of people in order to handle emergencies. As such, the day shift involves more far intensive communication in terms of complexity and volume.

Interviews with Medical Staff. At the beginning of the first day of stage one an open interview was conducted with the head of the department in order to gain insight into his use of current communication technology. The interview took place in his office and lasted for 30 minutes. Notes were taken during the interview but it was not recorded. Additional interviews were recorded at a later date with four other members of the department. Two of them were residents and two were attending physicians including one that was serving as the head of the department while the department head that was interviewed previously (during the participatory observation sessions) was on leave. These interviews were conducted with one person at a time in a break room at the hospital and lasted for 30 minutes each. The initial focus of the interviews was the current use of phones and pagers and this was this followed by additional questions regarding attitudes towards the introduction of text messaging. The interviews were conducted in a mixture of Norwegian and English. Quotes have been translated from Norwegian to English when necessary. Each quote is labeled to identify the individual source of the quote and their work role.

3.2 The Work Practice at the Oncology Department

This section provides a description of the routine and mobile communication devices used at the oncology department in order to provide the reader with a basis for context of usage for many of the issues discussed in the next subsection.

Description of routine. On the first day the observers met early with the department head and were briefed about routines and some of the main activities at the department. They then conducted the interview with Department Head – A, and afterwards joined the rest of the physicians at a briefing at the Radiology Department. This is where the physicians (and the head nurse) usually begin their working day. During this meeting the physicians get a report on the CT-scans (and other radiology methods) for all patients examined the previous day.After this meeting the physicians and some of the head nurses have a short meeting. This can consist of lectures and/or discussions related to administrative work and assignments.

The department has three sections: The Outpatient Ward, The Inpatient Ward, and the Physics or Radiation Section (both terminologies are used by the staff). The physician(s) with responsibilities at the outpatient post and the radiation section leave to get started. The remaining physicians have a briefing of what has taken place in the evening and night; new submissions, emergency cases, changes to patient's conditions, and other issues related to the status at the inpatient ward are presented and discussed if necessary. The physicians for inpatient ward are then divided into two teams whom each have responsibility for a section of the department. The observers split up and followed the rounds of these two teams at the inpatient ward.

Each team starts with a pre-visiting rounds meeting at the inpatient ward. At the meeting each patient belonging to the team is discussed in detail. The meeting is led by a nurse who is a team leader and attended by a nurse assigned to that patient, a resident, and a physician. Each patient is discussed individually and a plan is laid down for each patient. This plan contains medication, variables that needs to be observed, messages and procedures the patient needs to go through, etc. This is also an opportunity for the nurses to interact with the physicians and communicate observations and evaluations related to each patient.

Visiting rounds follow this meeting. Each patient in the inpatient ward is visited by a physician and a nurse. The team visited approximately 10 patients. The rounds took one to four hours. After this the physician would get back to the office and go through email, mail and other messages. After lunch the physicians disperse to various tasks sometimes in their offices or in any of the specialties rooms.

On the second day the observers went to the physics section of the Oncology Department and saw how the physicists conducted the radiation treatment of the cancer patents after they had been through the simulator. The third day was similar to day one, while day four started as the other days, but instead of visiting rounds, observations were made of the physicians working at the radiation section. Two physicians from the department are assigned to the radiation section daily and are responsible for helping to plan out and simulate radiation treatment for patients from all departments in the hospital. The physics section acts as a self-contained unit and, besides the logistics of schedules for patients receiving treatment, there is little communication directly between the inpatient ward and the physics section.

Mobile Communication Devices. A variety of mobile devices made by the vendors Ascom and Ericsson are used by the staff. Pagers are the dominant mobile device but a few senior physicians also carry a wireless phone with them. A brief description of these devices and their intended use is given below.

Pagers. Pagers are the dominant mobile communication device used at the department with the vast majority of physicians not being reachable by wireless phone on a given day. Each doctor is assigned one individual pager, and occasionally one or more role-based pagers. The use of role-based pagers is interesting because it demonstrates a simple solution at the system level for supporting communication similar to role-based context-awareness as described by Munoz et al. [18]. Some roles that are assigned a pager, such as head of the department, are fairly static, while others are dynamically assigned on a shift-to-shift basis. One physician is always in charge of any emergency calls for example and at least one is in charge of the out-patient ward.

Pagers have unique numbers in the internal phone and pager system, listed in the hospital directory. Pagers are activated by placing a call to the number and allowing it to ring at least three times. The pager's display will then show the number of the caller. The pagers are charged in a special rack located in the nurse headquarters. Usually the same pagers are used for each role daily and the list of who is in which role is printed and hung next to the rack. This allows the rack to serve as a presence awareness mechanism by informing who is assigned to each role on a particular day and who is at work on each shift.

Using role-based devices frees staff members from needing to figure out who they need to locate in order to accomplish certain tasks. One interview participant explained … *"It is a problem that you never know if that person is available at the hospital or is on leave; on a three weeks holiday. All messages of any importance – they cannot just be sent somewhere to a pager that is not used for a week. That is the problem of connecting a message with a person and not a role."* [Department Head - B]. Dedicated role-based devices also make it easy for staff members to switch roles when necessary by handing their pager to a willing colleague.

DECT wireless phones. Some more senior physicians have a personal wireless phone available to them that is kept in their office when they are not using it. The physician can provide the number to the phone to anyone inside and outside the hospital and can also request that the number be listed in the internal hospital directories, or limit interruptions by keeping the number "unlisted" if they choose.

There are also some roles for which a wireless phone is assigned instead of a pager. This is done for roles associated with being "on-call", implying that the staff member must be immediately contactable at all times. There are two such on-call roles at the department, one for the inpatient ward and one for the radiation section. The residents interviewed in stage 2 both mentioned having a role-based wireless phone about one day per week. Role-based wireless phones were used at least occasionally by all of the physicians interviewed except for Department Head - B. When asked why she used a pager instead of a wireless phone in relation to her role as department head (Department Head - A used a role-based phone) she replied "*Because I haven't asked and no one has offered me [a phone].*" [Department Head - B]. This suggests that in her case a phone was not viewed as large enough of an improvement to warrant asking for one.

3.3 Balancing Interruptions and Communication Availability

Interview participants were asked about their use and opinion regarding pagers and wireless phones. As expected some users reported phones to have clear advantages. One of the attending physicians explained "*I have a pager, personal, and I have a wireless phone, as well. I got this [the phone] mainly because I have an office where there was no phone installed. So this is the phone I have in my office. It is this one that I use, always, and I find it incredibly neat to have a wireless phone. It eases – it is a considerable ease in my working day. Simply because there are many places where there are no available phones, so you save a lot of time by having your own phone, I think. Because, if you are on your way to the Radiology department and you need to call someone, or get paged, then you may talk while you are walking through the hallways. You don't have to search for a phone. For example, when you are at the radiology demonstration it is far to get to the nearest phone, so it eases the work to have your own phone and you can do the work directly – or just outside the door.*" [Attending - A]

The physician explained however that some other physicians did not share this view. *"There are several at the department who … use the phone only to phone out; where their number is not registered [in the directory]. Who do not want – who wants to get paged - and use the phone to call back. It is the switchboard who prepares the directory and in this case it only lists the pager and the office number and not the wireless phone number."* [Attending - A] This provides for a great deal of flexibility on controlling interruptions. Some of the physicians had an even stronger view against carrying a phone and refused to carry one even when it was available to them… *"There are some of the attending physicians that absolutely don't want a private [portable] phone, because they feel they never get to be undisturbed. So there are always advantages and disadvantages with availability."* [Attending - A]

One of these physicians was Department Head - A. He was adamant about not being forced to carry a phone with him. He expressed concerns that this would lead to a loss of control regarding his availability, and an environment where he would need to constantly explain why he refused to answer certain calls.

Thus, although he uses a role-based phone as head of the department he rarely takes the phone outside of his office. Instead, he forwards calls from the phone to his individual pager so that calls to the phone will result in a page that he can choose to answer or ignore. He commented…"*I can imagine that it can be a problem to be accessible at all times. Sometimes it is necessary to have quiet and not get disturbed by phone calls. I do have a DECT-phone, but I don't bring it to the [inpatient] ward because it would disturb and I always have the pager.*" [Department Head - A]

Although attending physicians at the oncology department were able to avoid carrying phones in order to control their availability, this was not always the case with other personnel. The most dramatic documented example of phones leading to over-availability occurred in relation to the use of the "on-call" phone used in the radiation section. The work-practice in the radiation section is run in a highly efficient and assembly line manor with the nurses and radiation technicians having a larger role in planning the overall work practice than in the inpatient and outpatient ward.

As such, they have assigned a DECT-phone to the physicians from oncology that work there to make sure that one of them is always easily reachable to answer general patient questions. One of the residents commented on this phone while being interviewed and stated … *"when I am down there [radiation section], as I said, there I have the [portable role-based] phone; I get called a lot – we are talking about more than 20 times a day when I get a call on the phone..."* [Resident - B]. Observations conducted in the radiation section that were completed after the interview revealed the phone to be something that is viewed as extremely interruptive by the physicians working there. This has resulted in a policy where the least senior of the physicians assigned to the radiation section is always assigned this phone.

A resident (that was not interviewed) observed to be carrying the phone during observation sessions even referred to it as "*the interrupter*" phone when asked about it. She commented that she felt the phone was unnecessary and that it resulted in her constantly being "*nagged*" about small things related to patients such as a headache etc. that could easily be handled without contacting a physician. She explained that these interruptions often significantly increased the amount of time needed to plan out radiation treatment because of the high degree of concentration needed to complete this task, in combination with the frequency of interruptions. The other physician assigned to the radiation section during the observation session was in general agreement with these statements.

Such a negative view was not associated with the on-call phone used at the inpatient ward however. This phone was used specifically for handling medical emergencies, and for calls related to external relations with GPs and physicians from other hospitals. These calls are viewed as something that necessary must be handled by a physician. The problem of being "*nagged*" had been avoided at the inpatient ward because the doctors had a larger control over the work situation there, and as such had made it clear to the nurses that they should only contact them with the on-call number in the case of an emergency. This illustrates how physicians will accept a loss of control over their communication availability as long as the associated interruptions are limited to those that are justified by medical necessity.

Interruptions from Calls. An issue further exasperating this problem is that individual calls are deemed to cause a greater interruption than a page. This opinion was shared by all of the participants, including those that prefer a phone over a pager in many situations. There were several reasons given for the higher level of interruptibility associated with a call. One issue was design differences between pagers and the DECT-phones used by the department. A pager has a screen and button located on the top of the device so that a user can easily interact with it without removing it from their coat pocket. A phone on the other hand requires a user to pick up the device in order to view who is calling them, and to locate the button used in order to reject a call. Another issue is that DECT-phones made for hospitals are relatively simple in comparison to standard cellular phones. The phones used by the department for example do not contain a "meeting mode" feature that allows users to limit calls to one ring. One user explained this by using hand gestures to illustrate how much easier it is to deal with a page while saying … "*the pager I can with pressing a key once turn off the calling sound, while the phone rings and rings*" [Resident - A].

Issues that are not be so easily attributed to specific design differences between phones and pagers used at the department also seem to be an issue. Department Head -A explained that his preference for a pager instead of a phone is based on the belief that individual calls result in a more severe interruption than a page. He forwards phone calls to his pager instead of carrying his phone with him while in the inpatient ward because he feels that each call forces him to immediately make an active decision about whether or not to answer it. This creates more stress and greater mental interruption for him than a page because he can still wait to finish what he is doing before deciding if he will answer the page promptly or not. Another physician made a similar comment and explained …"*with a phone it is easier to take the call and explain that you will call back later. I think I would do so, if I have a phone. So, that could be a disadvantage with the phone; that you may get interrupted and allow yourself to get interrupted. You get more easily interrupted by a phone than a pager.*" [Resident - B].

Another problem associated with phones was that callers tend to call back repeatedly in a short period when a call is not answered, even if it is not for an important matter, whereas this behavior was not typical with pagers unless the page was for a serious problem that needed to be dealt with immediately. This suggests that carrying a phone can automatically imply to others that you *should* answer, at least in order to tell them you are busy and will call back later. It also shows how one advantage of pagers is that they make it easier for a person to keep a certain communication distance by delaying or refusing response to certain pages.

Refusing or Delaying Response to a Page. Selective response to pages based on their origin and frequency has been noted previously as a screening procedure used to limit interruptions [8]. During the interviews participants were asked about situations where they may choose delaying or refusing response to a page or call. The participants explained that this behavior was most common when meeting with patients because this was a situation where being contacted was considered to be particularly interruptive. One participant explained … "*If, for example, you are in an important consultation with a patient, talking with the patient about serious issues, then it would be wrong to interrupt that conversation to answer a pager that may not* *be important."* [Resident - B]. An important clarification is that this behavior only occurred when contacted on devices assigned to them as an individual, or for certain roles such as department head where being immediately contactable at all times is not viewed as an aspect of the role. This was considered unacceptable however for devices attached to a role which required them to be "on-call" as explained by one of the attending physicians… "*In general you can say that if you are on call or have the secondary on call duty, you are obliged to respond immediately. Otherwise, I don't think there are any special [rules]."* [Attending - A]. This suggests that this behavior will be much less common in departments where emergencies are more common than the oncology department.

Selective response was not generally viewed as negative by the participants as there was a clear understanding of the need for others to ignore some pages in order to control their availability. The participants were asked if they knew of any hospital policy requiring them to answer pagers within a certain time frame. No such policy was known to any of the participants and they often took a defensive tone when responding to the question. One of the participants explained … *"I don't think any would invent such a rule. It can't be like that. … there are no such rules that you need to obey. I hope not, and I have never heard of such, and it is not reasonable to have them."* [Resident - B]. This further emphasizes that clear need of the staff members to manage their own communication availability.

Prioritization of pages received on a dedicated role-based device illustrates that one advantage of using multiple devices is that it enhances the ability of users to discern the importance of certain pages and calls in order to manage interruptions. It also suggests that a mechanism allowing users to quickly identify which number has been called in order to contact them (i.e. a role-based number or their individual number) will be a critical feature of any pervasive device that integrates role and individual based contact.

The "Tragedy of the Commons". One important trend that showed up in the interviews was the perception that, as a physician gained more responsibility, interruptions from calls would become a larger problem. The two residents both commented that while they would prefer to have an individual phone now, they are unsure if they would still want one once they have more responsibilities. When asked if he would prefer a phone one resident explained… *"It is easier with a phone. I think I would have selected the phone, but I think one may change one's opinion when you have been here, in 'the game', a while, perhaps as you rise in position and get more responsibility and more people are interested in getting hold of you to consult with you, then it may – I think there may be a lot of bothersome requests."* [Resident - B]

The other resident stated … *"I see that the attending physicians that carry a phone get an extreme amount of calls. Whether this is because they are more easily available to others, to other physicians, or to other people in the hospital who know that you just need to punch in the number and then you get hold of them immediately and you don't have to wait for the person to reply on the pager. It is good to be available, but if it is too easy I think you would get a few requests that are not necessary to deal with then and there."* [Resident - A], and when asked if she thought that she would be interrupted more if she had a phone she jokingly stated… *"No, as a resident I* *wouldn't mind having a phone. Then I could give it away when I become an attending. No, I wouldn't do that."* [Resident - A]

This may further explain why Department Head - A was so against carrying a phone with him at all times.

Text Messaging. As stated earlier Department Head - B was the only person we interviewed that never had a wireless phone available to her. When asked about the possibility of being provided a phone her initial response was *"I think I would prefer a phone or a pager with possibility to have more messages."* [Department Head - B] She continued by stating that she would prefer to have a phone instead of a pager if she were forced to choose, but her response clearly showed that mobile text-messaging was a technology she was more interested in being provided than wireless phones. Text messaging is an interesting technology to explore because it may serve as a nice "middle ground", allowing for a wider variety of communication than a simple page without creating interruptions similar to that from a call.

After discussing the use of phones and pagers the interview participants were engaged in a discussion about the introduction of mobile text messaging. Many of the staff interviewed viewed the idea positively although a few had reservations. Interestingly the participants that were positive about text messaging expressed that its main advantage would be that it could help them manage availability and more effectively deal with pages and calls. One participant explained … *"...if you could give a message that you are occupied for a quarter of an hour, or let them know what you are doing – I think that could be useful. It seems to be of some help. Absolutely! Then you would know what the situation is and when you can reach that person, and how long that person is unavailable. That would have been progress, if you had managed to something like that."* [Resident - B].

Another comment was …*"Yes, that could be useful. You could use this [a messaging device] if you are in the outpatient ward and dismiss any calls or send a message to say you will call back later."* [Attending - A]

Other comments suggested that this could be accomplished by using the messaging system in order to provide meta-information about the origin and urgency of pages. One comment was …*"Today we only get a number. If you don't know that number it is not possible to know where the page comes from. If you could get a more of – at least what department it comes from or which room the call is made from, then you would, often, get a notion of what this is about."* [Resident - B] and another was *"One thing I have thought of is that it would be nice if I could get some more information [when paged] to understand the urgency of the call: If this must be replied to immediately, or can wait 10 minutes or needs a reply today. … to have some priority – do you understand what I mean?"* [Resident - A]

Some participants were skeptical of the idea but this seemed more because they were worried about being bothered by messages from patients or with having to carry an additional device. One respondent stated "*It is our experience that new systems that come as an addition to existing systems are not going to be used. We don't want it to be so that we get a solution in parallel with the existing systems, it must be integrated in what we already have. Let me take an example: We participated in testing [an experimental service] where we got an additional email client [for secure email] – and it was never used.*" [Department Head - A]

Another explained *"It depends on who sends the messages. Should that be the switchboard or the administrative personnel? I think that it is important that it is filtered as we don't want the patients to be able to send us messages directly. We are in general opposed to the idea that the patients should be able to send us email directly. It would be very much to reply to. Then they would expect an availability – then you are committed to reply. I think that some would consider this useful and some who definitely don't want it, just as with the phone."* [Attending - A]

The association with being contacted by patients is understandable since text messaging is an increasingly pervasive way of communicating and interacting in private situations but does not see much professional use in hospitals. Using technologies that are internal to the hospital and not directly compatible with popular personal devices should create a clear distinction between professional and private devices, and thus avoid this problem.

An additional topic covered related to text messaging was if participants thought that menus of short predefined messages would be a useful feature. Previous studies have estimated that standard messages would be sufficient 90% of the time for hospital workers [3]. While the quotes above suggest that standardized messages for coordinating communication and controlling availability would be useful the people we interviewed were skeptical that they would be able to fulfill other communication needs. One participant explained …*"To be able to send a text message could be nice. But I think it should be so that you type the message each time. There are so many situations... "* [Resident - B]. Another comment was *"I don't think we could answer by predefined messages. If I don't get hold of a person, then: 'Call me back'. I don't think we can reply these kinds of requests by two or three words."* [Attending - A]

4 Discussion

Previous studies of wireless phone use in hospitals have focused on settings where the technology is introduced to nurses [17],[25], [24]. These studies reported generally positive results but did mention the possibility of making nurses "fatally available" when deploying the technology in practice [17]. This study has used participatory observations and semi-structured interviews of physicians at an oncology department that have independently started to adopt the technology in their work practice. This has allowed for a more in-depth study of the problem in a naturalistic setting in order to identify effects on physicians, which are a hospital's most valuable information source.

The interviews and observations conducted at the oncology department revealed a general concern among the physicians about wireless phones leading to more frequent and more severe interruptions than pagers. This suggests that if not carefully deployed, they may have negative overall effects on the work practice for at least some physicians. One specific instance of over-availability was documented for example, where physicians had been assigned a phone by the team of nurses and technicians at the radiation section of the department in order to make it easier to access them for general questions. This "interrupter phone" was viewed in an extremely negative way by the physicians assigned it. They stated that it was not justified in terms of improving patient care, and that it constantly interrupted work that requires a high level of concentration.

Whenever possible the physicians at the department adopted various strategies for obtaining benefits from wireless phones while avoiding interruptions. This included limited phones to use for outgoing calls by refusing to provide other staff members with their wireless phone number. Some physicians also showed a general preference for pagers by avoiding carrying a wireless phone altogether.

This suggests that, while wireless phones do provide certain advantages, further study is needed on how hospitals can use them most effectively. For example, it is becoming increasingly common for workplaces to switch completely over to wireless phones, since the cost of deploying and maintaining wireless networks can now compete with landline phones in many situations. Our data suggests that abandoning office phones in favor of a "pervasively wireless" hospital could have risks, and that this should not be recommended without further study. In particular this may create problems with over-availability, especially for more senior physicians.

The limited functionality of the paging system does suggest however that the staff would benefit from increased use of messaging and phones. This could provide a number of benefits including improving the ability to discern information about the urgency and origin of pages, making it possible to contact other staff members directly without needing to locate landline phones, and making it easier to quickly inform others about their communication availability by sending messages such as "Call you back in 15 minutes."

Clear resistance to non-integrated solutions requiring users to carry additional devices however suggests that a single device integrating text, voice and paging services would be advantageous. The critical nature of hospital work however will require that such a device have a high quality of service reputation if it is going to replace pagers. This limits the choice of hardware to vendors that create specialized equipment that has a reputation for reliability when deployed in hospitals. One example of such a device that does support voice, text and paging is the "Ascom 9d24" wireless phone.

The data suggests however, that some specific technical and organizational features will need to be considered when deploying such a device. A critical component of such a system will be mechanisms that allow users to manage their communication availability. Without such mechanisms devices that support voice services will be resisted by many staff members, and if widely deployed may have negative consequences on work practice.

In addition, the study also revealed some surprisingly rich elements of the existing paging infrastructure including the use of role-based devices that provide functionality similar to role-based context-awareness, and a public charging rack that provides basic presence information about who is assigned to each role. These features will also need to be considered in any future system.

4.1 Design Issues

From the perspective of pervasive computing it is interesting to discuss potential design features for a single integrated device for paging, voice and text services that would be pervasively deployed in the place of the current system based on office phones, and multiple mobile devices for each user. This would free users from having to keep track of multiple devices, and reduce the total number of devices that need to be purchased by a hospital. This could bring significant cost savings until the cost of more complex mobile devices drops to a level similar to that of pagers today. A system based on such a device would need to retain features of the current system while allowing users to effectively manage their communication availability and interruptions. Some features that may help reach this design goal are:

- The ability to configure the ringer separately for individual and role-based communication. This will allow users to continue screening calls and pages in a similar way to the behavior noted with multiple devices.
- The ability to configure the ringers separately for voice, paging and text messaging. This will allow users to control their availability by for example, turning the ringer off for voice services while still being reachable via the paging system.
- The ability for users to control the distribution of their individual number for voice services. This will require the use of a separate number for voice services than pages or messages so that the person can still be contacted by those that do not have access to their voice number.
- A presence awareness mechanism that allows users to easily identify who is assigned to each role on a daily basis. This functionality would be similar to that of the charging rack that is currently used at the department.
- An easy way for users to transfer role and have communication requests for the role re-routed to the new person. Today this is possible by simply handing off a role-based pager. One possible way of accomplishing this with a system based on a single device for each user would be to create wearable role-associated "badges". Using this method the system would automatically route calls to a person's device once they clip the badge on their clothing, allowing staff members to hand off a role by giving the person the role-associated badge. If desired this also could serve as a presence indicator since the badges could give a visual indication of the roles each person is assigned to. Electronic methods for role transfer that do not require handing off of a physical device may also be interesting to explore. This may introduce a number of new issues such as handling of electronic confirmation of delivery, and acceptance of responsibility.
- Predefined short messages for managing availability and coordinating communication. This would include messages such as "Call you back in X minutes" and messages that improve the ability to identify the origin and urgency of pages.
- Support for at least some limited interaction without removing the device from a coat pocket.

An additional strategy that may prove useful for managing interruptions and coordinating communication is to augment the system with context-awareness. Context-awareness is closely associated with ubiquitous and pervasive computing [9] and can be applied to communication systems in a number of ways. The data from our study suggests that one of the most useful applications of context-awareness in mobile communication for hospitals would be to apply it in order to limit interruptions from calls. Previous work suggests however, that users tend to use contextual information regarding the availability of others as a presence awareness tool instead of to reduce interruptions [12]. This suggests that in order to manage interruptions effectively context will need to be applied in a way that automatically configures devices, for example by turning the call ringer on/off when appropriate.

Statistical models based on sensor systems that collect information such as position, time of day, the number of occupants in a room and the existence of speech have shown to be as effective as people at estimating interruptibility in office environments [11]. These metrics have not been investigated in a hospital setting however, and further studies are needed in order to measure how effective they will be for managing interruptions during real use.

Automatic detection of communication availability may also be advantageous when used in combination with the messaging system in order to send status messages such as "Busy. Call back in 20 minutes" when blocking calls. This would provide presence and among other things discourage multiple call back attempts and/or pages for non critical matters.

4.2 Future Work

Our eventual goal is to design a communication system based on a single pervasive device for voice, paging and text that can be deployed hospital wide at University Hospital of North Norway. This will require consideration of technical, social and organizational aspects of the system. The data collected in this study points to a number of interesting directions for future work in this regard.

As a first step we plan on deploying and studying the use of text messaging services. This is viewed as a first step since a basic implementation without advance features controlling interruptions would not introduce as many risks as wireless phones. Many of the design issues described above can also be viewed as future work and will need to be investigated before recommendations about their affect on usability can be made. For example, a wearable badge system for switching roles may create problems if users end up forgetting badges more often than they forget communication devices. Another interesting problem worth investigating is the necessary level of accuracy needed when detecting the appropriate moment for interruptions.

In addition, we will need to expand this study to include the input of other staff members at the hospital including nurses, technicians and individuals from other departments. Another aspect of mobile communication in hospitals that is not examined in this paper is the impact that the introduction of wireless phones may have on the privacy of patient data. Allowing doctors to communicate via wireless phones walking down the hallway for example increases the possibility that someone may overhear a discussion that would interfere with doctor patient confidentiality.

References

1. Acuff, R.D., Fagan, L.M., Rindfleisch, T.C., Levitt, B.J. and Ford, P.M., Lightweight, Mobile E-Mail for Intra-Clinic Communication, AMIA Annual Symposium, (1997).
2. Anderson, J.G., Aydin, C.E. and Jay, S.J., Evaluating health care information systems: methods and applications. Sage Publications, Thousand Oaks, (1994).
3. Bardram, J.E. and Hansen, T.R., The AWARE Architecture: Supporting ContextMediated Social Awareness in Mobile Cooperation, ACM CSCW '04, (2004).
4. Beyer, H. and Holtzblatt, K., Contextual Design. Morgan Kaufmann, (1998).
5. Blum, N.J. and Lieu, T.A. Interrupted care: the effects of paging on pediatric resident activities. Am J Dis Child, 146, (1992), 806-808.

6. Coiera, E., Guide to Health Informatics. Arnold publishers, (2003).
7. Coiera, E. When Conversation Is Better Than Computation. J Am Med Inform Assoc., 7, 3 (May-Jun 2000), 277-286.
8. Coiera, E. and Tombs, V. Communication behaviours in a hospital setting: an observational study. British Medical J, 316, (Feb. 1998), 673-676.
9. Dey, A.K. Understanding and Using Context. Personal and Ubiquitous Computing, 5, 1 (2001), 4-7.
10. Eisenstadt, S.A., Wagner, M.M., Hogan, W.R., Pankaskie, M.C., Tsui, F.-C. and Wilbright, W., Mobile Workers in Healthcare and Their Information Needs: Are 2-Way Pagers the Answer?, AMIA Annual Symposium, (1998).
11. Fogarty, J., Hudson, S.E., Atkeson, C.G., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J.C. and Yang, J. Predicting Human Interruptibility with Sensors. ACM Transactions on Computer-Human Interaction, 12, 1 (March 2005), 119-146.
12. Fogarty, J., Lai, J. and J., C. Presence versus availability: the design and evaluation of a context-aware communication client. International Journal of Human-Computer Studies, 61, 3 (Sept. 2004), 299 - 317.
13. Grudin, J. Groupware and social dynamics: Eight Challenges for Developers. Communications of the ACM, 37, 1 (1994), 93-105.
14. Katz, M.H. and Schroeder, S.A. The sounds of the hospital: paging patterns in three teaching hospitals. N Engl J Med, 319, 24 (Dec. 1988), 1585-1589.
15. Kidd, A.G., Sharratt, C. and Coleman, J. Mobile communication regulations updated: how safely are doctors' telephones used? Quality and Safety in Healthcare, 13, 6 (2004), 478.
16. Mendonça, E.A., Chen, E.S., Stetson, P.D., McKnight, L.K., Lei, J. and Cimino, J.J. Approach to mobile information and communication for health care. International J of Medical Informatics, 73, (2004), 631-638.
17. Minnick, A., Pischke-Winn, K. and Sterk, M.B. Introducing a two-way wireless communication system. Nursing Management, 25, 7 (Jul. 1994), 42-47.
18. Munoz, M.A., Rodriguez, M., Favela, J., Martinez-Garcia, A.I. and Gonzalez, V.M. Context-aware mobile communication in hospitals. Computer, 36, 9 (Sept. 2003), 38-46.
19. Myerson, S.G. and Mitchell, A.R.J. Mobile phones in hospitals Are not as hazardous as believed and should be allowed at least in non-clinical areas. British Medical J, 326, (2003), 460-461.
20. Parker, J. and Coiera, E. Improving Clinical Communication: A View from Psychology. J of the American Medical Informatics Association, 7, 5 (Oct. 2000), 453-461.
21. Poon, E.G., Kuperman, G.J., Fiskio, J. and Bates, D.W. Real-time Notification of Laboratory Data Requested by Users through Alphanumeric Pagers. J Am Med Inform Assoc., 9, 3 (May-Jun 2002), 217-222.
22. Pratt, W., Reddy, M.C., McDonald, D.W., Tarczy-Hornoch, P. and Gennari, J.H. Incorporating ideas from computer-supported cooperative work. J of Biomedical Informatics, 37, 2 (2004), 128-137.
23. Reddy, M.C., Pratt, W., McDonald, D.W. and Shabot, M., Challenges to Physicians' Use of A Wireless Alert Pager, AMIA Annual Symposium, (2003).
24. Sammon, M.J., Brotman, L.S., Peebles, E. and Seligmann, D.D., MACCS: Enabling Communications for Mobile Workers within Healthcare Environments, MobileHCI '06, (2006).
25. Spurck, P.A., Mohr, M.L., Seroka, A.M. and Stoner, M. The impact of a wireless telecommunication system on time efficiency. J Nurs Adm., 25, 6 (Jun 1995), 21-26.

Using Ground Reaction Forces from Gait Analysis: Body Mass as a Weak Biometric

Jam Jenkins[1] and Carla Ellis[2]

[1] Western Carolina University
Department of Math and
Computer Science
Cullowhee, NC 28723 USA
`jam@cs.wcu.edu`
[2] Duke University
Department of Computer Science
P.O. Box 90129
Durham, NC 27708 USA
`carla@cs.duke.edu`

Abstract. Ground reaction forces generated during normal walking have recently been used to identify and/or classify individuals based upon the pattern of the forces observed over time. One feature that can be extracted from vertical ground reaction forces is body mass. This single feature has identifying power comparable to other studies that use multiple and more complex features. This study contributes to understanding the role of body mass in identification by (1) quantifying the accuracy and precision with which body mass can be obtained using vertical ground reaction forces, (2) quantifying the distribution of body mass across a population larger than has previously been studied in relation to gait analysis, and (3) quantifying the expected identification capabilities of systems using body mass as a weak biometric. Our results show that body mass can be measured in a fraction of a second with less than a 1 kilogram standard deviation of error.

Keywords: body mass, biometric, ground reaction forces, gait.

1 Introduction

Body mass is one of several weakly identifying biometrics used for classification. For example, descriptions of wanted criminals almost always include their height and build, both of which contribute to the helping rule in or rule out suspects. When applying for a driver's license, most states require the applicant to give information such as weight, height, and other physically identifying characteristics. In each of these cases, the body mass biometric is used less to identify the person individually and more to rule out other individuals. How would the weakly identifying ability of body mass change if it could be measured accurately to the kilogram, within a fraction of a second, and without requiring the active participation of the person being measured?

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 251 – 267, 2007.

This study addresses just how this measurement can be taken and explores the expected identification capabilities of using such a measure. Results indicate that the standard deviation of measuring body mass while the person is walking is less than 1 kg. Moreover, this measure can be taken in a fraction of a second and over a small footprint (two 0.5 x 0.5 meter force plates). Two methods are developed for obtaining this body mass. The gait cycle method is restricted to acting on the ground reaction forces experienced during normal gait. The steady state method can be applied more generally to other patterns of forces experienced such as those generated by people passing over the force plates in wheelchairs.

The remainder of this paper will be organized as follows. Section 2 will describe the related work which uses ground reaction forces during gait for identification. Section 3 will describe some of the benefits and drawbacks of using body mass as a biometric. Section 4 will describe two methods which can be applied to actual ground reaction force data to extract the body mass feature. Section 5 evaluates these methods to determine the level of precision and accuracy obtained on a set of clinical walking trials. Section 6 will quantify the expected identification power on three different populations. Section 7 will describe potential future work and section 8 will conclude.

2 Related Work

Gait analysis is emerging as a promising biometric identification technique [8]. A wide range of media and techniques have been developed to classify gait sequences. The two categories of gait analysis most directly related to the work presented in this study are those which are based at least in part body size and/or shape, and those which analyze ground reaction force patterns.

In the system by Bobick and Johnson [10], three body measures are obtained from video: overall height, leg length, and torso length (including head). These measures are combined with a stride length parameter to create a feature vector. In the system by BenAbdelkader et al [11], stride length, cadence, and height are used as the features extracted from video sequences.

Ground reaction forces measured using force plates have also been used in a variety of gait recognition systems. The ORL Active Floor [1], the Georgia Tech Smart Floor [2], and the floor developed by Middleton et al at the University of Southampton all use only the impact of footsteps for identification purposes. In the biometric identification system developed by Cattin [12], data collected from force plates is combined with video recordings in order to improve the recognition rate over using ground reaction forces alone.

The Active Floor uses hidden Markov models to identify individuals. This system was able to achieve a 91% identification rate on 15 individuals. In later experiments the ground reaction forces were scaled by the inverse of the average force experienced during one footstep. This normalizing procedure was aimed at removing the influence of body mass on the identification rate. With the scaled data, the identification accuracy dropped below 50%. The Active Floor paper suggests that using different segments of the gait cycle may produce more accurate identifications.

The Smart Floor extended this work in many ways. The testing using the Smart Floor was much more extensive because the experiments involved the same people wearing different shoes in order to evaluate the effect of footwear on the recognition rate. Their results indicated that different footwear had a negligible effect on the identification rate. Nearest neighbor applied to ten different features of ground reaction forces for a single footstep was able to achieve a 93% identification rate on 15 individuals.

In the floor developed by Middleton et al, three gait features unrelated to body mass were examined for their identifying capabilities [3]. This system used 1536 pressure switches arranged on a 3-meter long by 0.5-meter wide floor mat. Fifteen people walked up and down the mat 12 times each. The features extracted from these trials were the stride length, stride cadence, and time on toe to time on heel ratio. The third feature alone, the toe to heel ratio, was able to achieve 60% identification accuracy on 15 individuals using four footsteps. When all three features were combined, the identification accuracy rose to 80%.

Cattin improves upon the ground reaction force only systems by capturing video simultaneously with the force plate data. This approach is more robust than other approaches which synthetically combine different biometrics by assuming they are independent. The model used by Cattin does not need to make such a claim because any correlation between the video sequence identification and the force plate identification will be present in his identification algorithm. Through fusing three video sequences with force plate data, a recognition rate of 99.74% is achieved when using 6 walking sequences as the training set and 4 walking sequences in the test set from 17 individuals. The force plate recognition rate alone (not fused with the video sequences) was 93.4%.

All four of these ground reaction force studies confirm that gait has some degree of identification information when measured by floor sensors. What is not clear from the above studies is how much influence body mass has on the level of identification. The Active Floor makes very clear that body mass is indeed a significant part of their identification scheme, but it does not fully characterize the amount of identifying information that remains after the body mass factor has been removed, nor does it address how accurately the body mass biometric can be measured. The Smart Floor incorporates features related to body mass, but it does not evaluate the degree to which these features contribute to the identification rate. The system by Cattin et al attempts to avoid using body mass as a biometric by focusing instead on the dynamic aspects of gait present within the force plate signatures.

Our work focuses on characterizing the contribution that body mass can provide for identification purposes. We neither promote nor dissuade from using body mass as a biometric. Instead, we provide a detailed analysis of how body mass can be obtained from ground reaction forces, how accurate and precise this extracted body mass feature is, and how accurately the body mass feature identifies over larger populations than are typically examined with respect to gait recognition. This analysis may be independently evaluated to examine the potential gains of using body mass as a biometric within different systems with different needs.

3 Benefits and Drawbacks of Using Body Mass as a Biometric

While body mass does not share the strongly identifying characteristic with other biometrics such as fingerprints and iris scans, the body mass biometric does have characteristics that make it an attractive feature to obtain within the context of a biometric fusion system. One of the strongest benefits of measuring body mass as opposed to other more identifying features is that body mass can be measured without the active participation of the person being measured. Simply walking over the force plates registers the person's body mass. While some methods of measuring this body mass require that the person be walking normally over the plates, other methods exist which can measure the body mass regardless of the manner of traversing the plates (these methods will be described in the next section.)

One of the greatest disadvantages of using body mass as a biometric is that the person can easily alter his or her body mass to some degree by carrying or depositing objects. If the typical mass of objects carried (or deposited) is much less than the typical differences in body mass, then a reduced identifying power remains even when using the altered body mass. Such changes in body mass by carrying or depositing objects are difficult to characterize generally and tend to be more domain specific. For example, people entering and exiting a classroom will likely carry in about the same load as they carry out but people entering a grocery store are likely to be carrying larger loads upon exiting than they were upon entering. Section 6 will describe how these differences in the expected load people will carry can affect the expected accuracy of the body mass biometric.

Finally, actual body mass has the property that it can gradually change over time. The amount of time between readings is a good indicator as to the amount of expected change that can occur. For example, over the course of 12 hours, it is very unlikely that a person will be able to gain or lose 5 kg, but this amount of change becomes more and more feasible when the time is extended to weeks or months. For this reason, the body mass measure must be regularly taken in order to remain accurate. Including body mass as a static biometric on documents such as passports will have a weaker ability to identify because of the long duration these documents remain valid. If however the passport were to contain a chip which could be regularly and securely updated, body mass could prove a useful biometric.

4 Methods for Extracting Body Mass

This study examines two basic methods for extracting body mass from the forces produced while walking. The gait cycle method assumes that walking is a periodic motion in which the forces produced are roughly identical for each period. One period of the cycle is observed and the average force over this period is divided by the acceleration of gravity in order to calculate the mass. The steady state method assumes that over a long enough duration of time, the average force generated by any object with any behavior will come to a steady state. This steady state force is then divided by the acceleration of gravity to again yield the mass. The steady state method has the advantage of being able to calculate mass over a broader set of scenarios such as people traversing the force plates in wheelchairs or people using crutches while passing over the force plates. The gait cycle motion method would have to incorporate more complex patterns in order to identify these patterns and appropriate periods.

In order to evaluate the accuracy and precision of these two methods, ground reaction force data were acquired from clinical gait trials [4]. This data contains anthropometric, kinematic and force plate data from 62 children walking normally across two consecutive force plates. Three trials are available per child and the mass of each child is reported directly. To the authors knowledge this database is the largest publicly available collection of ground reaction forces with corresponding body mass measurements.

The force plates used to collect the data for this study are OR6-7 Force Platforms manufactured by Advanced Mechanical Technology, Inc [13]. These force plates are specially designed for use in a clinical gait analysis setting and retail for approximately $10,000 each [14]. Much more appropriate and less expensive force plate designs are described in the Active Floor and Smart Floor papers. Data collected using these custom plates is not readily available, but the complete ground reaction force and body mass data used in this study are well documented [4].

4.1 Gait Cycle Method

The pattern of ground reaction forces during the gait cycle is shown in Figure 1. The first part of this cycle is called *double support*. This is when the vertical force is split between the feet. At the start of the recorded forces one foot is on the force plate and the other is off the force plate. When the weight is transferred onto the heel, this is called the *load response*. Sometimes before the weight is entirely transferred, there is a transient force experienced just before the load response. The force during the load response is typically a force exerted which exceeds the normal standing force. Between the heel coming down and the toe pushing off is what is called the *midstance*. Typically the vertical ground reaction forces experienced during the midstance are below that of normal standing force. When the toe pushes off, this is called the *terminal stance* because this is the last part of the gait cycle where the person's weight is entirely on one foot. Following the terminal stance is another period of double support and the cycle repeats itself.

Ideally, each gait cycle would generate identical ground reaction forces when walking on level ground at a constant pace [4]. If this were the case, then two complete footsteps would be sufficient for characterizing the walking pattern (assuming the period of time of double stance is symmetric). In addition, the average of the left and right foot ground reaction forces over each period would be the same. Since there is no net rise or fall over time during normal walking, this average ground reaction force is the same as the force which would be experienced if the individual were standing still. The gait cycle mean in Figure 1 and 2 would be computed using the forces from time a to d. By dividing this force by the acceleration of gravity, the body mass can be determined. Figure 2 shows this force over 1.5 gait cycles (3 steps).

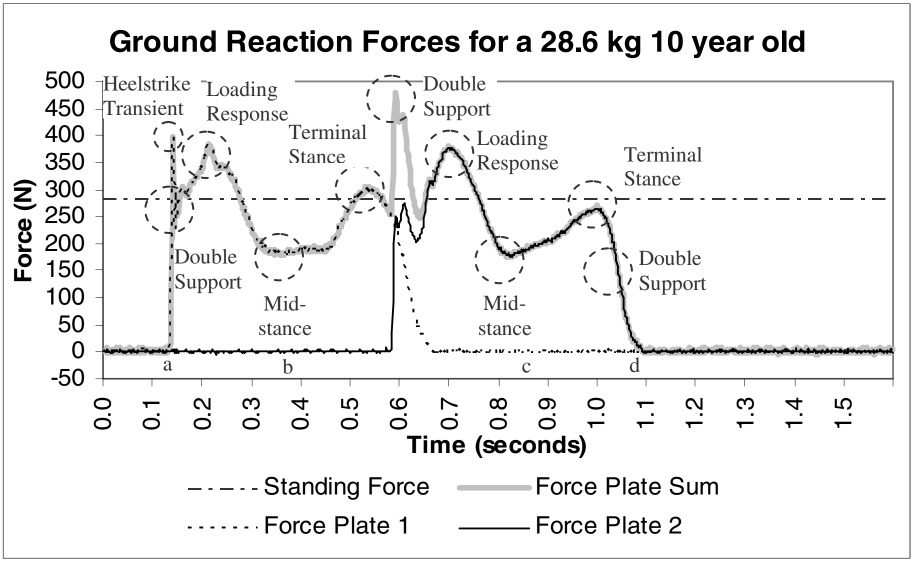


Fig. 1. The ground reaction forces represent two steps from dataset aa_10_01 in [4]. The first step fell directly on the first force plate and the second step fell directly on the second force plate.

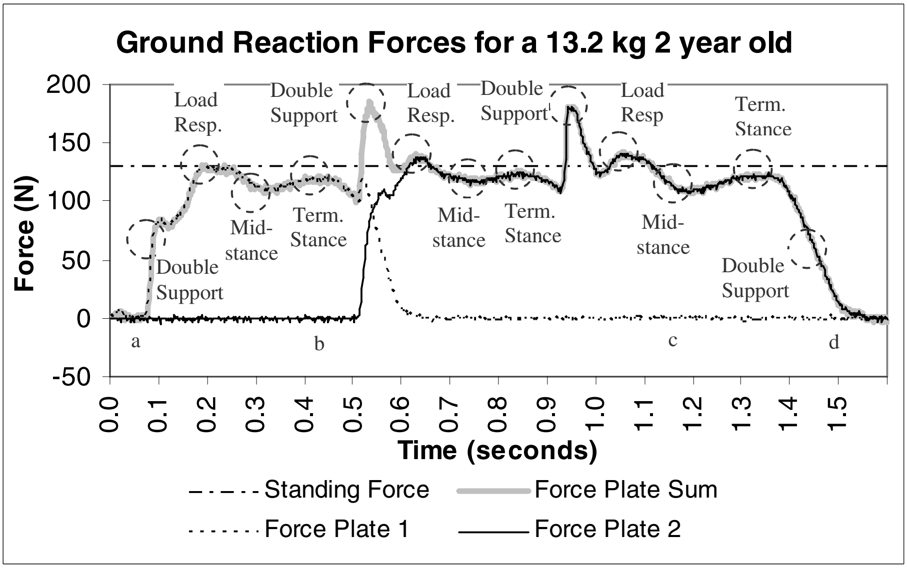


Fig. 2. The ground reaction forces represent three steps from dataset am_02_01 in [4]. The first step fell directly on the first force plate and the second two steps fell directly on the second force plate.

4.2 Steady State Method

During the gait cycle, the sum of the left foot and right foot ground reaction forces vary above and below the force that would be experienced if the person were standing still. Typically, the load response, the terminal stance, and double support experience forces above average standing force and the midstance experiences forces below average standing force. However over time, the average force must tend toward the average standing force regardless of the pattern of forces unless there is a net vertical change in the center of gravity.

When the person steps onto the plate and off of the plate, he or she is transferring forces which may not be measured because they are off the measuring area. This can be seen in Figure 1 where the sum of the forces for the first and last double support is much lower than the middle double support when the person is entirely over the force plates. For the period of time the forces exerted by the person are entirely over the measurement area, average force experienced should be the same as the average standing force if there is no net vertical change in the center of gravity.

There are two competing potential sources of error when using this method to calculate the average force. The first potential error is to start taking measurements before the person is entirely over the measurement area. This would lead to measuring an average force below the standing force. In order to avoid this potential error, as many of the measurements when transitioning onto and off of the force plates should be discarded. The second potential error can occur if there is asymmetry with respect to the gait cycle when the measurements begin and end. Because there is normal rise and fall during walking, if the measurements begin and end at different phases of the gait cycle, then there may be a net vertical change in the center of gravity. This error can be minimized by making the reading range as large as possible so as to make insignificant the contribution of this error to the mean force experienced. Thus the minimizations of the two errors compete. The minimization of the first error requires as many of the start and end of the measurements to be discarded while the second potential error requires as large of a read area as possible. For this reason a balance must be reached between how many of the beginning and ending measurements to discard and how many to retain. As the area of the force plates increases, each source of error becomes easier to minimize. For Figure 1 and Figure 2, the steady state mean would be computed by discarding the first and last quarters of the measurements (time a to b and c to d) and retaining the middle half of the measurements (time b to c).

4.3 Implementing the Methods

The gait cycle method requires the segmenting of the force signals into footsteps. If every step were to directly fall onto the center of each plate then segmentation of the footsteps would be a simple task. However, there are many occurrences of individuals stepping more than twice (such as in Figure 2), stepping partially on a force plate, and stepping between force plates. While this complicates the segmentation, this type of activity must be handled if the force plates were to actually be deployed in a commercial application.

The two consecutive periods of activity in the force plates are considered an integral number of footsteps with no footstep falling between the force plates. For example, in Figure 2, the first force plate receives one footstep while the second force plate receives 2 footsteps. The period of time when both the first and second force plates have active readings is considered the double stance phase. The beginning and end of the signal are also considered part of the double stance phase in which one foot is on the force plate and the other is off the measurement area. Therefore, if the gait cycle is symmetric, then the sum of the beginning and ending of the signal would correspond to a single double stance phase while the other parts of the measure are directly over the force plate. In this way, the sum of all forces over both plates includes one less double stance phase than actually occurred. Therefore, to compute the gait cycle mean, the sum of the forces is divided by the period of activity less the period of time both force plates are active.

The problem of segmenting the forces into footsteps does not exist in the steady state method because there is no assumption about the pattern of signals experienced. The only requirement is that the forces begin and end measurements when the vertical center of gravity is the same upon starting to measure and finishing the measures. Determining this duration of the beginning and ending to discard and how much of the center to keep must be found experimentally. For the purposes of this study, the first and last 27.5% of the active measures were discarded for all individuals. This amount to discard reached a balance between discarding the forces when they are being transferred onto the plates and retaining an amount which makes changes in the vertical center of gravity insignificant with respect to the duration of measures taken.

5 Accuracy and Precision of the Methods

In order to evaluate the ability of using body mass to identify, the accuracy and precision of the different measurements need to be characterized. The accuracy will determine how close to the true value the measures are, and the precision will determine how close consecutive measures are to each other. Both the accuracy and precision are important to evaluate independently. Because the true measure of body mass may be unavailable, the consistency of the body mass estimate determines in part how the measure may be used to classify individuals into groups. If the measure were for example, always 10 kg less than the true measure, it would have perfect precision but poor accuracy. This precision and a predictable accuracy though would be sufficient for classifying individuals. Of course, precision alone is not sufficient. Measuring all body masses as 1 kg would give perfect precision, but the accuracy of the measure is so far from the true value that it could not be used. In this way, it is also necessary to measure the accuracy of the body mass measure.

5.1 Accuracy of the Gait Cycle and Steady State Methods

To the author's knowledge, no previous study identifying individuals by the ground reaction forces has reported the body masses of the individuals. Because the true body mass is known for all of the 62 children from whom the ground reaction forces were recorded, the accuracy of the body mass measure for the two methods can be evaluated. Each of the children made three passes over two consecutive 0.5 meter by 0.5 meter force plates. The mass measures were computed for both the gait cycle and steady state methods. The actual mass was then subtracted from the measured mass in order to determine the errors experienced.

For the gait cycle method, the mean error was -1.18 kg with a standard deviation of 1.78 kg. The maximum absolute error experienced was 11.32 kg below the actual mass and all measures were within a 14.45 kg range (from 3.13 kg above actual mass to 11.32 kg below actual mass). For the steady state method, the mean error was -0.67 kg with a standard deviation of 0.96 kg. The maximum absolute error experienced as 5.66 kg below the actual mass and all measures were within a 7.00 kg range (from 1.33 kg above the actual mass to 5.66 kg below the actual mass). Because the steady state mean experienced far better accuracy and would function across a wider range of force plate signal patterns (such as if a wheelchair would roll across the force plates), only the steady state method will be examined from this point forward.

In order to characterize more fully the errors experienced using the steady state method, the errors were examined to determine if the distribution of error is close to normally distributed. In order to do this, the Kolmogorov-Smirnov statistic was evaluated [5]. This statistic evaluates the maximum deviation from the expected order statistics for any fully specified distribution. For the statistic to be valid, the distribution parameters must not be drawn from the sample being tested itself. For this reason, the first two trials were used to compute the mean and variance parameters for the normal distribution, and the third trial was evaluated to see if there is evidence of deviation from this normal distribution. This experiment was repeated using the first and third trials to compute the parameters and test with the second trial, and finally with the second and third computing the parameters and the first being evaluated. The 90%, 95% and 99% confidence levels for this statistic are 0.155, 0.173, and 0.207 respectively when examined on a set of 62 samples. Values below these numbers indicate there is no strong evidence of deviation from the tested distribution while values above indicate there is evidence of deviation. The three Kolmogorov-Smirnov statistics computed for the errors of the steady state method were 0.124, 0.189, and 0.204. These results indicate that the distribution of error shows slight deviation from what would be expected from a normal distribution. Qualitatively, the errors tend to have a longer tail in the underestimating direction.

In order to evaluate the effect of body mass on the error, correlations between the actual body mass and the error were computed over each of the three trials. These correlations are -0.71, -0.65 and -0.47. They indicate there is a trend toward underestimating the mass more as the body mass increases. In addition, the correlation was computed between the actual body mass and the absolute error. These correlations are 0.67, 0.60, and 0.45. These positive correlations indicate there is a trend of increasing magnitude of error as the body mass increases. These correlations are important to heed because they indicate that the accuracy with which body mass can be measured may be slightly less in adults than it can be in children because of the increased average body mass. The correlations imply that further testing should be done in order to evaluate if this correlation is indeed causation.

5.2 Precision of the Steady State Method

While the accuracy of the steady state method evaluates to some degree the level of identification that measured body mass can provide, the precision of the measures also can also influence the identification level. For example, it is possible to achieve the level of accuracy described in section 5.1 with every body mass measure coming out exactly the same per individual or vastly different per individual (constrained to achieve overall identical mean, standard deviation, max and min over the population).

For each individual, the range of the three values was computed (maximum measure minus minimum measure). The average range of body mass measures over the 62 children was 0.91 kg and the standard deviation of this range was 1.00 kg. The maximum range experienced was 5.05 kg. In addition, the correlation of errors between trials was computed. The correlation of first and second, first and third, and second and third respectively are 0.58, 0.38 and 0.38. This seems to indicate that there is some level of clustering of the errors which will make the measures more precise than the standard deviation of the overall accuracy indicates.

6 Identifying with Body Mass

There are two primary factors that influence the ability to identify using body mass – the distribution of body mass within the population, and the combination of precision and accuracy with which this measure can be obtained. The distribution of body mass in the population will not only affect the overall identification performance, but this distribution will also affect how identifiable each individual is. For example, uniformly distributed body masses provide the greatest overall level of identification due to the maximal dispersion and every individual is equally identifiable (with the exclusion of those at the extremes). If the body masses were instead normally distributed, then the individuals with masses in the tails will be the most identifiable and there would be a lesser overall identification accuracy when compared to uniformly distributed masses. This overall and individual identification level needs to be exposed in order to understand fully the amount of identification information provided by body mass.

The precision and accuracy achieved in clinical trials was explored in Section 5. The actual experience in the field may be quite different. People may carry more or less load on their person causing the body mass measure to experience greater magnitudes of error. Moreover, behavior outside of controlled trials may introduce other unexpected errors. For this reason, the identification power of body mass must be explored in a manner which addresses these domain specific sources of error apart from the measurement error.

These interactions of the distribution of body mass in the population and the model of error are explored in Figures 3-6. Figures 3-5 describe the level of identification on individuals with a given mass while Figure 6 describes the overall expected identification power given a continuous range of error models. Figure 3 shows the percentage of other children in the clinical trials who have an actual body mass within a given range. For example, the solid black line shows that there are about 5% of the children between 19 and 21 kg (a 2 kg range centered at 20 kg). If children being identified were rejected when the measured mass was more than 1 kg different from what is expected, then there would be a 5% false accept rate for a 20 kg child. However the false accept rate under the same scenario but for a 50 kg child would be near 0%. In both cases, if the body mass measures were within +/- 1 kg 90% of the time, then the false reject rate for all children would be 10%. Figure 4 explores this same interaction, but over a larger population of children the same age as those in the [4]. Figure 5 explores this interaction over about 4000 adults [6].

Figure 6 explores the overall expected identification power for a continuous set of ranges for the three populations in Figures 3-5. This graph can be used to describe the percentage of the population that would be expected to be in the given range if a person were selected randomly from the population. Stated another way, this is a mapping of the expected false accept rate for a given false reject rate and given a measurement error model. For example, if 90% of the masses measures are +/- 5 kg from the true body mass (a 10 kg diameter of range), then approximately 30% in the child population would be accepted in this range. This roughly translates into mapping a 10% false rejection rate to a 30% false acceptance rate for the population described. Figures 3-6 are all functions of the distribution of body mass within the population, but they can be used to estimate the expected false rejection and false acceptance rates under general measurement error models so long as the error is symmetrically distributed around the true measure.

For the purposes of directly evaluating the identification power of the ground reaction forces in the child database in particular [4], the following experiment was performed. For each of the three trials, the steady state mass was computed. The body mass measures for the first two trials per child were averaged to create a baseline mapping of body mass to child. The body mass computed in the third trial was then ranked as to how close to the baseline measure it was. A rank of 1 would mean the baseline measure for the child who generated the ground reaction forces was the closest to the measured mass in the third trial. A rank of 2 would mean the baseline mass was second closest to the measured mass of the child. The procedure was repeated for the remaining pairs of trials generating the baseline measure and testing against the measure left out.

The results from this experiment are shown in Figure 7. Of the 62 children, 39% of the time, the correct child was the one with the baseline body mass closest to the measured mass. 85% of the time, the correct child was within the 4 closest baseline body masses. The Active Floor and Smart Floor studies have shown approximately 90% identification accuracy on 15 individuals. The identification accuracy is defined as when the person who generates the footstep is the rank 1 individual. The trials in the Active Floor and Smart Floor studies clearly segmented individual footsteps by requiring the participants to place each step directly onto the center of the force plates.

19 of the 62 participants in the trials examined in this study had at least one trial where at least one footstep did not fall directly on a single force plate. The remaining 43 participants had footsteps that could be clearly segmented in the same way the footsteps were segmented in the Active Floor and Smart Floor studies.

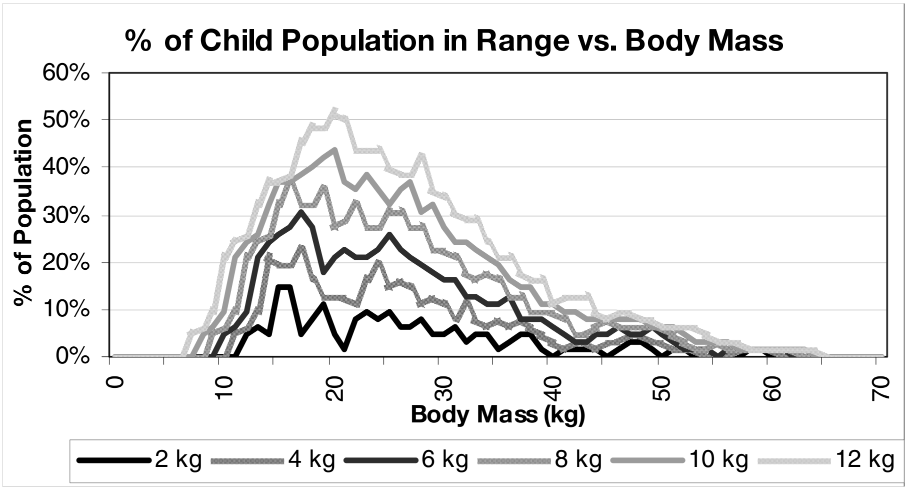


Fig. 3. This graph covers the population for which ground reaction forces were available. The horizontal axis represents the center of the range and different lines represent the range magnitudes.

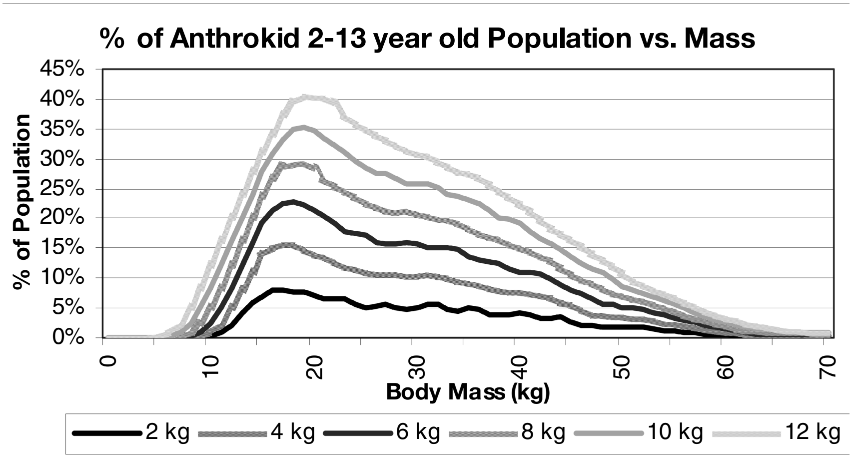


Fig. 4. This graph shows the distribution of body mass over 2924 children the same age as those in Figure 3. The data is drawn from a publicly available anthropometric database [7].

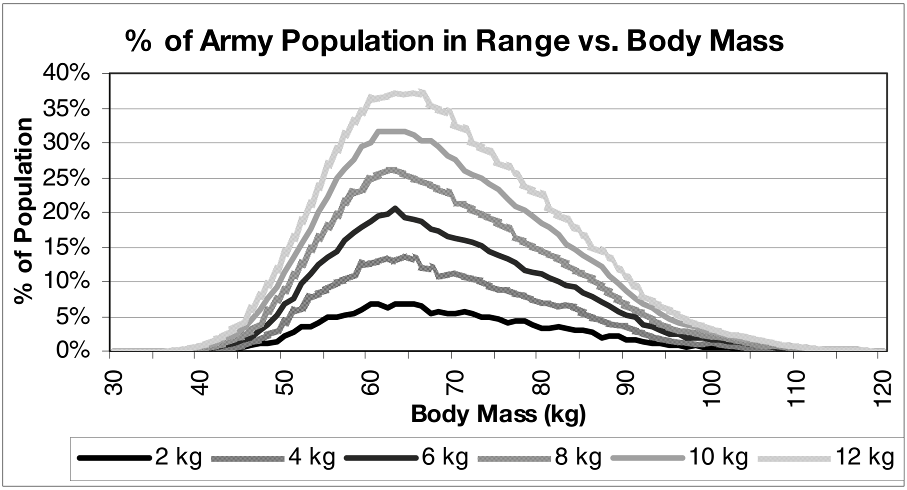


Fig. 5. This graph covers a 1988 anthropometric survey of 3982 men and women in the United States Army [6]. To the author's knowledge, this is the largest public domain database of adult anthropometric data.

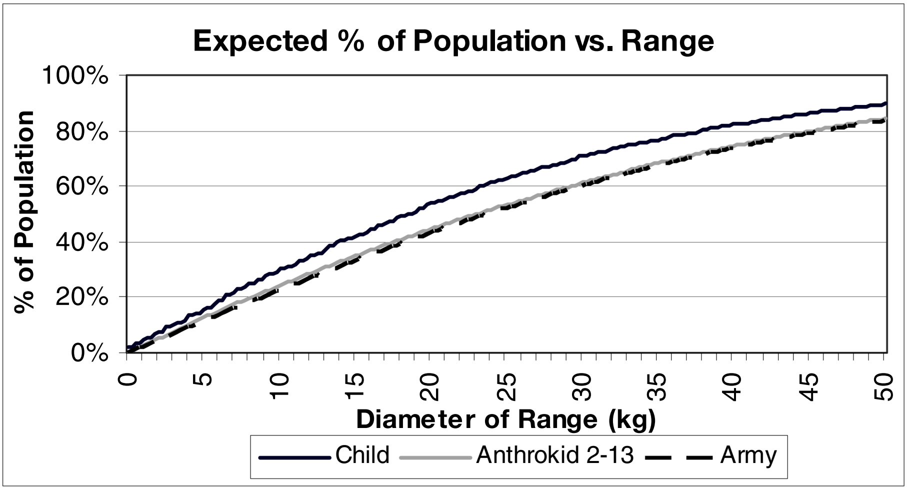


Fig. 6. This graph represents the expected percent of the population whose body mass would fall in the given range. The three populations shown are the same as the ones used in Figure 3, 4, and 5 respectively. Given a distribution of the error in determining a person's body mass, this graph can be used to approximate the interaction between the false accept and false reject rates.

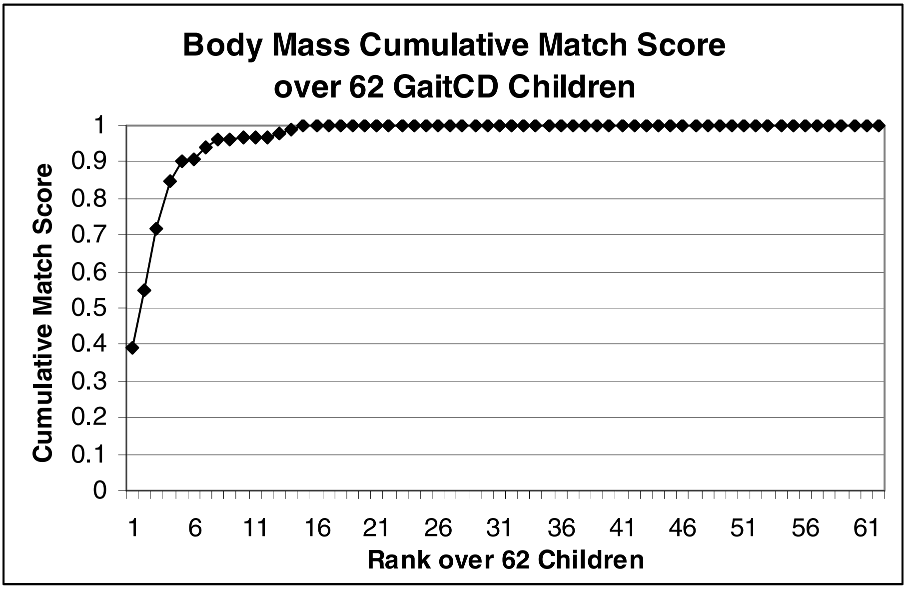


Fig. 7. This shows the identification rank of the 62 children when identified by body mass only. The cumulative rank is how much of the population was ranked at that level or less. No person was ever ranked higher than 15.

The identification algorithm used in the Smart Floor was applied to the data from the 43 participants who had clearly segmented footsteps. The steady state method for measuring the body mass was also evaluated on these same 43 individuals. The Smart Floor recognizes individuals using a single footstep whereas the steady state method uses a single pass over the two force places. For this reason, the distance measures for each footstep described in the Smart Floor algorithm were combined in four different ways in order to match a pair of footsteps to an individual. The measures were combined by taking the minimum of the two measures, maximum of the two measures, sum of the two measures, and product of the two measures. The identification accuracy reported is the best performing combination of the steps per trial. The Smart Floor algorithm was also tested when the forces were scaled by the inverse of the average standing force in order to remove the effect of differences in body masses between individuals.

Figure 8 shows the cumulative match curve for the steady state body mass method and the Smart Floor algorithm. The steady state body mass method outperforms the Smart Floor algorithm at all ranks. In addition to computing the ranks, the correlation between the Smart Floor feature distance and the difference between body masses is 0.93. This seems to indicate that the majority of the identification in the Smart Floor algorithm is due to differences in body mass. The correlation between the measured mass via the steady state method and the actual mass is nearly 1.

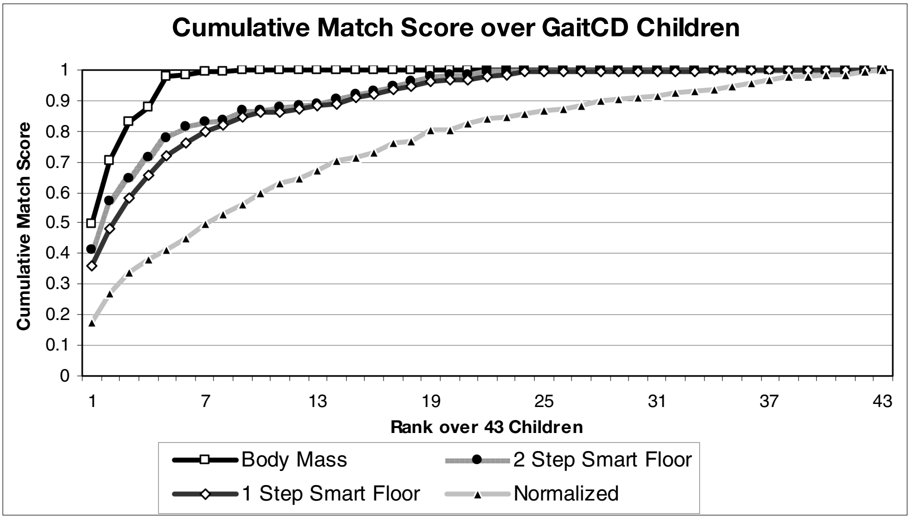


Fig. 8. The figure shows the identification rank of the 43 children when identified by body mass or the features used in the Smart Floor. The normalized curve is the 1-Step Smart Floor algorithm applied to forces scaled by the inverse of the average standing force of the children.

7 Future Work

While this study does establish the accuracy with which body mass can be measured using ground reaction forces on a population of children in a clinical setting, it is unclear exactly what precision and accuracy can be obtained on a population of adults in a normal environment. While the results from the study on children seem to indicate this precision and accuracy will be well within the range necessary to retain identification power, the degree of the identification power will remain uncertain until tested on a population of adults using a more economically designed force plate. Researchers at the Duke University Home Depot Smart Home have constructed a custom force plate and are designing studies to determine how effectively the adult body masses can be measured.

Beyond indicating a person's body mass, the pattern of ground reaction forces may contain other information that can be used to classify and identify individuals. This identifying information could be combined with body mass information to improve identification accuracy. Simply scaling the forces experienced by the inverse of the standing force should be sufficient for removing part of the influence of body mass on the ground reaction force patterns. The Active Floor indicated that this was done and the identification rates fell below 50%, but exactly how much identification power remained was not described in detail. Future work on identifying using ground reaction forces should separate out the influence of body mass before identifying further patterns. If the identification algorithms for using ground reaction forces include a feature distance measure between individuals, one simple measure which could quantify the influence of body mass in the identification algorithm is the correlation between the feature distances and the differences in body mass. The closer this correlation is to 1, the more likely the body mass may be exerting a greater influence in the identification algorithm.

Extracting these patterns from the footstep ground reaction forces requires the segmentation of footsteps from the forces recorded. Previous studies have either required users to step directly onto separate force plates or have selected from a larger pool of footsteps those which are easily segmented from each other. Moreover, previous studies have tended to test only one person at a time. When multiple people are present, the ground reaction forces must be assigned to a particular individual. These segmentation tasks are non-trivial and depend on the granularity of measurement. The larger the force plates are, the more likely ground reaction forces will come from different individuals, but the smaller the plates, the less likely a single footstep will fall on a single force plate. Designs and methods need to be developed for assigning the ground reaction force to a particular person with a group.

Finally, the inclusion of body mass into biometric fusion systems based at least partially on anthropometric measures should be explored. Large anthropometric databases exist for both children and adults. These body measures may be combined in such a way as to provide a relatively strong level of identification. Once the methods for measuring these body parameters have been developed and the measurement error has been characterized, the databases can be used to predict the possible identification power of the system on large populations.

8 Conclusion

The results from this study indicate that ground reaction forces during normal walking are sufficient for measuring body mass in a fraction of a second and with a standard deviation of less than 1 kg. This result is based upon the analysis of data obtained during clinical gait trials on children. If this level of accuracy can be obtained within the general population, then body mass has sufficient discriminating power to be included in biometric fusion systems. The difference in body mass has been shown to be highly correlated with the feature distance of another gait recognition system. Improved performance was achieved by measuring body mass directly and using that feature alone for recognition. Because both the application domain and the selection of force plate hardware impact on the degree of accuracy and precision realized in a given system, this study quantifies the level of recognition possible under more generalized error models.

References

1. M. Addlesee et al. The ORL Active Floor. *IEEE Personal Communications*, vol. 4, no. 5, pp. 35--41, October 1997
2. R. Orr and G. Abowd. The smart floor: A mechanism for natural user identification and tracking. In *Conference on Human Factors in Computing Systems*, The Hague, Netherlands, April 2000
3. Middleton, L., Buss, A. A., Bazin, A. I. and Nixon, M. S. A floor sensor system for gait recognition. In *Proceedings of Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, Buffalo, New York, USA, 2005

4. Christopher L. Vaughan. GaitCD. Available from http://www.kiboho.co.za/GaitCD/. Kiboho Publishers, South Africa, 1999
5. NIST/SEMATECH e-Handbook of Statistical Methods. http://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm, accessed March 2006
6. Final Report, 1988 Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics, NATICK/TR-89/044
7. AnthroKids - Anthropometric Data of Children. http://www.itl.nist.gov/iaui/ ovrt/projects/anthrokids/, accessed March 2006
8. Mark S. Nixon, Tieniu N. Tan and Rama Chellappa. *Human Identification Based on Gait*. Springer-Verlag, 2005
9. Sudeep Sarkar, P. Jonathon Phillips, Zongyi Liu, Isidro Robledo Vega, Patrick Grother, Kevin W. Bowyer. The HumanID Gait Challenge Problem: Data Sets, Performance, and Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* ,vol. 27, no. 2, pp. 162-177, February, 2005
10. A. Y. Johnson and A. F. Bobick. A Multi-view Method for Gait Recognition Using Static Body Parameters. In *3rd International Conference on Audio- and Video Based Biometric Person Authentication*, pages 301-311, Halmstad, Sweden, June 2001
11. Chiraz BenAbdelkader, Ross Cutler, and Larry Davis. Person Identification using Automatic Height and Stride Estimation. *IEEE International Conference on Pattern Recognition*, 2002
12. Philippe Cattin. Biometric Authentication System Using Human Gait. PhD thesis, ETH Zurich, 2001
13. Advanced Mechanical Technology, Inc. OR6-7 Force Platform Brochure http://www.amtiweb. com/brochures/OR6-7lowres.pdf, accessed January 2007
14. Advanced Mechanical Technology, Inc. Personal communications. January 2007

Weight-Sensitive Foam to Monitor Product Availability on Retail Shelves

Christian Metzger[1], Jan Meyer[2], Elgar Fleisch[1], and Gerhard Tröster[2]

[1] Information Management, ETH Zurich, Zurich, Switzerland
[2] Wearable Computing Lab, ETH Zurich, Zurich, Switzerland
`cmetzger@ethz.ch`, `jmeyer@ife.ee.ethz.ch`, `efleisch@ethz.ch`, `troester@ife.ee.ethz.ch`

Abstract. The retail industry, which is characterized by highly complex supply chain processes, still faces stockout rates of 5-10%. This results in sales losses of up to 4% which corresponds to hundreds of millions of dollars for large retailers. The most significant cause for stockout situations is inefficiencies in in-store logistics due to the lack of inventory visibility. In this paper, we present a product availability monitoring system, which anticipates stockouts before they occur and triggers the personnel to replenish the shelf. Our monitoring system is based on inexpensive polyolefin foam, which serves as mount for capacitive sensing elements. Our sensor system is designed for roll-to-roll based manufacturing, which suggests low production costs. Preliminary tests suggest that the system offers sufficient sensitivity to accurately and reliably detect low quantities of stocks. This will not only reduce losses of sales but also increase customer satisfaction.

Keywords: Retail logistics, shelf replenishment, product availability monitoring, capacitive sensors, pervasive computing.

1 Introduction

Providing high product availability at minimal operating costs is a major factor for economical success in the retail industry. Even though retailers have invested in the improvement of their supply chain management processes, recent studies show that unsatisfactory product availability on retail shelves remains a large problem. An ordinary retail store faces stockout rates of 5-10%, which results in a loss of sales of up to 4% [5]. This accounts for a diminution in revenue of hundreds of millions of dollars per year for large retailers such as Wal-Mart, Target or Metro. Customer dissatisfaction and dwindling customer loyalty may further leverage the loss.

In Europe, the two most significant causes for “Out-Of-Stock (OOS)” situations are due to ordering problems and replenishing problems. Situations in which products are in the store but not on the shelf (e.g. in the backroom) cause 38% of all stockouts. Incorrect forecasting and ordering practices account for 32% of OOS situations [5]. Other reasons for stockouts are found in planning-related causes and manufacturing capacity constraints.

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 268 – 279, 2007.

While demand forecasting will always remain subject to uncertainty, sufficient stocks on retail shelves can be achieved through precise product availability monitoring and timely replenishment. Even though most stores are already equipped with electronic point of sales (EPS) systems and take record of deliveries made to the store, retailers still fail to separate backroom inventory from sales floor inventory in their information systems [19], [15]. Therefore, in-store practices such as stock-taking are still carried out manually. Employees visually inspect retail shelves and take written notes on ordering quantities, which are later entered into an electronic ordering system. However, performing manual inspection is both, slow and expensive; and the process is highly susceptible to human error [19].

A study on the duration of OOS shows that only 20% of the products that become OOS are replenished within 8 hours. 61% are replenished within 8 hours to 3 days, while 19% are replenished in more than 3 days [5]. The duration of OOS is the measure that is most meaningful from the perspective of the customer and correlates with the amount of loss. An automatic monitoring system that seamlessly updates the stocks on a retail shelf and informs the personnel before an OOS situation occurs, allows considerably reducing the time span for replenishments. Furthermore, the ability of guiding the personnel directly to the shelf that needs to be replenished saves expensive labor time of visually inspecting the quantity of stocks on shelves.

The retail industry represents a large market with potential to benefit from enabling technologies to reduce losses of sales. However, this industry is hesitant to make larger investments in technology, and therefore mainly economical factors will decide on the successful adoption of pervasive computing technology for product availability monitoring. Another critical factor is unobtrusive integration – customers' purchasing behavior should not be affected by the presence of monitoring technology.

Sophisticated solutions that detect the presence of objects by measuring their weights are already available. Most of these systems are designed and developed with a limited number of highly sensitive sensors to meet the needs of specific applications (e.g. seating and bedding assessments). The retail industry has different requirements - moderate sensitivity at very low cost for wide-area adaptation. In order to keep manufacturing costs low, the design of such sensors must account for large-scale manufacturing processes. Further, the sensor mounting must be flexible for roll-to-roll manufacturing and the sensors need to be mounted at high speed during a continuous casting process.

This paper is structured as follows: In the next section, we review and summarize existing approaches. In Section 3, we describe foam that serves as mount for sensing elements and present a design for capacitive sensors that allows for low-cost manufacturing. Before we conclude in Section 5, we evaluate and discuss the limitations of our design and suggest further improvements.

2 Related Work

Only a limited number of commercially available products address the problem of out-of-shelf situations (e.g. Sainsbury's SAM and the e-Replenishment by IBM). These systems aim at making better use of existing data (e.g. EPS) to track the flow of goods and to estimate on-shelf product availability. However, these solutions rely on accurate sales-floor inventory information, which requires the additional recording of goods that are moved onto the sales floor. This data gathering is expensive because it requires significant manual interaction (e.g. bar code scanning of each product) [19]. Furthermore, this approach disregards inaccurate inventory due to shrinkage (e.g. theft) and the time span between taking a product from the shelf and having it scanned at the check-out.

Currently, leading retailers introduce radio frequency identification (RFID) into their supply chains at case and pallet levels for automatic product identification and tracking. In-store practices such as product availability monitoring, stock-keeping, detection of shrinkage (e.g. theft, dead stock, and misplacements) and fast check-out could also significantly benefit from RFID provided that all items are equipped with tags [6]. However, due to very small margins in the consumer goods industry, tag costs would devour a significant portion of the sales returns and therefore, a high market penetration of RFID item-level tagging is not to be expected within the near future [21], [4].

There are several other technologies, which can provide product availability monitoring on retail shelves. Research on computer vision has developed powerful algorithms for pattern recognition, but these algorithms still require considerable processing power and the technology shows high susceptibility to strong illumination or blocking of view. Infrared sensors can be manufactured at low-cost but photoelectric barriers, which may be arranged in a matrix around the shelf, only provide limited resolution. This approach provides information on whether a specific row is empty but fails to estimate the remaining stocks. Furthermore, bars that would serve as mount for the photoelectric elements require a robust installation on the shelves which increases integration costs. Ultrasound transceivers may offer some degree of information about stock quantities due to multiple reflections from objects. However, these transceivers are expensive and come in large form factors. By contrast, pressure sensitivity suggests high potential due to low manufacturing costs, sufficient resolution, low data processing, and simple integration. Additionally, capacitive sensors offer the ability to be attached to flexible mounts, which allows for inexpensive roll-to-roll manufacturing. Pressure sensitivity is a well-established technology in a wide range of applications, and sophisticated measurement devices are available [2], [7], [14]. However, the commercially available systems are designed for specific applications such as biomedical applications where high sensitivity and high resolution on a limited area are required (system providers are e.g.: Novel, XSensors, Measurement Specialities, and TekScan). These systems are optimized for best performance, which lowers scalability and increases manufacturing costs e.g. Novel's AT-25A pressure-sensitive mat at the dimensions of 70cm x 30cm x 2cm and 2 sensors/cm^2 costs approximately $7,500 [11].

Academic research has developed weight-sensitive floors based on load sensors to identify and track people [1], [12], [3]. However, these applications require large, robust, and inflexible sensors. The "Magic Carpet" by [13] uses piezoelectric wires, which offer some degree of flexibility, but only generate outputs when there is a variation in the applied force. Hence, objects that remain still cannot be detected. Schmidt et al [16] propose the use of load cells in table-legs for context acquisition in an everyday environment. They demonstrate that the position of an object on a surface can be accurately and reliably determined. However, the flexibility of their approach is limited to determining the static position of a single new object, and the integration of robust and therefore expensive load cells requires significant modifications to the environment such as tables, furniture, or floors. Research on capacitive sensitivity for flexible mounts has developed smart textiles where conductive fibers are weaved into the fabrics or electrodes are stitched onto textiles [18], [10], [8]. However, even though these textiles offer high flexibility, this approach comes at significant manufacturing costs. Furthermore, textiles provide instable stands to products with a high center of gravity due to the fabric's compressibility.

3 A Pressure-Based Product Availability Monitoring System

Our design for a monitoring system consists of two parts – the weight-sensitive foam and the measurement hardware. The electronics are separated from the foam so that the same hardware can be connected to multiple foams that are attached to different shelves. In this section, we present the characteristics of the capacitive sensors and the design for low-cost manufacturing, before we describe the specifications of the measurement electronics.

3.1 Weight-Sensitive Sensor

Products on a retail shelf are usually arranged in rows. If the quantity of items in a single row is low, a product is likely to be low on stocks and should be replenished. In order to detect low quantities, we install polyolefin foam on the display area of the shelf and place the products on top. Electrodes on the foam provide capacitive pressure sensitivity to detect the presence of items.

We use capacitive sensors because they provide higher accuracy than resistive sensors, in which particles are often not uniformly distributed [20]. Capacitive sensors also show better linearity, less affection to the number of pressure cycles and history, and lower dependence on temperature and humidity than resistive sensors. Additionally, capacitive foil sensors may be fabricated with off-the-shelf materials in a simple manufacturing process while resistive sensors require special materials and more complex manufacturing.

Electrode leads that run in parallel across the foam on both sides but with perpendicular directions form the capacitive elements at their crossover points (Figure 1) [18], [9]. The weights of the products lead to local deflections of the foam. Such a compression reduces the gap between electrodes on opposite sides and results in an increase in capacitance. Accordingly, the number of products on a shelf is determined through the detection of diverging capacitances.

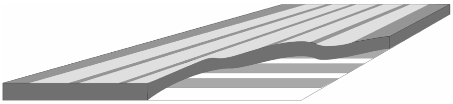


Fig. 1. The leads run in parallel across the foam and form capacitors at their crossover points

We use ALVEO-Soft polyolefin foam as mount for the electrodes. This foam features moderate compressibility and shape recovery for stable stands. Test results for tensile strength (ISO-1926), breaking elongation (ISO-1926), compression load deflection (ISO3386-1), compression set (ISO-1856-C), and shore strength (868-1985, ASTM D2240) suggest good durability [17]. In contrast to non-generic foam such as PVC, PUR or EPDM, it is stable in its web direction, which is an advantage during conversion (e.g. coating, lamination). Polyolefin foam is manufactured as continuous roll foam in very large quantities for the automotive and construction industry. The large-scale production leads to low prices per square meter.

The foam's thickness and compressibility significantly influences the degree of derivations in capacitance, which is inverse-proportional to the gap between two electrodes. We tested numerous polyolefin foams to determine the most suitable foam. While moderate thickness and high compressibility are desirable to achieve high sensitivity for light products, these parameters lead to low resilience and unstable stands for heavy products (Figure 2).

The graph on the upper left of Figure 2 shows the deviation in capacitance when a weight of 11g is placed on a 10mm x 10mm electrode that is mounted on 250µm-thick TEE0300. The weight leads to a steep edge representing a change of about 25fF. When the weight is removed, the capacitance immediately returns to a value, which is 2.5fF higher than its original value. On the lower left of Figure 2, the experiment is repeated with 200µm-thick TEE0400. It shows a slower rising edge, which reaches a change of 15fF. When the weight is removed, the capacitance drops quickly but returns to a value, which is 4fF higher than the original value. We assume that the weight leads to a permanent deformation of the foam. The same effect was observed to be more severely with softer and thicker foams. On the right, we measured the capacitance of 5mm x 5mm electrodes, which are mounted on 200µm-thick TEE0300

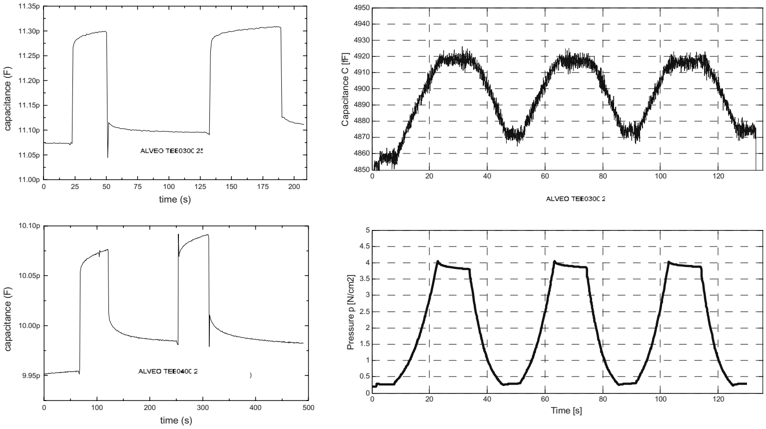


Fig. 2. The graphs on the left show the deviation in capacitance for 250µm-thick ALVEO TEE300.25 and 200µm-thick TEE400.2 with a load of 11g on 10mm x 10mm electrode pads. The graph on the right shows the deviation in capacitance according to a linear change in weight from 0-100g/cm^2 for 200µm-thick ALVEO TEE0300.2 with 5mm x 5mm electrode pads.

foam, according to varying weight. We steadily increased and decreased the weight from 0g/cm^2 to 100g/cm^2 which results in a maximal deviation in capacitance of 50fF. Our measurement hardware can resolve up to 4fF, which corresponds to a weight resolution of 8g/cm^2. However, the noise level limits the actual resolution to about 5fF, which corresponds to 10g/cm^2. Additionally, the graph for 200µm-thick TEE300 shows high resilience (variations remain below the noise level) and sufficient sensitivity. In addition, crosstalk and creep do not exceed the noise level for this design where two neighboring electrodes are separated by 20mm. Therefore, this combination of 200µm-thick TEE300 foam with a sensor grid of 20mm and electrode pads of 25mm^2 suggests meeting the requirements of the retail application.

This design, where the leads not only form the capacitors but also connect several electrodes, reduces the required amount of connection lines (Figure 3) [18]. Therefore, the maximum resolution, which corresponds to the minimal gap between two parallel input leads, is not limited by connection lines that have to be routed in between. Our leads are hand-painted with conductive silver that show a resistance of approx. 2Ohm/meter. However, during manufacturing, a simple continuous casting process would superimpose the leads on the foam.

Fig. 3. Conductive silver leads run in parallel across the flexible polyolefin foam to form the 96 sensors. When multiplexed, the leads on different sides of the foam form the capacitive elements.

3.2 Sensor Measurement System and Data Processing

We built a demonstrator that is 30cm x 20cm in size with 96 capacitive sensors in 12 columns and 8 rows. Each electrode is 5mm wide and separated by 20mm from the next electrode (Figure 3). The electronics contain an AD7745 Sigma-Delta converter from Analog Devices, which measures the capacitances. This converter offers high accuracy of 4fF with an update rate of 90Hz by direct conversion of the capacitance to a digital signal. A multiplexer circuit switches sequentially row by row through all the leads. While the capacitance between two leads is measured, all other leads are grounded to minimize crosstalk and to keep the stray capacitances at a stable value. Parasitic capacitances are compensated directly in the converter.

The system is initialized by measuring each sensor in an unloaded state to compensate for variations in capacitances due to inaccuracies in manufacturing and due to different lengths of the connection lines. These individual values are stored to serve as reference values. Further measurements are compared to the individual reference values, and if a measured sensor exceeds its reference value by a certain threshold value, the system assumes the presence of an object. Once an entire measurement cycle is completed, all loaded and unloaded sensors are displayed. For our demo application, the color of loaded pads changes. To determine the quantity of stocks on a shelf, the loaded sensors per row are summed up. Products are not necessarily displayed in a consecutive order but may be arbitrarily distributed. If the number of loaded sensors per row is small, the shelf needs to be replenished. Our application uses a graphical interface to inform the staff of low stocks (Figure 4). Other ways to alert the personnel such as turning on a light above the shelf or sending a text message to the cellular phone of the sales floor manager may also be applicable.

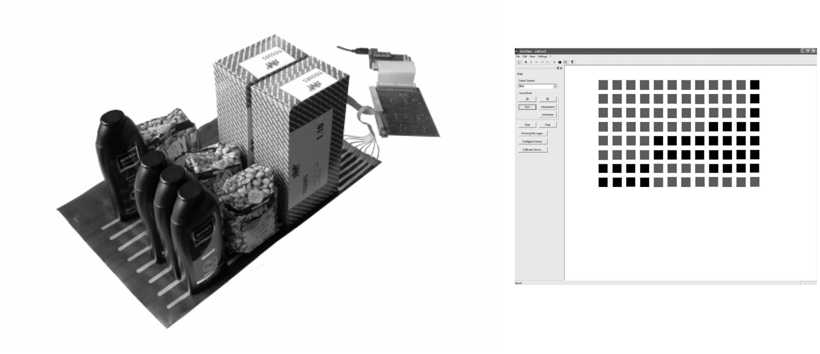


Fig. 4. On the left: Products with highest out-of-stock rate are placed on weight-sensitive foam. On the right: The graphical representation of loaded and unloaded capacitive elements that allow concluding the relative quantity of products on a shelf.

3.3 Cost Considerations of Manufacturing Processes

Economical considerations of system costs are one of the most critical factors for successful adoption of monitoring technologies in the retail industry. In order to estimate the costs of our monitoring system, we accumulate estimated individual costs. These costs contain material cost for polyolefin foam, conductive silver, and protective layers, set-up costs for the manufacturing machines, and costs for the electronics and power supply (note, integration costs significantly depend on the backend inventory system and are not considered here). Manufacturers give prices for polyolefin foam of about \$20/kg. This results in \$0.1/running meter for foam with a density of 25kg/m^3 and the dimensions of 1m x 0.6m x 200e-6m. Conductive silver comes in containers of 60ml (90g) at \$90. If leads are printed according to the grid we suggested above, 40 leads on each side of the foam with widths of 5mm and lengths of 0.6m (depth of a retail shelf) are required. Manufacturers suggest that current casting processes can achieve lead thicknesses as thin as 4µm-20µm for carrier materials with a thickness of 200µm. This results in a volume of 9.6e-7m^3 to 4.8e-7m^3 of conductive silver per running meter or costs of \$1.4-\$7.2. Prices for protective layers of PE foils are about \$0.1/running meter. The manufacturing of the sensors requires about 4 different processes, and each process comes at set-up costs of about \$10. However, we assume that at least 100m of foam are processed at once in roll-to-roll based manufacturing. Therefore, we can conclude total set-up costs of \$0.4 per running meter. Chip manufacturers quote prices for chips that contain measurement electronics and WLAN modules of \$5. It is important to note that a single chip can cover an area of at least two square meters, and thus, the electronics contribute to the total costs with about $2.5. Two AAA batteries for about $1.5 are expected to be sufficient to power the system for years. The total of these individual costs results in system costs of $6.1-$12[1] per running meter. However, it can be expected that prices will drop due to large-scale production.

We believe that our design bears the potential to meet the economical requirements by the retail industry because the costs of the proposed system are significantly lower than the ones for manual data collection and are also lower than RFID item-level tagging where each product is equipped with a disposable 20-cent-tag.

4 Evaluation and Future Work

Most retail stores offer an enormous variety of products. In order to test our design, we restricted the evaluation to products that show significant stockout rates. According to [5], hair care products show highest OOS rates with an average of 9.8%. These products also account for the largest share of estimated retail loss of 4.5%. Out-of-stock rates for hair care products are followed by laundry products (7.7%), diapers (7.0%), feminine hygiene products (6.8%), toilet tissues (6.6%), and salty snacks (5.3%). From these categories, we picked those products that seem to be representative for a specific category and those that suggest the highest challenges for detection. Therefore, we tested our design with products that show significant weights on small footprints on the one hand and light weights on large footprints on the other hand. The characteristics of these products are listed in Table 1.

Table 1. The table shows the weight of the products, the covered area on the foam, the footprint as seen by the sensors (actual pressure area), and the weight per area

Product	Absolute Weight [g]	Covered Area [cm^2]	Footprint [cm^2]	Weight [g/cm^2]
Shower gel	260	2.5 x 6.5	4.8	54
Laundry	2200	8 x 17.5	54	41
Diapers	3000	10 x 22	140	21
Tissues	220	7 x 11	28	8
Peanuts	240	4.5 x 6.5	5.5	44

Our product availability monitoring system does not aim at determining the explicit number of products on the display area, but instead, the personnel is informed when only a small number of sensors per row are loaded because this corresponds to an almost empty shelf (Figure 4). Therefore, it is critical that the system accurately detects and displays the sensors that are exposed to the weights of products. Errors

---

[1] $6.1-$12 (total costs) = $1.44-$7.2 (conductive silver) + $0.1 (foam) + $0.2 (PE foil) + $0.4 (set-up) + $1.5 (batteries) + $2.5 (electronics).

occur when completely or partly covered sensors do not report a load (false negatives) or when sensors show the detection of a product even though they are not loaded (false positives). We determined the error rate for false negatives and false positives by first arranging the previously selected products on the display area and then removing individual items in three different ways – front-to-back, back-to-front, and randomly. After each removal, we visually determine which sensors are covered and compare the result to the actual output of the system. We performed 10 test runs for each product and each different way of removing items (results are listed in Table 2). The tests for shower gel show that 1.1% of all covered sensors do not show a load (false negatives) and 0% of the sensors report a weight detection even thought they are not exposed to any weight of a product (false positives). For laundry bags, the same test shows 2.8% false negatives and 0% false positives, and for diapers, the results are 7.0% false negatives and 1.2% false positives. Peanuts (salty snacks) show the highest error rates with 9.7% of false negatives and 4.1% of false positives. The declared weight per area for tissues is around the minimal resolution of the system, which leads to inaccurate detection. However, in retail stores tissue boxes are usually stockpiled. A minimum of two stacked boxes can be reliably detected by our hardware with an error rate of 2.5% of false negatives and 3.3% of false positives. Note that customers usually take products from the front so that the products in the back of the shelf remain piled. We also tested the sensors' resilience by first loading each sensor and then measuring the false negatives and false positives after the complete removal of all loads. The design shows high robustness with an error rate of 0% for both. This evaluation of the detection of selected product suggests error ranges of 1.1%-9.7% for false negatives and 0%-4.1% for false positives.

Table 2. The table shows the false negatives and false positives for the tested products

Product	False Negatives	False Positives
Shower gel	1.1%	0%
Laundry	2.8%	0%
Diapers	7.0%	1.2%
Tissues	9.7%	4.1%
Peanuts	2.5%	3.3%

False negatives result in early triggering of the replenishment process. They are mainly the result of uneven footprints where the covered area is not equal to the area to which pressure is applied. Especially peanuts, which come in small bags, show uneven footprints. This explains the significant number of false negatives that occurred. All other products show moderate error rates of false negatives. However, the more significant measure for the evaluation of the system is the error rate of false positives. False positives lead to higher accounted stocks than actual stocks.

Therefore, the system risks to not anticipating out-of-shelf situations because it records a stock quantity that is too high. Nevertheless, our preliminary tests show moderate error rates for false positives. In addition, we can account for these inaccuracies by increasing the threshold value at which the system concludes an almost empty shelf. Therefore, the results of this evaluation suggest that the system can accurately anticipate out-of-shelf situations and inform the personnel before an OOS situation occurs.

The reduction of out-of-shelf situations in retail stores is a multi-tier problem. Our system only provides visibility to the quantity of stocks on retail shelves but cannot account for slow replenishment by the personnel. However, in order to compensate for slow response time the threshold value at which the system displays the replenishment notification can be increased. Since the threshold value is hardware independent, it can be adjusted after installation at the retail store. We plan on equipping a complete shelf at a retail store to determine the optimal threshold value on the one hand and to exactly determine the reduction of out-of-shelf situation on the other hand. The in-store shelf testing is also necessary to prove the durability of the system.

Our design offers potential to increase the resolution by narrowing the gaps between leads. A higher number of weight sensors on a smaller area not only allows concluding the exact weight of a product but also allows deriving its footprint. Therefore, we plan on incorporating pressure pattern analysis to determine individual footprints. This will allow deriving the exact number of products on a shelf.

Misplacements account for sales losses of approximately 2% [6]. A system that detects weight and footprint of individual products can locate misplaced items if their weights and footprints differ from those of the other products on the shelf. However, the detection granularity is often limited to product categories because many products are sold in standardized packaging (e.g. cans). Thus, a system based on weight and footprint sensing will not significantly reduce the number of misplacements.

5 Conclusion

Retailers rank among the leaders when it comes to supply chain management. However, this industry still faces significant stockout rates of 5-10%, which result in sales losses of up to 4%. In Europe, the most significant cause of out-of-stock situations is due to unsatisfactory shelf replenishment. Early detection of low stocks combined with timely replenishment by the store's personnel could significantly reduce out-of-stock situations. In this paper, we presented a pervasive product availability monitoring system for retail shelves that increases the visibility of stocks, anticipates out-of-shelf situations, and informs the personnel when a particular shelf needs to be replenished. This monitoring system is based on weight-sensitive foam which allows inexpensive roll-to-roll based manufacturing. The electrodes of the capacitive sensors are formed by parallel leads, which are printed on 200μm-thick polyolefin foam. Our system differs from existing solutions by integrating low-cost manufacturing techniques into the design, and our cost considerations indicate low system costs of about $9/running meter. Preliminary tests suggest that the system offers sufficient sensitivity of 10g/cm^2 to reliably detect low stocks of products. The error rates of false negatives and false positives are moderate and we account for those errors by increasing the threshold value at which the personnel is informed. Anticipating out-of-shelf situations through continuous automatic monitoring suggests shortening the duration of OOS situations. This will not only reduce losses of sales but also increase customer satisfaction.

References

1. Addlesee, M.D., Jones, A., Livesey, F., Samaria, F.: ORL Active Floor. IEEE Personal Communications, Vol.4, No 5, 1997, pp. 35-41.
2. Ashruf, C.M.A.: Thin Flexible Pressure Sensors, Sensor Review, Vol. 22, No. 2, 2002, pp. 322-327.
3. Cattin, P.: Biometric Authentication System Using Human Gait, Dissertation No. 14603, ETH Zurich, Switzerland, 2002.
4. Goth, G.: RFID: Not Quite Prime Time, But Dwadle at Your Own Risk, IEEE Distributed Systems Online 1541-4922, Vol. 6, No. 2, 2005.
5. Gruen, T.W., Corsten, D.S., Bharadwaj, S.: Retail Out-of-Stocks: A Worldwide Examination of Extent, Causes and Consumer Responses, Report, Grocery Manufacturers of America, The Food Marketing Institute and CIES – The Food Business Forum, Washington, D.C., 2002.
6. Lee, H., Özer, Ö.: Unlocking the Value of RFID. Graduate School of Business, Stanford University, Palo Alto, CA, 2005.
7. Lowe, M., King, A., Lovett, E., Papakostas, T.: Flexible Tactile Sensor Technology: Bringing Haptics to Life, Sensor Review, Vol. 24, No. 1, 2004, pp. 33-36.
8. Lumelsky, V., Shur, M., Wagner, S.: Sensitive Skin, Sensors Journal, Vol. 1, No. 1, 2001.
9. Metzger, C.: Indirect Object-sensing Technology to Prevent Out-of-stock at Retail-level, In Proceedings of the Workshop on Smart Object at the International Conference on Ubiquitous Computing, Tokyo, 2005.
10. Meyer, J., Lukowizc, P., Troester, G. Textile Pressure Sensor for Muscle Activity and Motion Detection, Int. Symposium on Wearable Computers (ISWC), Switzerland, 2006
11. Novel, Product Information "emed AT-25A", http://www.novel.de/productinfo/prices-emed.htm.
12. Orr, R. J., Abowd, G. D.: The Smart Floor: A Mechanism for Natural User Identification
13. and Tracking, Conference on Human Factors in Computing Systems (CHI 2000), ACM/SIGCHI, The Hague, Netherlands, 2000.
14. Paradiso, J, Abler, C., Hsiao, K., Reynolds, M.: The Magic Carpet: Physical Sensing for Immersive Environments, Conference on Human Factors in Computing Systems (CHI'97), Atlanta, GA, 1997.
15. Prosser, S., Schmidt, E.: Smart Sensors for Industrial Applications, Microelectronics Int., Vol. 16, No. 2, 1999, pp. 20-23.
16. Raman, A.: Retail-Data Quality: Evidence, Causes, Costs, and Fixes. Technology in Society Vol. 22, No. 1, 2000, pp. 97-109.
17. Schmidt, A., Van Laerhoven, K., Strohbach, M., Friday, A., Gellersen, H.W. Context Acquistion based on Load Sensing, In: Boriello, G., Holmquist, LE (eds.): Proceedings of Ubicomp 2002, Lecture Notes in Computer Science, Vol. 2498, Springer, Berlin, 2002, pp. 333-351.
18. Sekisui Alveo, Physical Characteristics – data sheet on Alveolit TEE 0300.25, ID-number: MDS-577/B4, September, 2004.

19. Sergio, M., Manaresi, N., Cami, F., Canegallo, R., Canegallo, R., Guerrieri, R.: A Dynamically Reconfigurable Monolithic CMOS Pressure Sensor for Smart Fabric, Journal of Solid-State Circuits, Vol. 28, No. 6, 2003.
20. Tellkamp, C.: The Impact of Auto-ID technology on process performance – RFID in the FMCG supply chain, Dissertation No. 3182, University of St. Gallen, Switzerland, 2005.
21. Wijesiriwardana, R., Mitcham, K., Dias, T.: Fibre-Meshed Transducers Based Real Time Wearable Physiological Information Monitoring System, Proceedings of the Eighth IEEE International Symposium on Wearable Computers (ISWC'04), Arlington, 2004.
22. Woods, J., Piszczalski, M., Davison, J., Steenstrup, K., Vining, J., Rozwell, C., Maoz, M., LeHong, H., Burt, M., Reynolds, M., Jones, N., Mahler, M., Hieb, B.R., Landry, S., Harris, M., White, A., Miklovic, D.: Hype Cycle for Radio Frequency Identification, Report G00129258, Gartner Research, p. 9, 2005.

Assessing and Optimizing the Range of UHF RFID to Enable Real-World Pervasive Computing Applications

Steve Hodges, Alan Thorne[1], Hugo Mallinson[1], and Christian Floerkemeier[2]

Microsoft Research, 7 JJ Thomson Avenue,
Cambridge, CB3 0FB, UK
`shodges@microsoft.com`
[1] Cambridge Auto-ID Lab, Institute for Manufacturing, University of Cambridge
Mill Lane, Cambridge CB2 1RX, UK
`ajt@eng.cam.ac.uk, hfm21@eng.cam.ac.uk`
[2] ETH Zurich, Institute for Pervasive Computing
CH-8092 Zurich, Switzerland
`floerkem@inf.ethz.ch`

Abstract. Radio frequency identification (RFID) may be used to automatically detect, locate and/or identify objects, making it an ideal candidate for many pervasive computing applications. As RFID technology improves in terms of cost and performance, it is increasingly being explored in a variety of applications, ranging from eldercare through to the smart supply chain. However, while passive UHF RFID has many benefits over other RFID variants, reliable operation as the tag moves in the environment is inherently difficult to predict and can represent a significant challenge. In this paper, we present a novel and practical experimental method called attenuation-thresholding which may be used to characterize the operating range of such RFID systems. The results presented demonstrate the advantages of our method over the conventional read-rate approach. We also demonstrate a novel approach to collecting the measurements in range characterization experiments using robotic automation. Finally, we show how the application of attenuation-thresholding in combination with robotic automation can be used to optimize tag placement on an object. In addition to the clear relevance of this work to the many RFID-based pervasive computing applications reported in the literature and currently under development, it also has broad applicability in other RFID application domains. We conclude with a number of ideas for future extensions to this work.

1 Introduction

Radio-frequency identification (RFID) provides a relatively simple and cheap way to associate an electronic identity with a physical object. An RFID tag is attached to the object to be identified, and an associated RFID reader can then detect the presence of that object and determine its identity. This ability to augment objects (and even people) in a relatively light-weight manner and then use RFID readers in the operating environment to detect, identify and to some extent locate them, means that RFID technology is increasingly used in a number of pervasive computing application scenarios.

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 280 – 297, 2007.

One of the most significant applications for RFID is supply chain automation through the replacement of barcode technology. The potential benefits of RFID tagging individual items and logistical units (boxes and pallets) in the consumer goods supply chain are huge [10], because the identity, location and authenticity of those items can be much more easily monitored [14, 30]. In turn, this creates the potential for increased efficiencies and cost savings [6, 19]. Whilst supply chain automation may be the largest single application for RFID, a huge number of other pervasive computing applications are made possible through the technology; examples include smart shelves, desktops and medicine cabinets [31, 8, 32], “Reminder Services” [4] and location monitoring of elders [25].

While passive RFID technology has many benefits over other technologies that might be used to detect, locate and/or identify objects [12, 15], reliability of operation as the tag moves in the environment is inherently difficult to predict and can represent a significant challenge. Although there are techniques for predicting the range of an RFID system which would in theory help designers to build systems with predictable operating performance, many of these are limited in practice. Instead, a typical UHF RFID system is designed very conservatively so that during use the tag will always be close enough to the RFID reader to ensure consistent operation.

In this paper, we present a novel and practical experimental method called attenuation-thresholding which may be used to characterize the operating range of UHF RFID systems. The results presented demonstrate the advantages of our method over the conventional read-rate approach. We also demonstrate a novel approach for the automation of measurements in range characterization. Finally, we show how the application of attenuation-thresholding in combination with robotic automation can be used to optimize tag placement on an object. In addition to the clear relevance of this work to the many RFID-based pervasive computing applications reported in the literature and currently under development, it also has broad applicability in other RFID application domains.

2 The Difficulties of Deploying UHF RFID Systems

In this paper we define the ‘range’ of an RFID system as being the volume of space in front of an RFID reader antenna in which an RFID tag can be reliably detected and identified. We do not want to formally define ‘reliable’, but we understand it to mean that a typical pervasive computing application will operate as envisaged by the designer whenever the tag(s) in the system are within ‘range of the RFID reader’. More specifically, we actually refer to the range of the reader *antenna* – the antenna may be built into the same enclosure as the reader electronics, but with the UHF systems under consideration here, is often a separate unit connected to the reader with a cable.

The difficulties associated with determining the range of an RFID system stem from the complex nature of the RF field generated by the reader antenna. This field must power the RFID tag in order for the system to operate, and this is only possible if it is sufficiently powerful at a given location and if enough of the power available is transferred to the tag. Once the tag is powered, it must then transmit a signal back to the

RFID reader. It has been demonstrated that the limiting factor in this two-step process is typically delivering power to the tag [26], i.e. if the tag can be powered then the return path will be largely error-free. So modeling or measuring the strength of the RF field generated by an RFID reader is in essence the key to understanding its range.

At the simplest level, the strength of an RF field drops off as you move away from the antenna that is generating it according to the Friis Transmission Equation:

$$\frac{P_{tag}}{P_{reader}} = G_{tag} G_{reader} \left(\frac{\lambda}{4\pi R}\right)^2 \tag{1}$$

This says that the ratio of power received by the tag antenna (P_{tag}) to power input to the reader antenna (P_{reader}) is proportional to the gain of both the tag and reader antennas, (G_{tag} and G_{reader} respectively) and to the square of the wavelength (λ), and inversely proportional to the square of the distance between reader and tag (R). If you double the distance between reader and tag, the power available becomes one quarter of its previous level. However, the above equation applies only under ideal conditions. Whilst these ideal conditions can be artificially created in a purpose-built anechoic chamber[1], in reality a number of factors combine to change the actual field strength in a complex, non-intuitive, and hard to predict manner. These factors are:

Absorption: Any material between tag and reader will reduce the power available to the tag; the amount of degradation depends on the amount and nature of that material.

Multipath fading: Even if there is line-of-sight between reader antenna and tag, so-called fading effects can sometimes decrease, and sometimes increase, the read range. Fading is caused by interference between two or more versions of the transmitted signal, which travel along multiple (different) paths and combine at the receiver to result in a signal with widely varying strength. A localized area of particularly low signal strength within a region of generally higher strength is called a *null*; and in this paper we refer to the opposite condition as an *outlier*.

Polarization losses: The ability to power a tag is further significantly reduced by polarization losses, which occur when the RF energy from the reader is not polarized in the optimal orientation for the tag.

Impedance mismatch: Similarly, any impedance mismatch at the tag antenna will reduce the power available. Typically tags are designed to be impedance matched when operating in free space, and the proximity of the object to be tagged (no matter what it is made from) will have a de-tuning effect. Different materials will exhibit different effects.

As will be appreciated, the nature of the object to be tagged can have a dramatic effect on the range of a UHF RFID system – not only when that object comes between the tag and the reader, but also when the tag is in line-of-sight of the reader antenna, due to antenna de-tuning. Similarly, the orientation of the tag may change its ability to pick up RF energy at a certain location. However, the most insidious effect is multipath fading, where other objects in the environment, possibly items many metres away from both the tag and the reader, cause reflected signals that conspire to create

---

[1] This is a special isolated environment, typically the size of a small room, which is designed to be free from sources of interference in order to simplify the analysis of RF devices.

nulls and outliers, resulting in very unexpected results. This is particularly relevant for pervasive computing applications, where the environment is often uncontrolled.

It is possible to extend the Friis Transmission Equation to account for the additional factors listed above, and this approach is sometimes used to address specific issues (such as the de-tuning effects of the object to be tagged). However, it is often impractical to do this due to the complexity of the resulting model, the complexity of the data needed for the calculation, and the complexity of the calculation itself.

The simplest approach to deploying a reliable RFID system is to be conservative in the system design. However, it is nearly always desirable to maximize operating range because that usually leads to a less constrained user experience. This is especially true when using the more sophisticated tags that are being developed for some pervasive computing applications (such as [29]), due to the increased tag power consumption which acts to reduce range. Unfortunately, not many tools to accurately characterize or improve the range of an RFID system are available. Typically, the system designer will simply try out a specific configuration, and if it appears to work reliably, tweak it for optimal performance. A better approach to understanding and optimizing the range of RFID systems would be very valuable.

3 Previous Work Related to Assessing RFID Range

3.1 Building More Sophisticated Models

The obvious way to improve on the basic model presented in the previous section (Equation 1) is to analyze one or more additional factors from first principles and extend the model appropriately. One example is the use of a technique known as radar cross section (RCS) analysis to examine the performance of the RFID tag antenna [18, 24]. An RCS model can not only model the ability of a particular antenna design to communicate back to the reader, but may also account for impedance matching effects [24]. However, these factors are not typically dominant in determining the range of an RFID system.

Another common approach to extending the basic Friis Transmission Model is to run a number of experiments to characterize certain factors that influence RFID range and then incorporate additional terms in the model to reflect the effects that have been observed. For example, the reader and tag antennas will rarely have constant gain in all directions (as assumed in the Equation 1). Instead, they will frequently be *directional*, and the easiest way to characterize this is by measurement in an RF anechoic chamber. The resulting data (see Figure 2(b) later in the paper for an example) may be used to extend the Friis Equation. Since there are no external reflections or noise sources during the evaluation of the antenna, the readings can be assumed to be free from errors [17].

Another extension based on data collected in a series of experiments has been proposed in a recent study [13]. Here, the RF energy received by an RFID tag antenna in free space was measured, and this experiment was then repeated with the antenna attached to a number of different materials. For ease of testing, this was done using RF test equipment, rather than an RFID reader and tag. The authors not only present the measured data, which shows to what extent different materials reduce a tag's ability to receive RF energy, but they also demonstrate how their experimental results may be generalized to give different 'gain penalty' figures for those different materials. In this way, they suggest that the effect of tagging objects made from these materials may be predicted without the need for further measurements. However, it would appear that the experimental approach reported does not (sufficiently) model the de-tuning effect because the authors used a tunable impedance transformation network to reduce the effect of impedance mismatch. Also, the technique require that new experiments are run for additional materials (not previously analysed) or for complex objects. This cannot be done in the field at the time of RFID deployment without access to suitable RF test equipment.

More sophisticated models of RFID range are not reported in the literature, presumably due to the difficulties of generating and using them. Even the simple models presented above are largely in their infancy and have not been widely adopted.

3.2 Measuring the Strength of the RF Field

Since the delivery of power from the reader antenna to the tag is typically the limiting factor in an RFID system [26], the strength of the RF field itself is a useful metric for predicting range. This may be measured relatively simply using a field probe in conjunction with the appropriate RF test equipment. It also has the advantage that it becomes straightforward to introduce objects into the environment, such as the object that the tag is attached to, and to monitor the associated effect on field strength and therefore range. However, there are also a number of drawbacks. It is complicated to set up: the field probes may not have the same performance as the tags themselves; and ultimately only part of the RFID system is being evaluated.

A number of pieces of previous experimental work to measure RF field strength are reported in the literature. In [22] field probes are used to sense the RF energy available to an RFID tag when the tag and field probe are attached to a case of medical ampoules. To overcome inaccuracies due to the difference between the field probe antenna design and the RFID tag antenna, anechoic chamber antenna characterization is carried out. However, some difference between the performance of the probe and that of the tag will always remain. A different approach is used in [28] – here the field probe is essentially incorporated into a custom-built high performance RFID ‘tag’ which communicates to the RFID reader in the same way as a standard tag, but records RF field strength at the same time. This technique again introduces inaccuracies through the use of a custom antenna design which will differ from the chosen tag antenna.

3.3 Measuring the Extent of the Reader Field Using Read-Rate

The techniques reviewed so far are only of practical use for evaluating performance of *parts* of the RFID system. An ideal measure would take all factors affecting range in a real operating environment into account. There is actually one very simple way to assess RFID operating range, using a measure known as *read-rate* – the rate at which a reader can identify a tag [1, 16]. The essential idea is that the reader is put in a mode where it continually scans for the presence of an RFID tag. If a tag is detected, its identity is determined and recorded by the reader, whereupon the scan operation immediately repeats. Any error during the detection and identification stages will result in failure for that particular scan operation. When the tag is sufficiently powered, the number of errors will be (very close to) zero. However, as the tag-reader separation is increased, the chances of a communications error increase, and scan cycles will start to fail. Eventually the number of errors becomes so great that the reader will consistently fail to identify the tag.

The exact definition of ‘read-rate’ varies; typical approaches are to measure the number of successful reads per second or the ratio of successful cycles to the total number of cycles. The latter is more formally defined by:

$$r = N_r / N \tag{2}$$

where N represents the number of cycles, N_r is the number of successful cycles, and r is the read-rate (a dimensionless scalar between the values of 0 and 1) [16].

The advantage of read-rate is that it is very easy to measure – it is natively supported by (nearly) all RFID readers[2]. As a result, read-rate is by far the most widespread method used in practice in the authors’ experience, across a wide variety of different application areas. It may also be used with tagged objects, and hence provides a straightforward way to incorporate any effects introduced by the object, which may of course be very significant. Indeed, [16] concludes that read-rate is a useful metric for determining reader-to-tag distance. However, as we show later in this paper, read-rate does not always correlate with distance between the reader and the tag as might be expected, and can therefore sometimes be quite misleading.

4 Evaluating Existing Approaches to RFID Range Characterization

4.1 Basic Theoretical Modeling for RFID Range Characterization

As a baseline for the experiments which follow in this paper, we used the Friis Transmission Equation in conjunction with a model of the reader antenna based on its measured radiation pattern to calculate the strength of the field generated by the reader. By making an assumption about the level of power required to operate a tag in the field, it is possible to predict the extent of the operating range across a simulated 3D space. Figure 4 (which occurs later in the paper to facilitate comparison with other figures) depicts this simulated data, based on the radiation pattern of a Cushcraft S9028PC antenna (see Figure 2 and [7]) and a fixed operating frequency of 915MHz.[3] Note the position of the reader antenna, also shown in the plot.

4.2 Experiment to Characterize RFID Range Using Read-Rate

As described in Section 3.3, read-rate has been used as an indicator of field-strength and/or range of an RFID system. However, we have not seen a systematic measurement of read-rate across the entire operating space of a UHF RFID system, which would result in a plot similar to the one in Figure 4. Perhaps one reason for this is simply the scale of such an experiment – previous experiments which have characterized RFID

---

[2] Generating read-rate data is very straightforward with all UHF readers that the authors have experience with.

[3] In practice the RFID reader will actually frequency-hop across a number of channels in the 902-928MHz band; we simplified calculations by approximating to the centre frequency.

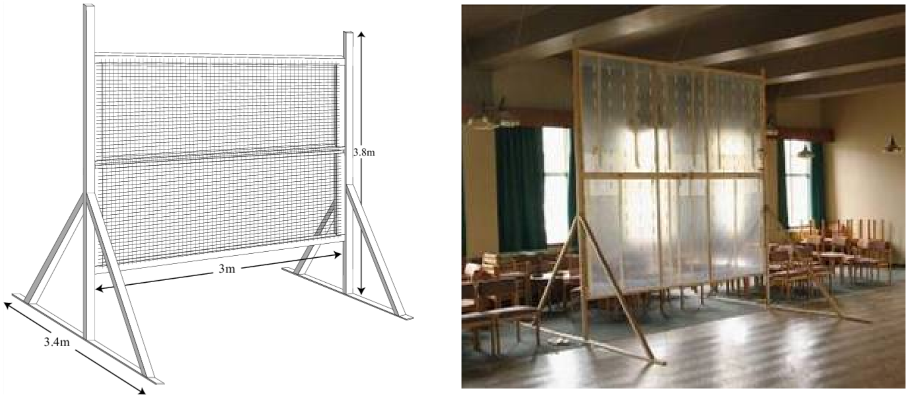


Fig. 1. The frame which supported the grid of tags was constructed of wood, with plastic overlaid for rigidity, and a sheet of thin plastic to which the tags were attached. The wooden chairs behind the frame in the photo were assumed to have negligible effect on RFID operation.

read fields [1, 9, 13, 16] have tested one tag position at a time. In particular, [1] used a grid marked out on the floor at set intervals, and a height-adjustable stand to which the tag was attached. The stand is manually placed at a certain location on the grid, a reading is taken; the stand is then moved and the next reading is taken. This has to be repeated for all of the grid locations and for all the tag heights under consideration, and is therefore considerably time-consuming. To minimize effects that may be introduced by the presence of a human body, it is important that the area under test is vacated by the operator(s) for each test, further increasing the testing time.

In the experiment reported here, read-rate was recorded for every point on a 20cm grid throughout a 6m x 3m x 3m volume in front of the reader antenna, for a total of 7,932 points. The measurements were carried out in a large, open room on the top floor of a two-storey building of steel-reinforced concrete construction. In order to reduce the time required to characterize the entire field, a 14x14 array of tags which could be tested in one operation was used. This array was supported by a 3m x 3m wooden frame with semi-rigid plastic netting attached to it to provide stability and with thin plastic sheeting stretched over the front to form a smooth surface for attaching the tags. The frame is depicted in Figure 1. No metal was used in its construction to minimize any effect of the frame on the RF field. All tags were at least 10cm from any of the wood in the frame, to minimize adverse de-tuning effects the proximity of wood might have on them.

An ALR-9750 915MHz RFID reader and a circularly polarized antenna supplied by Alien Technology were used. This reader runs the EPC class 1, generation 1 air interface protocol [2]. The circularly-polarized antenna, which is used for transmission and reception, results in consistent performance independent of the orientation of the tag (with respect to the reader antenna) in the plane parallel to the reader antenna, and so is a popular choice for many application scenarios[4]. Although an Alien part number or

---

[4] Despite the flexibility afforded by the circularly polarized antenna, all the data presented in this report was collected with the tag in a consistent orientation, namely vertical.

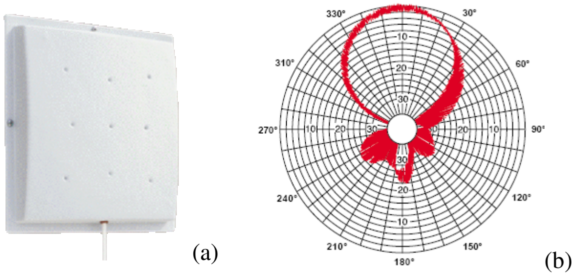


Fig. 2. Cushcraft S9028PC circularly polarized antenna. (a) photo and (b) H-plane radiation pattern. This antenna has a +7.5dBic gain, a VSWR of 1.5:1 and a -3dB cut-off at 65° for both H-plane and E-plane.

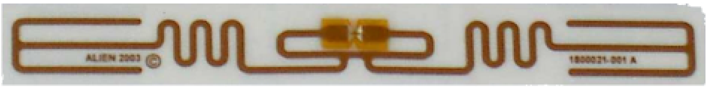


Fig. 3. Alien ALL-9340 ‘squiggle’ tag

datasheet was not available for the antenna, it believed to be very similar to the Cushcraft S9028PC antenna [7] depicted in Figure 2. The tags were Alien ALL-9340, 98.2 x 12.3 mm ‘squiggle’ tags, using their Omega revision silicon (Figure 3). Since the experiments were carried out in Europe before the availability of UHF RFID readers operating in the 868MHz band, a 915MHz test license was obtained.

The results of this extensive experiment are shown in Figure 5.

4.3 Discussion

The two approaches to characterizing the range of an RFID system presented in this section give quite different results. The Friis Transmission model leads to a well-defined prediction of read range – the sort of envelope that is often assumed in the literature. Of course, effects such as multipath fading, impedance mismatching and polarization changes are not taken into consideration in the model. On the other hand, the read-rate data naturally incorporates all these various factors, because it is collected from an operational RFID system. In particular, nulls and outliers are clearly present, presumably due to constructive and destructive interference from reflections; these are accentuated by the non-linear nature of read-rate data [26]. Whilst this data must be an accurate indication of which tags could and could not be read at different locations in the environment during the course of the experiment, it is not a realistic indicator of the extent of the operating range. Experience with real RFID deployments shows that outliers such as those shown in Fig. 5 are hard to replicate reliably and can in no way be predicted or relied upon. Similarly, whilst there may be nulls close to the reader antenna, small relative movements between the tag and the reader antenna (as would be expected in most real-world applications) will likely eliminate these.

We conclude that neither of the approaches outlined here are good candidates for predicting RFID range, for quite different reasons. It is possible that an approach based on field probes would be more successful, but this would have disadvantages such as the need for specialized test equipment and/or instrumented tags. The next section introduces a new technique developed to overcome weakness of read-rate testing but without requiring expensive test equipment or instrumented tags.

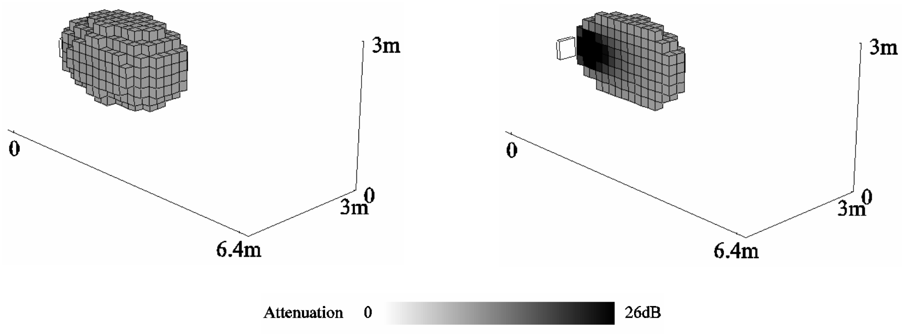


Fig. 4. The RF field strength as predicted using the Friis Transmission Equation. The cut-away view clearly shows the field strength variation and depicts the reader antenna location.

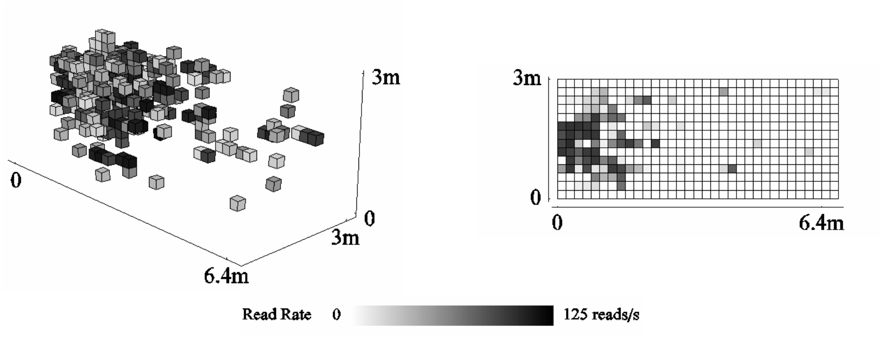


Fig. 5. Read-rate data, 3D view and a 2D view of a horizontal slice through the data

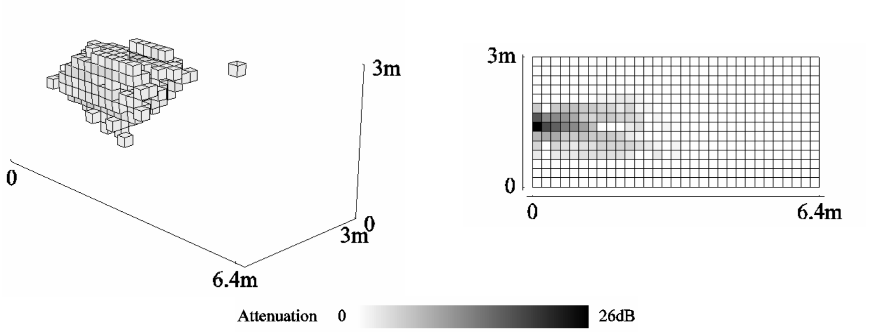


Fig. 6. Attenuator-thresholding gives a much more realistic indication of operating range

5 Introducing Attenuation-Thresholding: A New Approach to RFID Range Measurement

5.1 Experimental Setup for Attenuator-Thresholding

We have developed a new approach for measuring the performance of an RFID system, which we believe is nearly as easy to use as the read-rate metric, but is a much better indicator of RFID system range. We use a programmable attenuator to selectively degrade the signal that passes between a standard RFID reader and its antenna. By increasing the attenuation automatically (under computer control) until the tag read-rate drops below a chosen threshold, it is possible to determine and record the RF margin (i.e. how much power is available in excess of the minimum required to operate the tag) for each location tested. We call this technique *attenuation-thresholding*.

For the attenuation-thresholding experiments reported here, we used the same RFID equipment as outlined in Section 4.2, in conjunction with a Pasternack PE7011-6A programmable attenuator [23]. This was connected between the RFID reader's single antenna port and the reader antenna. It has a DC to 1GHz operating range, a maximum VSWR of 1.4:1, an insertion loss of 2dB and can be digitally controlled to insert up to 63dB in 1dB steps (±0.3dB). The power rating of the attenuator is 0.5W average, 50W peak, which is compatible with the EPC Class 1 Generation 1 specifications [2]. The total level of attenuation introduced is controlled via six separate stages which are connected in series inside the unit. There are six inputs to the device; applying 12V to an input enables the corresponding attenuation stage. A microcontroller-based custom interface was constructed to allow the attenuator to be controlled from a PC via a serial port. We used a threshold read-rate of zero for these experiments.

5.2 Results

We ran an antenna-thresholding experiment using the tag array of Section 4.2 in order to compare the technique directly with the read-rate analysis presented in the previous section. The results are shown in Figure 6. As can be seen, even though it intrinsically captures factors such as multipath fading and polarization losses, the data is much less noisy than for the read-rate testing.

5.3 Discussion

The comparison of the read-rate measurements with the attenuation-thresholding approach shows that the data captured with the latter technique is much less noisy, even though it intrinsically captures factors such as multipath fading and polarization losses. We believe that this is at least partly caused by the fact that attenuator introduces an additional loss of 2dB. This will simply remove some of the outliers due to the overall reduction in field strength. However, much more importantly, attenuation-thresholding much more effectively shows the reduction in power margin as the tag is moved away from the antenna. The read-rate approach is not suitable for determining the power distribution in the vicinity of the reader, since read-rate vs. power at the tag is a highly non-linear function [26]. The true power distribution inherently measured with attenuation-thresholding is essential for predicting the operating range under different conditions, e.g., in the presence of interference from other transmitters, or tag detuning. Attenuator-thresholding is thus an easy-to-use technique, which requires little additional hardware beyond a standard RFID system set-up and yet produces much more useful information about the operating range of an RFID system than read-rate analysis, which is commonly used today.

5.4 Limitations of Evaluating a Large Workspace

The experiments presented so far use an array of 196 tags to map out the size and shape of the usable workspace for an RFID system. However, there are some limitations to this approach and we have subsequently developed an experimental technique which completely automates RFID system range testing using robotic automation. An RFID tag is attached to the end-effector of a robot using a non-conductive mount (to reduce the influence of the robot itself on tag operation) and the robot is controlled so as to move that tag sequentially between each of the positions to be evaluated. At each position, a read-rate or attenuation-thresholding test may be performed. In this way, the entire range characterization may be carried out without any manual intervention no matter how many different test positions are required. By using the same tag at each location, there are no discrepancies due to variations between tags (which is likely the case with the array of tags used in Sections 4.2 and 5.1). Since the robot can move at speeds of over 1ms^{-1}, the method is fast even though data is collected one point at a time. The programmed locations for the end-effector of the robot (and hence the tag under test) will be very accurate, typically within 1mm.

In order to evaluate this technique, we attached a tag to the end-effector of a Fanuc M6i industrial robot [11]. This is an anthropomorphic style robot arm with ±0.1mm end-effector positioning repeatability and 6kg lifting capacity, designed for general purpose factory automation tasks such as arc welding and loading/unloading parts. Under computer control, we instructed the robot controller to move to each test position in turn, running a read-rate test (in this case) at each location. A fixed 10dB attenuator was inserted inline between the reader and the reader antenna in order to artificially limit the range of the RFID system without altering the shape of the field. This technique, which has been previously reported in the literature, limits the extent of the field to within the reachable workspace of the M6i robot. Figure 7 shows a dataset gathered using this technique.

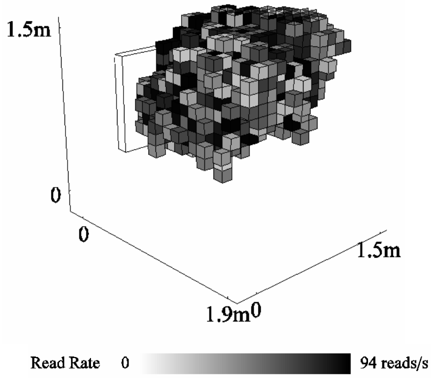


Fig. 7. Read-rate testing using robotic automation to move a tag sequentially between each test position. The reader antenna location is shown at the back of the plot.

Robotic range testing overcomes another major limitation common to much RFID system analysis, namely the labour-intensive and time-consuming nature of testing. The RFID evaluation system based on the use of robotic automation presented here alleviates these issues, and therefore provides a versatile platform for extensive characterization of RFID operation in new ways that have previously been impractical. In the next section, we demonstrate the use of attenuation-thresholding in combination with robotic automation to find the best place to put a tag on an object.

6 Using Robotic Attenuation-Thresholding to Determine Where to Tag an Object

6.1 The Importance of the Object to Be Tagged

In the analyses presented so far, the range of an RFID system has been characterized for a tag in free space i.e. not affixed to an object. Of course, in real applications the tag will be attached to an object of some description – and the introduction of such an object will in many cases have a significant impact on the performance of the RFID system. Whilst the theoretical model of the reader field strength cannot be trivially extended to incorporate all the effects the object to be tagged may introduce, the experimental techniques reported in this paper do support this obvious enhancement.

However, before assessing the usable workspace over which a tagged object may be reliably detected, it is important to choose the best location for the tag on that object. If the object is of non-uniform composition, the location of the tag on the object may be significant – some locations would lead to much more degradation in performance than others. Due to the complex nature of the interaction between the tag and the object, predicting the best location is not straightforward. In this section we present a new experimental technique to methodically determine the optimal tag position.

6.2 Assessing Different Tag Locations on an Object Using Tag Mapping

In order to support the choice of tag location, we have extended the robotic automation technique described in Section 5.4 to allow the position of the tag on the object to be varied automatically. In this case, the tag is not actually stuck down to the object as would normally be the case, but is instead held against the surface of the object using two thin, flexible plastic fingers, see Figure 8 for details. Note that this tag positioning arrangement has no measurable effect on the RFID system performance. The tag is held flat against the face of the product, even along curved faces or around edges. We call this technique *tag mapping*.

To demonstrate the ability to measure the performance of different tag positions on an object, we applied this technique to a case containing six bottles of wine, which was purchased from a local supermarket. This object was chosen due to interest in tag mapping from colleagues working on pervasive computing supply chain applications.

Fig. 8. The custom tag holder allows the RFID tag to be pressed against an object as if it were actually stuck down. Two compliant 'fingers' are made from loops of 9x.75mm polyethylene (to the left and right of the tag) and are attached to the robot end-effector using a fairly RF-transparent material. An RFID reader antenna can be seen in the background.

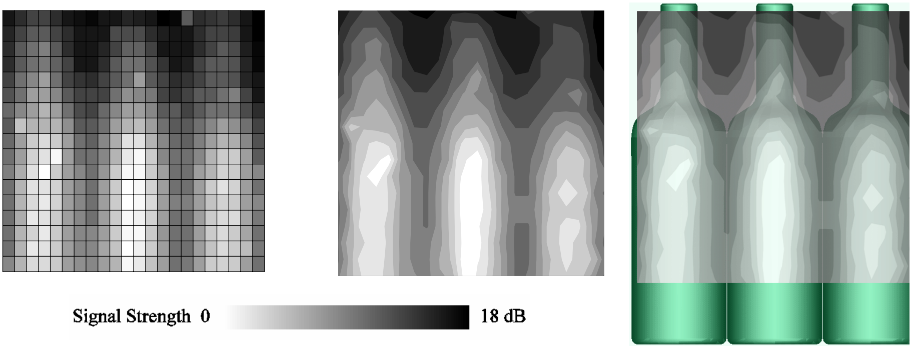


Fig. 9. The results of testing one face of the case of wine. The raw 'pixel' data is shown (left) along with averaged 'contour' data (centre) which more clearly shows how the shape of the wine bottles effects the performance of the tag. The same contour plot is superimposed on a 3D model of the wine bottles to demonstrate the close correlation.

The detrimental effect of proximity to water on the operation of an RFID tag is well established. We would therefore expect the tag to perform badly when placed towards the bottom of the case, where it is in closest proximity to the wine bottles, and to perform much better right at the top of the case, where the bottles narrow and there is much more free space inside the case. The generated tag map should reflect this.

The M6i robot was used, but in this experiment the tag was fixed in space and the case of wine was moved relative to it in order to simulate the different possible tag positions. A total of 374 candidate locations were evaluated across a single face of the case. The tag orientation was fixed throughout the experiments.

6.3 Tag Mapping Results

The tag map produced from the case of wine is shown in Figure 9. The raw 'pixel' data shown on the left represents the degradation introduced by the presence of the object when the tag is placed so that its centre is positioned at each pixel in turn.

6.4 Comparing the Performance of Different Tag Locations as an Object Moves Away from the RFID Reader Antenna

The final experimental results presented here show how robotic attenuation-thresholding can be used to determine any changes in the tag placement map as the tagged object is moved relative to the RFID reader antenna. For this type of experiment, a second robot is required – we used a Fanuc M16/iT robot arm mounted on a gantry, allowing robot arm itself to move through 11.2m (along the length of the gantry). Instead of moving the object to different locations in front of the reader antenna, we actually attached the reader antenna to the M16 end-effector. The M6i and M16i/T robots are shown in Figure 10.

Fig. 10. Fanuc M6i and M16i/T robots (left and right respectively). The M16i/T used in this work had a 12m gantry that allowed 11.2m of movement. The M6i was fixed at a height of 1m from the ground.

Modified control software was used to synchronize the movement of the two robots so that for each candidate object location, a separate tag placement map was measured. For this experiment, a different (although similar) object was tested, namely a case of twelve 50cl bottles of Vittel spring water. Figure 11 shows an external view of the case, as mounted on the robot end-effector, and a photograph of the contents of the case (as viewed from above).

6.5 Results of Tag Map Comparison Experiments

Two particular scenarios are reported here. The first of these investigates how the tag placement map varies as the tagged object moves away from the reader antenna, by

Fig. 11. A photo of one of the 50cl Vittel bottles (left), the inside of the box of Vittel under test, and robot end-effector mounting details for the box

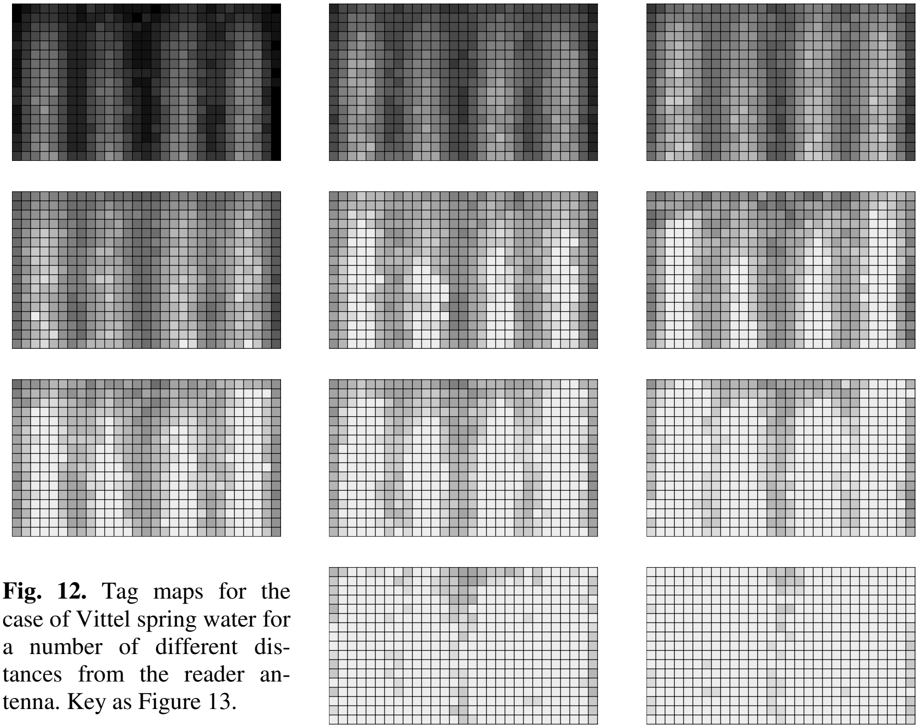


Fig. 12. Tag maps for the case of Vittel spring water for a number of different distances from the reader antenna. Key as Figure 13.

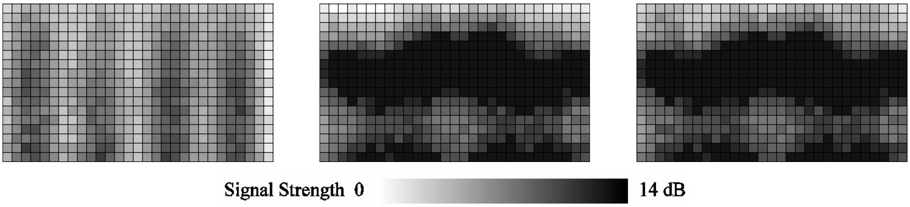


Fig. 13. The tag maps for the case of Vittel spring water when the tagged surface is facing towards the reader (left) and away from the reader (centre). Also shown is the 'worst-case' combination of these two tag maps (right), which demonstrates the minimum performance that could be expected if either case orientation is possible (and not known in advance).

recording a tag map at a series of distances from the reader antenna, each 100mm more than the previous. The results of this test are shown in Figure 12.

In many pervasive computing applications it is not possible to constrain a tagged object so that the tag is always facing the RFID reader antenna. The second test investigates what happens if the tagged object is rotated through 180° so that the tagged surface is facing *away* from the reader antenna, usually the worst-case scenario in terms of RFID operating range. As Figure 13 shows, the map of good and bad tag locations is very different when the tag is on the reverse side of the case. If the orientation of the case with respect to the reader is not known in advance, i.e. if the tag could be facing the reader or on the opposite side, then the map depicted on the right of Figure 13, which takes the worst-case elements of the two different conditions, should be used.

7 Conclusions and Future Work

We have presented two novel and practical experimental methods, namely attenuation-thresholding and the use of robotic automation. These may be used to assess and optimize the operating range of UHF RFID systems much more effectively than conventional approaches. Although we used custom-built hardware, we observe that the ability to programmatically control output power is increasingly being incorporated in off-the-shelf RFID readers. As a result, in many instances it is now possible to implement attenuation-thresholding using a simple script to control a reader remotely. We strongly encourage those deploying RFID in real-world applications to use this approach instead of using read-rate. The use of a robot to automate testing means that the effort required to measure RFID operating range is significantly reduced and the precision of the data collected is increased. Whilst the robot-assisted approach will be inaccessible to many researchers deploying RFID-based pervasive computing systems, we hope in time to characterize the performance of a range of different types of object, creating a kind of reference database which would be of direct relevance to practitioners. It may also be possible to create a third-party service of some kind (such as a test centre) to evaluate specific scenarios.

Additionally, our work on finding the best place to put a tag on a given object can deliver a significant improvement in RFID performance by systematically evaluating every possible tag position. This technique could be readily extended to compare the performance of different RFID readers or different tags, and in particular of different tag antenna designs. Variation in tag orientation could also be considered.

The limitations to our work are highlighted in the discussion section. While we can accurately assess the operating range of an RFID system in a given environment, it will not always be appropriate to assume the same operating range will apply in a different environment, due to the small-scale fading effects observed in the UHF frequency band. However, the data captured across different environments might help to parameterize statistical fading models, such as Rayleigh and Rician fading [27]. In this way, it would be possible to build extensions to the Friis Transmission Equation in order to model certain common scenarios faced during the design of RFID-based pervasive computing applications. We are currently actively working on this. We are also interested in the possibility of extending work in the literature (such as [20]) to build more accurate location systems built on RFID with attenuation-thresholding.

References

1. Alien Technology, "RFID training course materials," 2003.
2. Auto-ID Center, "860MHz-930MHz Class 1 Radio Frequency (RF) Identification tag radio frequency & logical communication interface specification", EPCGlobal Inc, http://www.epcglobalinc.com/standards_technology/Secure/v1.0/UHF-class1.pdf, 2002.

3. T. Balch, A. Feldman, and W. P. Wilson, “Assessment of an RFID system for animal tracking,” Tech. Rep. GIT-CC-04-10, Georgia Institute of Technology, 2004.
4. G. Borriello, W. Brunette, M. Hall, C. Hartung and C. Tangney, Reminding about Tagged Objects using Passive RFIDs, Ubicomp 2004.
5. L. V. Blake, Radar Range-Performance Analysis. Lexington Books, 1980.
6. G. Chappell, D. Durdan, G. Gilbert, L. Ginsberg, J. Smith and J. Tobolski, Auto-ID on Delivery: The Value of Auto-ID Technology in the Retail Supply Chain, Auto-ID Center whitepaper, ACN-AUTOID-BC-004, 2002.
7. Cushcraft, “S9028PC12NF product datasheet,” Cushcraft, Inc, http://www.cushcraft.com/comm/support/pdf/S9028PC12NF.pdf
8. C. Decker, U. Kubach and M. Beigl, Revealing the Retail Black Box by Interaction Sensing, Proceedings of the ICDCS 2003, Providence, Rhode Island, 2003.
9. D. M. Dobkin and S. W. Weigand, “Environmental Effects on RFID Tag Antennas”, Microwave Symposium Digest, 2005 IEEE MTT-S International Microwave Symposium, 2005.
10. The Economist, “The best thing since the barcode”, February 2003.
11. Fanuc Robotics, Fanuc M6 and M16 series product overview, http://www.fanuc.co.jp/ja/product/catalog/pdf/m6ib-je03.pdf
12. K. Finkenzeller, RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification. John Wiley & Sons, May 2003.
13. J. D. Griffin, G. D. Durgin, A. Haldi and B. Kippelen, “RF Tag Antenna Performance on Various Materials Using Radio Link Budgets”, IEEE Antennas and Wireless Propagation Letters, Volume 5, 2006.
14. J. Hightower and G. Borriello, Location systems for ubiquitous computing, IEEE Computer volume 34, 2001.
15. S. Hodges and D. McFarlane, “RFID: The concept and the impact”, pp 55-75, “The Security Economy”, Organisation for Economic Cooperation and Development, 2004. ISBN 92-64-10772-X. Available at http://www1.oecd.org/publications/e-book/0304031E.PDF
16. B. Jiang, K. P. Fishkin, S. Roy and M. Philipose. “Unobtrusive Long-Range Detection of Passive RFID Tag Motion”, IEEE Transactions on Instrumentation and Measurement, Vol. 55, No. 1, Ferurary 2006
17. J. D. Kraus, Antennas. McGraw-Hill Education, second edition, May 1988.
18. H. Kwon and B. Lee, “Evaluation of RFID Tag Antenna Performance Using Radar Cross Sections”, The European Conference on Wireless Technology, pp 491-493, 2005.
19. H. Lee, B. Peleg, P. Rajwat, S. Sarma and B. Subirana, Assessing the Value of RFID Technology and the EPC Standard for Manufacturers, EPCGlobal Whitepaper, 2005.
20. X. Liu, M. Corner and P. Shenoy, Ferret: RFID Localization for Pervasive Multimedia, Proceedings of the 8th International Conference on Ubiquitous Computing, September 2006.
21. H. Mallinson, S. Hodges and A. Thorne. Determining a Better Metric for RFID Performance in Environments with Varying Noise Levels, 12th IEEE International Conference on Methods and Models in Automation and Robotics, 2006.
22. J. Mitsugi and H. Hada, “Experimental Study on UHF passive RFID Readability Degradation”, Proceedings of the International Symposium on Applicatoins and the Internet Workshops, 2005
23. P7011-6A programmable attenuator DC to 1,000MHz datasheet, Pasternack Enterprises LLC, http://www.pasternack.com/Pdf/PE7011-6A.pdf

24. K. Penttilä, M. Keskilammi, L. Sydänheimo and M. Kivikoski, “Radar cross-section analysis for passive RFID systems”, IEE Proceedings of Microwave Antenna and Propagation, Volume 153, Number 1, February 2006.
25. M. Philipose, K.P. Fishkin, M. Perkowitz, D.J. Patterson, D. Fox, H. Kautz, D. Hahnel, “Inferring activities from interactions with objects” IEEE Pervasive Computing, Volume 3, Issue 4, Oct-Dec 2004 pp. 50 – 57.
26. K. M. Ramakrishnan and D. D. Deavours. “Performance Benchmarks for Passive UHF RFID Tags”, Proceedings of the 13th GI/ITG Conference on Measurement, Modeling, and Evaluation of Computer and Communication Systems, 2006, pp. 137-154.
27. T. S. Rappaport, “Wireless communications”, 2nd edition, Prentice Hall, 2002.
28. R. Redemske and R. Fletcher, “Design of UHF RFID Emulators with Applications to RFID Testing and Data Transport”, Proceedings of 4th IEEE Conference on Automatic Identification Technologies, October 2005.
29. J. R. Smith, A. Sample, P. Powledge, A. Mamishev, S. Roy. “A wirelessly powered platform for sensing and computation” Proceedings of Ubicomp 2006: 8th International Conference on Ubiquitous Computing. Orange Country, CA, USA, September 17-21 2006, pp. 495-506.
30. T. Staake, F. Thiesse and E. Fleisch, Extending the EPC network: the potential of RFID in anti-counterfeiting, SAC '05: Proceedings of the 2005 ACM Symposium on Applied Computing, pp 1607-1612, 2005.
31. D. Wan, Magic Medicine Cabinet: A Situated Portal for Consumer Healthcare, Proceedings of the International Symposium on Handheld and Ubiquitous Computing, Karlsruhe, Germany, 1999.
32. R. Want, K.O. Fishkin, A. Gujar and B.L. Harrison, Bridging Physical and Virtual Worlds with Electronic Tags, Proc. of ACM SIGCHI, pp 370-377, 1999.

Movement-Based Group Awareness with Wireless Sensor Networks

Raluca Marin-Perianu, Mihai Marin-Perianu, Paul Havinga, and Hans Scholten

University of Twente, Enschede, The Netherlands
{r.s.marinperianu, m.marinperianu, p.j.m.havinga, j.scholten}@utwente.nl

Abstract. We propose a method through which dynamic sensor nodes determine that they move together by communicating and correlating their movement information. We describe two possible solutions, one using inexpensive tilt switches, and another one using low-cost MEMS accelerometers. We implement a fast, incremental correlation algorithm, which can run on resource constrained devices. The tests with the implementation on real sensor nodes show that the method distinguishes between joint and separate movements. In addition, we analyse the scalability from four different perspectives: communication, energy, memory and execution speed. The solution using tilt switches proves to be simpler, cheaper and more energy efficient, while the accelerometer-based solution is more accurate and more robust to sensor alignment problems.

1 Introduction

Emerging applications of wireless sensor networks (WSNs) demand an increasing degree of *dynamics*. The sensor nodes are expected to take decisions autonomously, by using *context-aware reasoning*, and to provide an overall solution that is more reliable, accurate and responsive than traditional approaches. Examples of recent application domains include industrial processes, transport and logistics, user guidance in emergency situations [10]. In all these scenarios, we notice a growing interest in having many small, cheap devices that self-organize and cooperate, in order to supervise and actively support the actual processes. The challenges shift, accordingly, from small-scale user-to-device interaction towards large-scale device-with-device *collaboration*. In parallel, the design choices migrate from complex, centralized approaches towards simple, distributed techniques, that can be implemented on resource constrained devices.

We propose to construct dynamic groups based on nodes sharing a common context. We argue that such a method opens perspectives for a large variety of applications, ranging from user entertainment (people hiking or skiing together) and healthcare (body area networks), to smart vehicles carrying smart goods (in the field of transport and logistics, as we describe in Section 2). In this paper, the common context is the movement information. More specifically, we consider two nodes being *together* if their movement *correlates* for a certain amount of time. Nevertheless, constructing groups based on the correlation of the movement information raises a number of questions:

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 298–315, 2007.

1. How to extract and communicate the movement information?
2. How to compute the correlation, taking into account the resource limitations of the sensor nodes?
3. How does the method scale with the number of nodes?
4. How accurate is the solution and which are the benefits and limitations?

The contribution of this paper is a lightweight, fast and cheap method for correlating the movement data among sensor nodes, for the purpose of clustering nodes moving together. Each node correlates the movement data generated by the local movement sensor with the movement data broadcast periodically by its neighbours. The result of the correlation is a measure of the confidence that one node shares the same context with its neighbours, for example that they are placed in the same car. We focus in this paper on correlating sensor nodes carried by vehicles on wheels.

We describe two possible practical solutions, one using tilt switches, and another one using MEMS accelerometers. In order to answer the aforementioned questions in detail, we analyse the scalability from several different perspectives (communication, energy, memory and execution speed), and discuss the most relevant advantages and limitations. The analysis is based on the experimental results obtained from testing with real sensor nodes. We use the Ambient μNode 2.0 platform [1] with the low-power MSP430 micro-controller produced by Texas Instruments, which offers 48kB of Flash memory and 10kB of RAM. The radio transceiver has a maximum data rate of 100kbps. Figure 1 shows the sensors used for extracting the movement information and the sensor node platform.

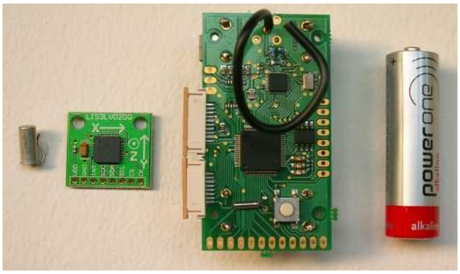


Fig. 1. Movement sensors and sensor node platform

In the following section we describe a concrete application setting in the field of transport and logistics, which best illustrates the idea of movement-based group awareness. Section 3 overviews the relevant related work. The general correlation method is described in Section 4. In Sections 5 and 6 we present the two practical solutions for autonomous group formation. Section 7 covers the analysis, advantages and limitations of both solutions, giving also comparative details whenever relevant. Finally, Section 8 formulates the conclusions.

2 Application Setting

Transport and logistics represent large-scale processes that ensure the delivery of goods from producers to shops. The distribution process starts at a *warehouse*, where an *order picker* gets an order list, assembles a rolling container (Returnable Transport Item - RTI), picks the requested products from the warehouse shelves, and loads them in the RTI. Next, the order picker moves the RTI to the *expedition floor*, a large area used for temporary storage (see Figure 2). The expedition floor is seen as a grid, where each cell of the grid is associated with a certain shop. At loading time, the *loading operators* place the RTIs into trailers, according to a *loading list*, derived from the delivery orders. Eventually, a truck pulls the trailer and delivers the goods to the shops.

Due to the large scale of the process, the transport company personnel (e.g. order pickers, loading operators) is prone to errors. It often happens that the order pickers make mistakes when filling the RTIs with goods, or that the RTIs are loaded in the wrong trailer. In addition, the products are sometimes stored in improper climate conditions, which is a serious problem in the case of perishable goods. WSN technology can be a solution to these problems, as sensor nodes offer precise control over the status (e.g. location, storage temperature) and history of the goods. Consequently, the RTIs and the products carried in them, equipped with sensor nodes, can check for errors and trigger alerts at the point of action. Movement-based group awareness is an essential component for achieving this vision of smart RTIs and goods.

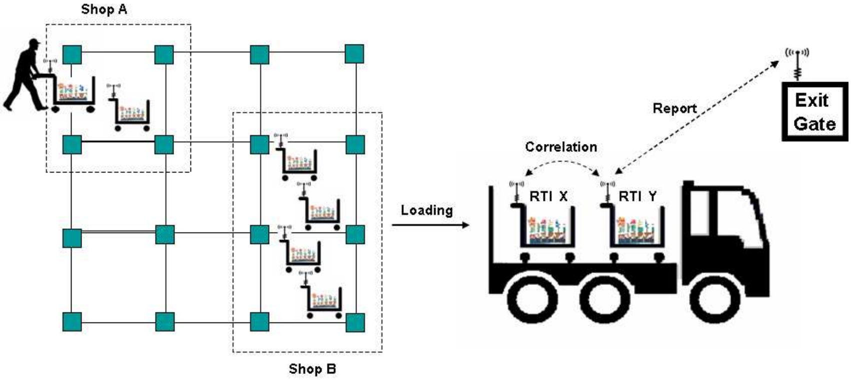


Fig. 2. Transport and logistics process diagram

The solution that we propose targets two specific problems. First, the goods from an RTI correlate their movement as the RTI is pushed, and report as a group to the device carried by the order picker. In this way, a missing or wrong item can be detected before arriving on the expedition floor. Second, any RTI placed in the wrong trailer should be signaled as the truck approaches the exit gate (see Figure 2). Since the distance between two RTIs or two trucks is quite short, the localization of the goods inside the RTI, or of the RTIs inside the truck cannot be done reliably with radio signal strength proximity techniques. We consider, however, highly probable that two different vehicles move differently in a certain time interval. Therefore, we propose to group the nodes based on the similarities and differences in the data generated by the movement sensors.

3 Related Work

Grouping of devices into clusters is a topic of interest in the field of wireless networks [11]. The clusterhead node is usually chosen based on different properties, such as the node capability, degree of dynamics, connectivity, etc. The main goal is to achieve energy efficiency at the networking layer, and therefore the application-level attributes are usually not of concern. Nevertheless, grouping based on application-specific attributes is studied in the field of service discovery [5]. Nodes organize into groups, in order to efficiently search for services. However, the grouping criteria is usually based on statically-assigned semantic descriptions, while we are interested in context-dependent, dynamic attributes.

Lam et al. [8] propose an algorithm for dynamic grouping based on the position and speed of mobile devices equipped with GPS sensors. Nodes within a certain area that move together (similar speed and direction) form a group. However, equipping each node with a GPS sensor is not a viable solution for WSNs, because of price and power consumption considerations.

In the project Smart-Its, Gellersen et al. [6] formulate the notion of context sharing. The idea is to associate two smart objects by shaking them together. As a result, the user can establish an application-level connection between two devices by imposing a brief, similar movement. In our work, we are interested in extending the idea of "moving together" at the group level, within large-scale industrial and business scenarios. Therefore, we propose a fast algorithm that correlates the movement over a larger time history, and analyse the accuracy, scalability, performance and limitation factors.

Lester et al. [9] use accelerometer data to determine if two devices are carried by the same person. Human locomotions represent a repeated activity that makes an analysis in the frequency domain possible. The authors use a coherence function to derive whether the two signals are correlated at a particular frequency. Our application domain poses, however, quite different challenges. There is no regularity in the movement of the RTIs that can facilitate an analysis in the frequency domain. Moreover, the computations involved in the frequency analysis can easily overcome the resources available on sensor nodes.

The Sensemble system [4] is meant to capture the expressive motion of a dance ensemble. Sensor nodes equipped with 6-axis inertial measurement units are worn at the wrists and ankles of dancers. The movement data is transfered at high speeds (1Mbps) towards a central computer, where a cross-covariance analysis is performed, in order to express the similarity of gestures and generate a musical feedback. Our solution is different, in the sense that sensor nodes compute the correlation online, autonomously. In addition, due to price limitations, we utilize low data rate radios and just one movement sensor per node.

4 General Method

The algorithm correlates the movement data generated by the movement sensors attached to different sensor nodes. Regardless of the movement sensor type, we use the same general method.

4.1 Computing the Correlation

Let **x** be one of the sensor nodes, which receives the sampled movement data from another sensor node **y**. Node **x** stores the latest sample values produced by the local movement sensor in a circular buffer X_C of size k. The buffer X_C is periodically transmitted to the neighbours at intervals $k\Delta t$, where Δt is the sampling interval. At step $i \geq 1$, **x** receives from **y** the buffer $Y_i = \{y_{(i-1)k+1}, y_{(i-1)k+2}, ..., y_{ik}\}$. Node **x** then copies the buffer X_C into a working copy $X_i = \{x_{(i-1)k+1}, x_{(i-1)k+2}, ..., x_{ik}\}$ and calculates the correlation coefficient over the last n sequences of data X_i and Y_i. More precisely, at each step i, the correlation coefficient is calculated over a sliding window of size $N = nk$, with the data $X = (x_{(i-n)k+1}, x_{(i-n)k+2}, ..., x_{ik})$ and $Y = (y_{(i-n)k+1}, y_{(i-n)k+2}, ..., y_{ik})$. Note that for $j \leq 0$, $x_j = y_j = 0$. If we denote the means of X and Y as $\bar{X}$ and $\bar{Y}$, respectively, the correlation coefficient is computed as follows:

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)\ var(Y)}} = \frac{\sum_{j=(i-n)k+1}^{ik} (x_j - \bar{X})(y_j - \bar{Y})}{\sqrt{\sum_{j=(i-n)k+1}^{ik} (x_j - \bar{X})^2 \sum_{j=(i-n)k+1}^{ik} (y_j - \bar{Y})^2}} \quad (1)$$

Table 1 shows the execution time for computing the correlation coefficient on one sensor node, with two sets of samples of size $N = 128$. We conclude that using the direct computation from Eq. 1 generates slow execution times, so it is not feasible for implementation on resource-constraint devices. Therefore, we propose a fast algorithm that updates the correlation coefficient at each step. For large data sequences (large k), the memory consumption is also reduced by storing only intermediate values (see Section 7.2 for an evaluation of the memory consumption).

The algorithm is the following. At step i, node **x** receives the buffer Y_i from node **y**. Node **x** then calculates the following sums:

1. $S_i^x = \sum_{j=(i-1)k+1}^{ik} x_j$ and $S_i^y = \sum_{j=(i-1)k+1}^{ik} y_j$
2. $\sigma_i^x = \sum_{j=(i-1)k+1}^{ik} x_j^2$ and $\sigma_i^y = \sum_{j=(i-1)k+1}^{ik} y_j^2$
3. $S_i^{xy} = \sum_{j=(i-1)k+1}^{ik} x_j y_j$

Table 1. Execution times on MSP430 microcontroller, for N=128, k=16

Method	**Operation**	**Time** [ms]
Direct computation	Average	68.91
	Variance, covariance and correlation coefficient	275.65
	Total	**344.56**
Incremental algorithm	Auxiliary sums	0.81
	Average, variance, covariance and correlation coefficient	5.47
	Total	**6.28**

Afterward, node **x** computes the following values:

$$\bar{X}_i = \bar{X}_{i-1} + \frac{S_i^x - S_{i-n}^x}{N} \tag{2}$$

$$\bar{Y}_i = \bar{Y}_{i-1} + \frac{S_i^y - S_{i-n}^y}{N}$$

$$var_i(X) = var_{i-1}(X) + \frac{\sigma_i^x - \sigma_{i-n}^x}{N} - (\bar{X}_i^2 - \bar{X}_{i-1}^2) \tag{3}$$

$$var_i(Y) = var_{i-1}(Y) + \frac{\sigma_i^y - \sigma_{i-n}^y}{N} - (\bar{Y}_i^2 - \bar{Y}_{i-1}^2)$$

$$cov_i(X,Y) = cov_{i-1}(X,Y) + \frac{S_i^{xy} - S_{i-n}^{xy}}{N} - (\bar{X}_i\bar{Y}_i - \bar{X}_{i-1}\bar{Y}_{i-1}) \tag{4}$$

Finally, node **x** computes the new value of the correlation coefficient:

$$\rho_i(X,Y) = \frac{cov_i(X,Y)}{\sqrt{var_i(X)\ var_i(Y)}} \tag{5}$$

The proofs of Eq. 3 and 4 are given in Appendix A. We make the following observations:

- For all $j \leq 0$, S_j^x, S_j^y, σ_j^x, σ_j^y, S_j^{xy}, $var_j(X)$, $var_j(Y)$ and $cov_j(X,Y)$ are 0.
- If $var_i(X) = 0$ and $var_i(Y) = 0$, we take $\rho_i(X,Y) = \rho_{i-1}(X,Y)$.
  If $var_i(X) = 0$ and $var_i(Y) \neq 0$ or the other way around, we decrease $\rho_i(X,Y)$ with a value proportional to the positive variance.

The algorithm proves to be much faster (by a factor of about 55) than the direct computation of Eq. 1, as shown in Table 1. This result makes the implementation of the online correlation on sensor nodes possible.

4.2 Experimental Setting

We perform two types of experiments, in which we test the proposed method:

1. The first type of experiment is intended to reproduce the movement pattern of the smart goods, in which items equipped with sensor nodes are placed in RTIs maneuvered by people. Throughout the tests, we use two RTIs on wheels, which we push on a flat surface. For detecting joint movement, two sensor nodes are placed on the same RTI, while for separate movement, each sensor node is placed on a different RTI.
2. The second type of experiment maps to the setting where RTIs are loaded into and carried by trucks. We use instead two regular cars. Our experiments include the following types of movement: normal driving, accelerating and breaking, forward and backward maneuvers, curves, driving on even and uneven surfaces. Two sensor nodes are placed in the same car for joint movement, while for separate movement nodes are placed inside different cars.

Each experiment lasts approximately 10 minutes. The sensor nodes broadcast the movement data together with the correlation coefficient calculated locally. A gateway node logs the coefficients and the samples from both sensor nodes to a computer through a serial interface.

4.3 Parameters

Table 2 lists the values of the parameters used in the experiments, which are chosen considering the platform constraints (sampling interval Δt and data size k) and the scenario particularities (time history T).

Table 2. Experimental parameters for correlating the movement data

Parameter	**Explanation**	**Value**
k	Size of current data sequence	16 (2s)
n	No. of data sequences in data queue	8
$N = nk$	Size of queue	128
Δt	Time unit (sampling interval)	125ms
$T = N\Delta t$	Time history	16s

4.4 Synchronization

The synchronization between two sets of data to be correlated is very important for accurately calculating the correlation coefficient. We assume that the communication delay, plus the time for processing the incoming and outgoing buffers is $<< \Delta t = 125ms$. Therefore, we ignore the time spent on communication, and we make sure that each node copies its last k samples in the local working buffer, at the moment it receives the samples from the neighbouring nodes. If a transmission error occurs, the next message contains again the latest k samples. Using this method, we can achieve *implicit synchronization* between the two sensors.

5 Solution I - Tilt Switches

A ball-contact tilt switch (also referred to as ball switch or tilt switch) is a simple and cheap sensor, used in a large range of applications for coarse movement detection. Usually, the sensor is expected to provide binary information on the status of the device it is attached to (e.g. stationary/moving).

5.1 Extracting the Movement Information

In our experiments, we are using the ASSEMTECH CW1300-1 tilt switch [2]. The price is below 2 EUR and the power consumption is approximately 2μW. Our solution is based on counting the number of contacts made by the switch ball per time unit, as the object is moved. We make the following observations:

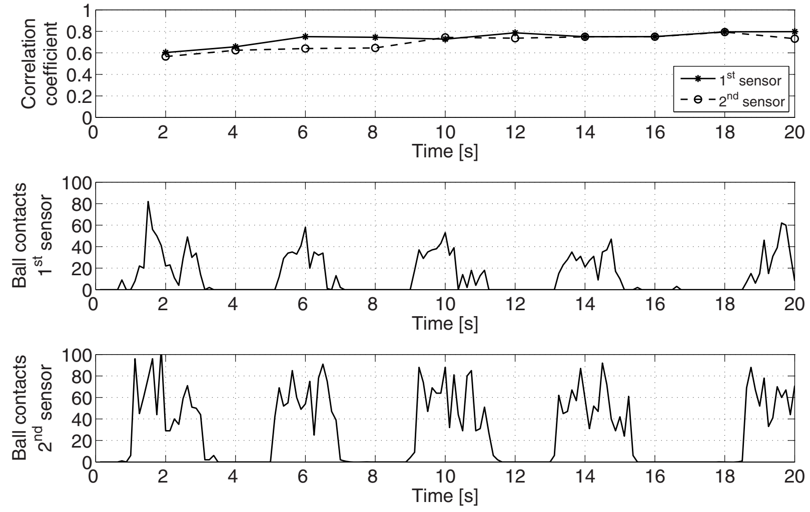
1. It is possible to distinguish the starting and stopping states (acceleration and deceleration) from the constant movement.
2. The sensitivity depends on the position of the ball switch.
3. The results are reproducible with other switches of the same type. Although the actual values vary due to the inherent sensitivity differences and imperfect alignment of the sensitive axis, the movement pattern remains similar.

5.2 Experimental Results

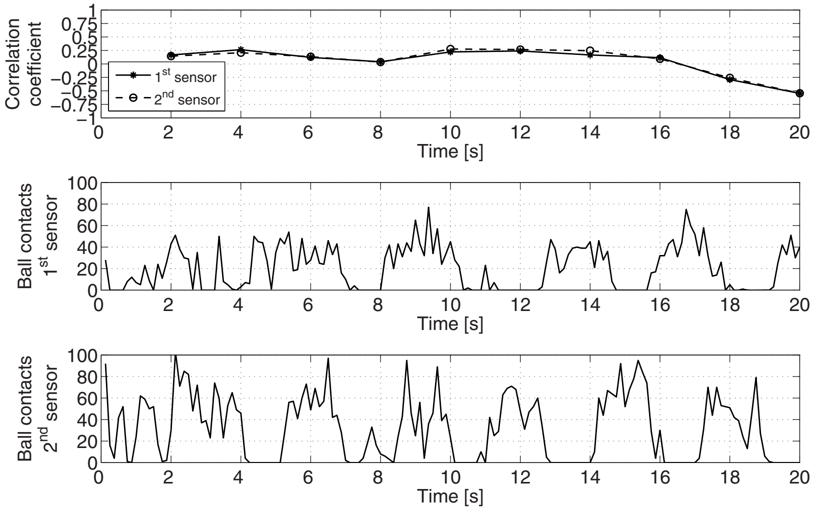
Figure 3 shows the typical behaviour of the algorithm for joint and separate movements, over a period of 20 seconds. The plots at the top of the figures show the correlation coefficients calculated by the sensor nodes over the time history T, while the two bottom plots show the sampled data from the tilt switches. We make the following observations:

- The sampled data from Figure 3(a) show a pattern, corresponding to the alternate stationary and movement periods.
- There is a clear distinction between the moving and stationary cases: when the sensor nodes are static, the number of ball contacts is 0. Therefore, in a static situation, nodes may not need to send the whole movement buffer, but just a short indication of their state, saving thus energy.
- The method is successful in distinguishing between correlated and uncorrelated movements, during both types of experiments. A high correlation coefficient indicates that the sensor nodes move together (Figure 3(a)), while a low correlation coefficient shows a separate movement (Figure 3(b)).

We represent the histograms of the correlation coefficients obtained on the entire duration of the experiments (≈10 minutes) in Figure 5 left side, normalized to a percentage scale. We notice the difference between the correlation coefficients computed when sensors move together (Figures 5(a) and 5(c)) and separately (Figures 5(e) and 5(g)). A more detailed analysis of the results is given in Section 7.1.



(a) Two nodes with tilt switches moving together.



(b) Two nodes with tilt switches moving separately.

Fig. 3. A typical behaviour of the algorithm for joint and separate movements, using tilt switches

6 Solution II - Accelerometers

MEMS accelerometers have become increasingly popular recently, due to their relatively low price compared with the performance offered. The range of applications is quite broad, from movement or free-fall detection to gaming or virtual reality, and inertial navigation systems (INS) [12]. The operating principle is based on measuring the displacement of a proof mass when an acceleration is applied. The accelerometer measures, therefore, the applied acceleration (including gravitation), and outputs the values of the projections along its sensitive axis.

6.1 Extracting the Movement Information

By using accelerometers, it is possible to extract elaborate information about the movement, such as the speed and distance. However, to calculate the speed and position accurately, information provided by gyroscopes has to be used for maintaining an absolute positional reference. In this way, the overall complexity and price of the system increase significantly. Moreover, the accumulation of errors require elaborated filtering and prediction techniques.

From these considerations, it appears that the resource-constraint sensor nodes are not yet capable of extracting and correlating speed or distance information. Therefore, we propose a simplified solution, which considers the magnitude of the acceleration vector $\| \mathbf{a} \| = \sqrt{a_x^2 + a_y^2 + a_z^2}$. The reason is that the magnitude of the sensed acceleration is the same in any frame of reference. Consequently, the alignment and orientation of the sensors are no longer important.

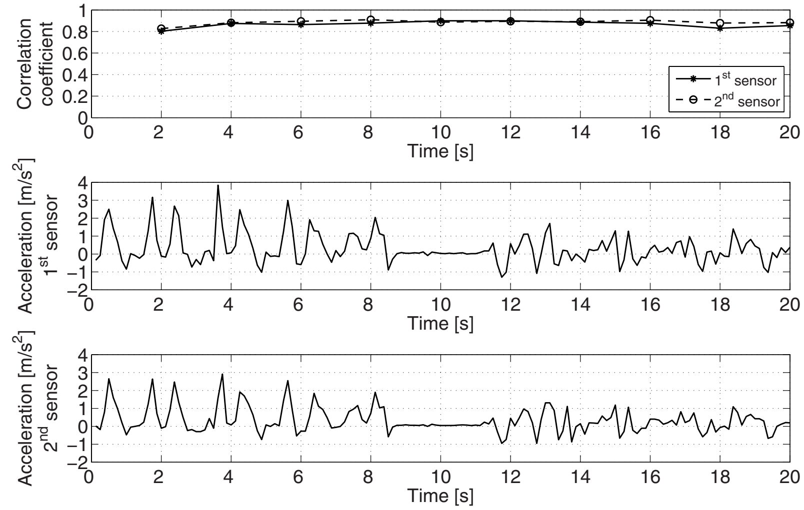
In our experiments, we are using the LIS3LV02DQ three-axis accelerometer from STMicroelectronics [3]. The price is around 15 USD and the typical power consumption is 2mW. The list of features include user selectable full scale of ±2g, ±6g, I^2C/SPI digital interface, programmable threshold for wake-up/free-fall and various sample rates up to 2.56kHz.

6.2 Experimental Results

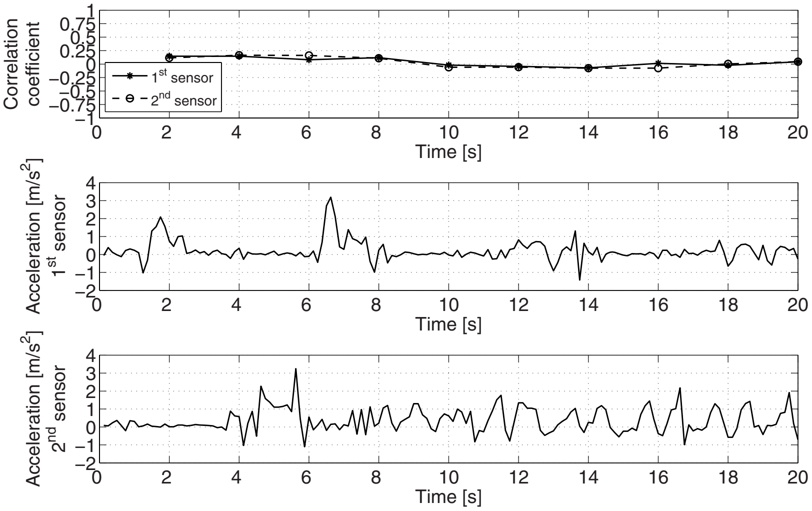
Figure 4 shows the typical behaviour of the algorithm for joint and separate movements, over a period of 20 seconds. The plots at the top of the figures show the correlation coefficients calculated by the sensor nodes over the time history T, while the two bottom plots show the magnitude of the acceleration calculated by the sensors, relative to 1g (the constant gravitational component). We make the following observations:

- A node can deduce that it is static by calculating the standard deviation over the current data sequence k: a relatively small standard deviation implies that the node is static. Therefore, similar to the tilt switch case, in the static situations nodes may just send a short indication of their state.
- The method is successful in distinguishing between correlated and uncorrelated movements, for both types of experiments. A high correlation coefficient indicates that the sensor nodes move together (Figure 4(a)), while a low correlation coefficient shows a separate movement (Figure 4(b)).

We represent the histograms of the correlation coefficients in Figure 5 right side, normalized to a percentage scale. We notice the difference between the



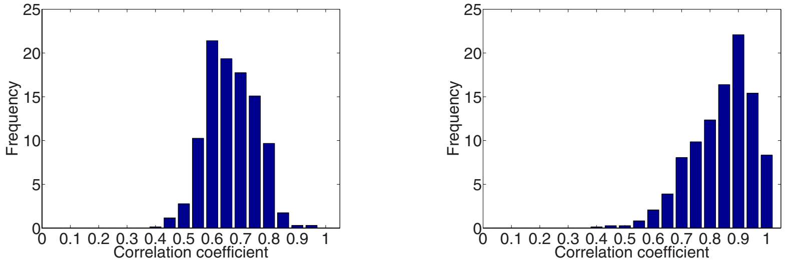
(a) Two nodes with accelerometers moving together



(b) Two nodes with accelerometers moving separately

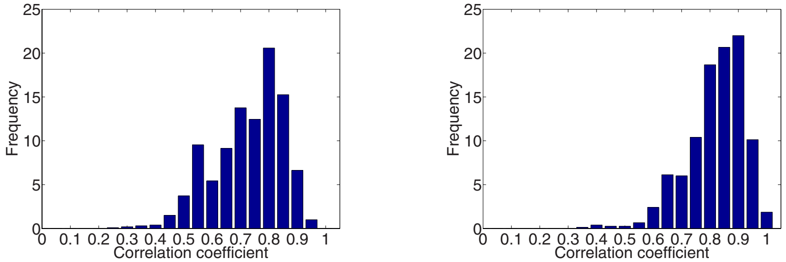
Fig. 4. A typical behaviour of the algorithm for joint and separate movements, using accelerometers

correlation coefficients computed when sensors move together (Figures 5(b) and 5(d)) and separately (Figures 5(f) and 5(h)). A more detailed analysis of the results is given in Section 7.1.



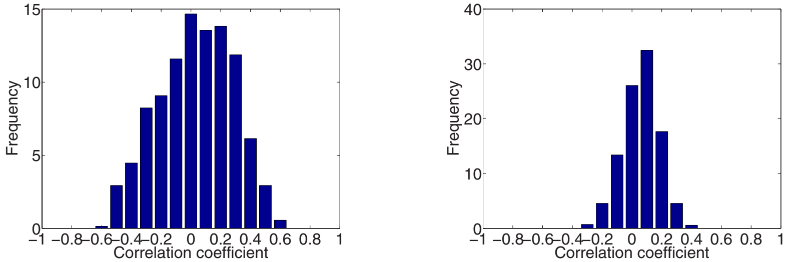
(a) Tilt switches on RTIs, moving together.

(b) Accelerometers on RTIs, moving together.



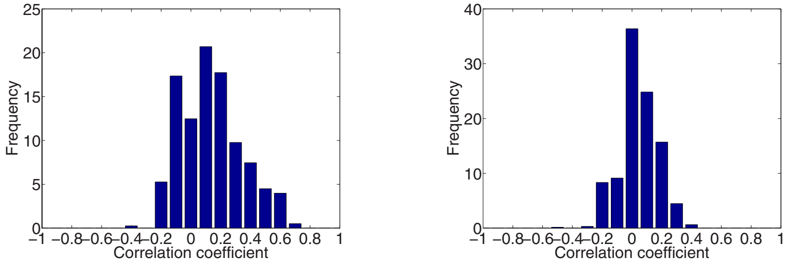
(c) Tilt switches in cars, moving together.

(d) Accelerometers in cars, moving together.



(e) Tilt switches on RTIs, moving separately.

(f) Accelerometers on RTIs, moving separately.



(g) Tilt switches in cars, moving separately.

(h) Accelerometers in cars, moving separately.

Fig. 5. Histograms of correlation coefficients for tilts switches and accelerometers

7 Analysis

In this section, we discuss the two proposed solutions, and analyse the accuracy and scalability problems, pointing out the advantages and limitations.

7.1 Accuracy

One of the major questions is how accurate are the proposed methods. Thorough tests within large scale settings as described in Section 2 are subject to future work (see also Section 8). In Table 3, we present a brief statistical analysis of the results obtained from our initial experiments with RTIs and cars, as explained in Section 4.2. The mean values indicate a constant difference of more than 0.6 between joint and separate movement, in any of the listed settings. The standard deviation values suggest that the accelerometers provide more precise results, fact confirmed by the histograms from Figure 5. The accuracy column shows the percentage of the correct decisions, where we consider a simple decision threshold $Th_C = 0.5$ (i.e. a correlation coefficient larger than 0.5 means joint movement and the other way around). The accelerometer-based solution proves more accurate, with 3.4% on average and a maximum of 5.2%. In addition, due to their better sensitivity, the accelerometers can identify reliably the separate movement situation.

Table 3. Statistical values

Sensor	**Setting**	**Movement type**	**Mean**	**Stdev**	**Accuracy [%]**
Tilt switch	RTI	joint	0.641	0.087	95.89
Tilt switch	RTI	separate	-0.017	0.249	99.45
Tilt switch	car	joint	0.700	0.121	93.77
Tilt switch	car	separate	0.086	0.208	95.50
Accelerometer	RTI	joint	0.817	0.106	99.31
Accelerometer	RTI	separate	0.009	0.124	100
Accelerometer	car	joint	0.796	0.102	98.93
Accelerometer	car	separate	-0.003	0.127	100

7.2 Scalability

We present the factors that influence the maximum number of nodes supported by our proposed correlation methods. We denote the maximum number neighbouring nodes as M. It follows that a node has maximum $M-1$ neighbours, for which it computes the correlation coefficients.

Communication (Medium Access). We estimate the maximum number of neighbouring nodes M as follows. Each node transmits a data sequence every $k\Delta t$. If a TDMA-based MAC protocol is used, then the frame length T_f has to be at most $k\Delta t$, so that each node has a chance to transmit the data. The slot time T_s of a node is therefore bounded by $MT_s = T_f \leq k\Delta t$. Depending on the radio chip used, the slot time for sending a data packet can be computed. In our experiments, $T_s = 20$ms, which leads to $M = 100$.

Memory. The available memory (RAM and FLASH) is usually a critical resource on sensor nodes. The FLASH usage is not a problem, since the code memory footprint of our implementation on the sensor node platform amounts to 2.1kB out of 48kB available. Considering the RAM, Table 4 shows the data structures required by the correlation method, and the associated sizes. In the case of recent low-power controllers equipped with 10kB RAM, the maximum number of nodes is $M = 106$.

Table 4. Memory requirements

Data structure	**Size** [bytes]
Data sequence to send (X_i)	16
Received data sequence (Y_i)	16
S_i^x, S_i^y	2
σ_i^x, σ_i^y, S_i^{xy}	4
$\bar{X}_i$, $\bar{Y}_i$, $var_i(X)$, $var_i(Y)$, $cov_i(X,Y)$, $\rho_i(X,Y)$	4
Auxiliary sums	n × M × 10
Correlation data (Eq. 3-5)	$M \times 16 - 8$

Execution Time. Since the correlation algorithm runs online, the nodes must have enough time within one slot to receive and process the incoming data. It follows that the execution time T_e must be much smaller than the slot time: $T_e << T_s = T_f/M \Rightarrow M << T_f/T_e$. For the values used in our experiments, we get $M << 318$. This shows that the speed of the algorithm is not a limiting factor from the scalability point of view.

Energy. Estimating the energy consumption is always important for the battery powered sensor nodes. We consider the radio communication and sensor operation as the most costly functions in terms of energy. For communicating the sampled data, a node performs $M - 1$ receptions and one transmission every frame T_f. Typical radio current consumption on Ambient μNodes is 12.8mA for reception and 11mA for transmission. In addition, the current consumed with operating the sensors is 0.64μA for tilt switches and 0.65mA for accelerometers. Figure 6 shows the operating time for a node with a typical 1000mAh battery. The running time is represented depending on the number of neighbouring nodes M. The maximum values for M are deduced from the previous analysis, as being $M = 100$. As an example, for $M = 50$, the system can operate for approximatively 156 hours of continuous movement when using tilt switches and 142 hours when using accelerometers. However, the overall lifetime of a node increases if the movement intervals are short: during stationary periods, a node needs to send only an indication of its status, not the whole movement buffer.

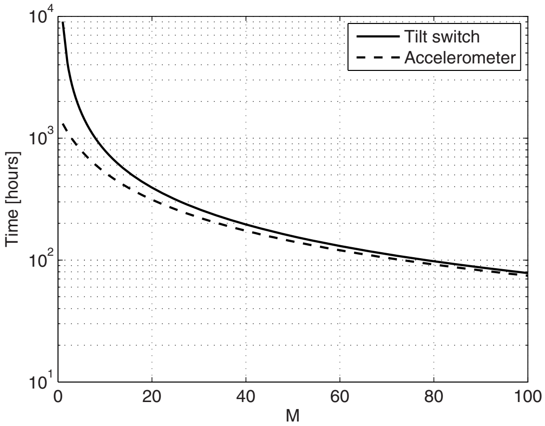


Fig. 6. The running time for continuous movement

7.3 Discussion

In what follows, we comment on the most important advantages and limitations of both solutions, giving also comparative details whenever relevant.

Advantages

1. *Autonomous group awareness.* The proposed methods aim at establishing groups autonomously, based on common dynamic properties of the group members. No infrastructure support is needed.
2. *Simplicity.* The overall system (hardware and software) is kept very simple. This implies both a low price range and the feasibility of the implementation on resource constrained devices.
3. *Robustness to constructive differences of sensors.* The correlation coefficient gives an indication on the degree of similitude of two signals. It is known that the result is neither affected by scaling the signals with a certain factor, nor by adding/subtracting a constant value. This makes our method inherently robust to constructive differences of sensors, such as: calibration factors, zero-offset values, differences in sensitivity.
4. *Distinction between ensemble and separate movements.* Figure 5 indicates a good behaviour, with separable thresholds for distinguishing between ensemble and disjoint movement. However, the solution employing accelerometers proves more accurate in all tested situations.
5. *Implicit synchronization.* There is one important factor that can adversely affect the correctness of the correlation result, and that is time synchronization. It is therefore essential that the data sequences X_i, Y_i are synchronized when computing the correlation coefficient. For this reason, X_C is implemented as a circular buffer, so that the incoming data from neighbours is correlated with the latest values sampled on the current node. Moreover, it is preferable not to use any retransmission mechanisms, since occasional packet losses do not affect the synchronization.

6. *Saving power while stationary.* In static situations, nodes can save energy by transmitting just a short indication of their status.
7. *Extended features.* More accurate results may be obtained by correlating extended movement features, such as direction or heading, speed, distance. In this sense, the accelerometer-based solution is much richer in possibilities.

Limitations

1. *Alignment and orientation.* Because movement is always relative to a frame of reference, different alignment or orientation of the sensors may produce misleading results. In the case of tilt switches, a similar alignment is necessary for obtaining a correct behaviour, such as in Figure 3(a). In contrast, for the accelerometer-based solution, no alignment is needed, since we are correlating the magnitude of the acceleration vector. This approach might also yield better results in the case of tilt switches, if a system such as Porcupine [7] is used, where each node is equipped with several switches oriented along different axis.
2. *Unpredictable delays.* Networking stack and sensor sampling delays can adversely affect the time synchronization of data sequences X_i, Y_i, eventually leading to errors in the correlation estimation. Timestamping the incoming Y_i at the receiver node $\mathbf{x}$, and choosing the corresponding X_i to correlate with, can alleviate the problem of networking stack delays.
3. *Placement on loose frames.* Throughout our tests, the sensor nodes are placed on the same rigid frame. We expect that the movement data becomes less correlated if the sensors are attached to loosely-coupled frames, such as the wagons of a train.
4. *Rounding errors.* In our incremental algorithm, rounding errors due to operations on floats may accumulate and affect the results in time. However, since the sampled data is bounded, the calculations can be done on integers and thus the accumulation of errors is prevented.
5. *Multihop networks.* Our solution is valid only for one-hop networks. A multihop network would impose a transitive correlation relation, implemented by border nodes or by a dominating set of nodes. This is subject of future work, see also Section 8.

8 Conclusions

This paper proposes a method for constructing dynamic groups based on movement information. Nodes are considered part of the same group if their movement correlate for a certain amount of time. For extracting the movement information we investigate two solutions, one using tilt switches, the other one using accelerometers. On the one hand, the solution using tilt switches proves to be cheaper, simpler and less energy consuming. On the other hand, the solution using accelerometers is more reliable in distinguishing between ensemble and separate movements and it does not need any sensor alignment. Nevertheless, the solution is more complex, as the magnitude of the acceleration has to be calculated from the three samples corresponding to the three axes. The scalability analysis shows a maximal network density of 100 nodes for both solutions.

For future work, a large-scale experiment is required, in order to obtain a thorough estimation of the accuracy and scalability of our method. Since such an experiment may imply a multi-hop network, and because it is not feasible to propagate the movement data over multiple hops, we plan to test a transitive correlation relation implemented by border nodes or by a dominating set. Having only a subset of nodes recompute the correlation, to check whether the group is still moving in unison, would also improve the energy and communication performance of our solution. We further intend to experiment with fusing information from multiple types of sensors, in order to improve the overall accuracy.

References

1. Ambient Systems. http://www.ambient-systems.net.
2. ASSEMTECH CW1300-1 ball-contact tilt switch. http://www.farnell.com/datasheets/67723.pdf.
3. STMicroelectronics LIS3LV02DQ 3-axis linear accelerometer. http://www.st.com/stonline/products/literature/ds/11115.pdf.
4. Ryan Aylward and Joseph A. Paradiso. Sensemble: a wireless, compact, multi-user sensor system for interactive dance. In *Proceedings of the 6th Conference on New interfaces for Musical Expression (NIME '06)*, pages 134–139, 2006.
5. Dipanjan Chakraborty, Anupam Joshi, Yelena Yesha, and Tim Finin. Toward Distributed Service Discovery in Pervasive Computing Environments. *IEEE Transactions on Mobile Computing*, 5(2):97–112, February 2006.
6. Lars Erik Holmquist, Friedemann Mattern, Bernt Schiele, Petteri Alahuhta, Michael Beigl, and Hans-Werner Gellersen. Smart-its friends: A technique for users to easily establish connections between smart artefacts. In *UbiComp'01*, pages 116–122, London, UK, 2001. Springer-Verlag.
7. Kristof Van Laerhoven and Hans-Werner Gellersen. Spine versus porcupine: A study in distributed wearable activity recognition. In *Proceedings of the Eighth International Symposium on Wearable Computers (ISWC'04)*, pages 142–149, 2004.
8. Gary Hoi Kit Lam, Hong Va Leong, and Stephen Chi fai Chan. Gbl: Group-based location updating in mobile environment. In *9th International Conference on Database Systems for Advanced Applications (DASFAA)*, pages 762–774, 2004.
9. Jonathan Lester, Blake Hannaford, and Gaetano Borriello. "Are You with Me?" - using accelerometers to determine if two devices are carried by the same person. In *Pervasive*, pages 33–50, 2004.
10. M. Marin-Perianu, N. Meratnia, M. Lijding, and P. Havinga. Being aware in wireless sensor networks. In *15th IST Mobile and Wireless Communication Summit, Capturing Context and Context Aware Systems and Platforms Workshop*, 2006.
11. Ivan Stojmenovic, Mahtab Seddigh, and Jovisa Zunic. Dominating sets and neighbor elimination-based broadcasting algorithms in wireless networks. *IEEE Transactions on Parallel and Distributed Systems*, 13(1):14–25, 2002.
12. C.W. Tan, S. Park, K. Mostov, and P. Varaiya. Design of gyroscope-free navigation systems. In *Intelligent Transportation Systems, Oakland, USA*, pages 286–291, 2001.

A Appendix

The proofs of Eq. 3 and 4 are the following:

$$\begin{aligned}
var_i(X) = & \sum_{j=(i-n)k+1}^{ik} \frac{(x_j - \bar{X}_i)^2}{N} = \sum_{j=(i-n)k+1}^{ik} \frac{x_j^2}{N} - \bar{X}_i^2 = \\
& = (\sum_{j=(i-n-1)k+1}^{(i-1)k} \frac{x_j^2}{N} - \bar{X}_{i-1}^2) + \sum_{j=(i-1)k+1}^{ik} \frac{x_j^2}{N} - \\
& - \sum_{j=(i-n+1)k+1}^{(i-n)k} \frac{x_j^2}{N} - \bar{X}_i^2 + \bar{X}_{i-1}^2 = \\
& = var_{i-1}(X) + \frac{\sigma_i^x - \sigma_{i-n}^x}{N} - (\bar{X}_i^2 - \bar{X}_{i-1}^2)
\end{aligned}$$

$$\begin{aligned}
cov_i(X,Y) = & \sum_{j=(i-n)k+1}^{ik} \frac{(x_j - \bar{X}_i)(y_j - \bar{Y}_i)}{N} = \sum_{j=(i-n)k+1}^{ik} \frac{x_j y_j}{N} - \bar{X}_i \bar{Y}_i = \\
& = (\sum_{j=(i-n-1)k+1}^{(i-1)k} \frac{x_j y_j}{N} - \bar{X}_{i-1} \bar{Y}_{i-1}) + \sum_{j=(i-1)k+1}^{ik} \frac{x_j y_j}{N} - \\
& - \sum_{j=(i-n-1)k+1}^{(i-n)k} \frac{x_j y_j}{N} - \bar{X}_i \bar{Y}_i + \bar{X}_{i-1} \bar{Y}_{i-1} = \\
& = cov_{i-1}(X,Y) + \frac{S_i^{xy} - S_{i-n}^{xy}}{N} - (\bar{X}_i \bar{Y}_i - \bar{X}_{i-1} \bar{Y}_{i-1})
\end{aligned}$$

Zone-Based RSS Reporting for Location Fingerprinting

Mikkel Baun Kjærgaard[1], Georg Treu[2], and Claudia Linnhoff-Popien[2]

[1] Department of Computer Science, University of Aarhus, Denmark
`mikkelbk@daimi.au.dk`
[2] Mobile and Distributed Systems Group, Institute for Informatics,
Ludwig-Maximilian University Munich, Germany
`{georg.treu,linnhoff}@ifi.lmu.de`

Abstract. In typical location fingerprinting systems a tracked terminal reports sampled *Received Signal Strength (RSS)* values to a location server, which estimates its position based on a database of pre-recorded RSS fingerprints. So far, poll-based and periodic RSS reporting has been proposed. However, for supporting proactive *Location-based Services (LBSs)*, triggered by pre-defined spatial events, the periodic protocol is inefficient. Hence, this paper introduces zone-based RSS reporting: the location server translates geographical zones defined by the LBS into RSS-based representations, which are dynamically configured with the terminal. The terminal, in turn, reports its measurements only when they match with the configured RSS patterns. As a result, the number of messages exchanged between terminal and server is strongly reduced, saving battery power, bandwidth and also monetary costs spent for mobile bearer services. The paper explores several methods for realizing zone-based RSS reporting and evaluates them simulatively and analytically. An adaption of classical Bayes estimation turns out to be the best suited method.

1 Introduction

Location-based Services (LBSs) compile information for their users based on the position of one or several target persons. LBSs can be initiated on request by the user, e.g., for being informed about nearby *Points of Interest (PoIs)*, or they can be initiated on the arrival of certain spatial events, such as the target person entering or leaving a pre-defined geographic zone. Services of the first type are called *reactive*, while the latter ones are *proactive*.

Another distinction of fundamental technical concern is whether an LBS is used *indoors* or *outdoors*. So far, there is no single positioning system that supports both environments in an acceptable quality. While high-quality receivers for the *Global Positioning System (GPS)* are meanwhile integrated in mass market cellular phones, GPS only works outdoors and not inside buildings. The most popular indoor localization technique to-date is *Location Fingerprinting (LF)*, having the major advantage to exploit already existing network infrastructures,

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 316–333, 2007.

like IEEE 802.11 or GSM, which avoids extra deployment costs and effort. Based on a database of pre-recorded measurements of *Received Signal Strength (RSS)* values sampled from different locations within a building, denoted as fingerprints, a mobile terminal's location is estimated by inspecting the RSS values it currently measures.

Resource-constrained terminals which are unable to store the fingerprinting database, such as mobile phones or active badges, are supported by a central *location server*. The server accesses the database and estimates their location based on RSS measurements conducted at the terminal. So far, measured RSS values are either transmitted on request, or the terminal updates them periodically with the location server, according to a pre-defined update interval. The associated problem is that periodic updating generates an excessive number of messages, if the target person changes her location only sporadically.

The periodic protocol performs especially badly if it only needs to be observed when the target enters or leaves certain pre-defined update zones, which is the case for proactive LBSs: As it turns out, by automatically detecting update zones, not only proactive single-target LBSs can be realized, e.g., for notifying the LBS user as soon as she is near a PoI. Also proactive community services, which consider the positions of multiple targets, are possible. An example is proximity detection [1], which automatically detects when two mobile targets have entered below a pre-defined proximity distance. In this case the update zones for each target are dynamically configured based on the current distance to the other.

This paper explores a novel, more efficient approach for realizing zone detection based on LF: The location server dynamically configures the terminal with update zones defined in terms of RSS patterns. Only when the terminal detects a match between its current measurements and these patterns, that is, when it enters or leaves the zone, it notifies the server about the fact. The associated challenge is the adequate definition of RSS patterns, for which the paper proposes several methods and compares them with respect to message efficiency, computational overhead, and detection accuracy. Also, the methods' support for different shapes and sizes of the zones are evaluated. As it turns out, the approach strongly reduces the message exchange at the air-interface, which has the following advantages:

First, by avoiding excessive messages exchanged with the location server, the power consumption of the tracked terminals is significantly lowered. Second, valuable bandwidth is saved and monetary costs the targets have to spend for mobile data services are reduced. The latter aspect is of special importance for cross-organizational scenarios, when the update messages can not be directed over the network that yields the RSS measurements, but, e.g., only by using public bearer services like GPRS or UMTS packetswitched. Third, the approach avoids that the terminals need to continuously switch back and forth between communication mode for sending messages and scanning mode for observing RSS values, which is an actual problem for many 802.11 adapters. Finally, by reducing the general amount of location information collected about the terminal, privacy of the target person is enhanced.

The paper is structured as follows. The next section discusses alternatives ways of organizing LF systems and motivates and explains the chosen architecture and protocol for zone-based RSS reporting. Several methods for representing geographical zones in terms of RSS patterns are devised in Section 3 and compared analytically and by simulation in Section 4. Section 5 overviews related work. A conclusion and a discussion of further work is given in Section 6.

2 Architecture and Protocol

This work assumes LF systems to be organized in a *terminal-assisted* fashion, i.e., the terminal conducts the RSS measurements and the location server estimates its location based on the fingerprinting database. Alternatively, LF could also be done in a *network-based* as well as a *terminal-based* way, see [2] for a classification of positioning methods. This section first discusses the pros and cons of these two alternatives. Then, an overview about efficient position update methods devised for terminal-based positioning like GPS, which motivated this work, is given. Finally, the novel protocol proposed for terminal-assisted LF is presented.

2.1 Alternative LF Architectures

In network-based LF systems the base stations measure the RSS values of their clients and forward them to the server, which, in turn, estimates the terminal's location. Thus, the whole procedure, including measuring as well as location estimation, takes place in the network. Network-based LF, however, comes with several pitfalls. First, the base stations need to be especially configured and attached to the location server, which hinders cross-organizational operation. Second, the target person's privacy control is very limited, because all of her movements are observed at the location server. Third, there is no obvious way for saving the energy of the terminal, which continuously has to emit radio beacons for being tracked.

In terminal-based LF the RSS measurements and the location estimation takes place at the mobile terminal, which caches the fingerprinting database. The approach enhances the privacy of the target person, because less data is collected about her than in the network-based scenario. Also, terminal-based LF enables cross-organizational operation "in the wild" [3], i.e., base stations not controlled by the location server can be included. Finally, terminal-based LF can be combined with the existing position update methods described below, where the position is determined at the device and reported to the LBS only when needed. From an architectural viewpoint this is similar to using GPS. A drawback of terminal-based LF not present with GPS, however, is that the fingerprinting database has to be stored at the device, which is not an option for resource-constrained terminals like mobile phones and active badges. Also, sophisticated location estimation algorithms conducted at the device may overstrain its computational capacities. Finally, every time the fingerprinting database is changed the terminals have to be re-synchronized, which creates severe scalability problems, independent of the terminal type.

2.2 Existing Position Update Methods

For supporting proactive LBSs as well as services which continuously track the position of a target, different position update methods have been proposed and compared. The goal is to provide for an efficient transmission of position data between a location server in the Internet and a mobile device using terminal-based positioning like GPS [4,5,6]. The methods are motivated by periodic reporting, according to a pre-defined *update interval*, being inefficient. As it turns out, long update intervals increase the server's uncertainty about the mobile's position, which negatively affects the quality of the LBS. On the other hand, short intervals generate an excessive number of messages in case the target person changes her location only sporadically. Messages are also wasted when the target never approaches the locations that are relevant for interaction with the LBS.

A more efficient technique is distance-based position reporting: The terminal is dynamically configured with a certain *update distance*, which prescribes the line-of-sight distance between two consecutive position reports. A way to further reduce messages is dead reckoning: Based on observed movement parameters like speed and direction, the location server estimates the mobile's current position. The most flexible method is zone-based reporting: Position updates are only reported when the terminal enters or leaves a pre-defined geographical *update zone*.

2.3 Zone-Based Updating for Terminal-Assisted LF

This paper explores zone-based updating for terminal-assisted LF, enabling the efficient realization of proactive LF-based LBSs.

Figure 1 illustrates the proposed procedure: First, the mobile terminal registers with the location server (1) and then starts observing the RSS values of the surrounding base stations (2). An LBS application server can subscribe to zone-based updates by sending a respective request message to the location server (3). The request carries the zone definition, either in terms of geographical coordinates, e.g. as a circle or a polygon, or symbolically, e.g. as a floor section. The location server then translates the geographical update zone into an RSS-based representation, which parameterizes one of the *detection methods* presented in Section 3. The configuration is passed on to the mobile device (4), where it is continuously compared to measured RSS values. Only when the current measurements match the zone representation, they are reported (5). At the location server, it is checked whether the updated RSS values correctly correspond to entering or leaving the update zone. If so, a position update is sent to the LBS application server (6). If a position update request is canceled by the LBS (7), the location server notifies the terminal about the fact (8).

It can be seen that terminal-assisted LF in the described configuration has all the advantages of terminal-based LF, including update efficiency and enhanced

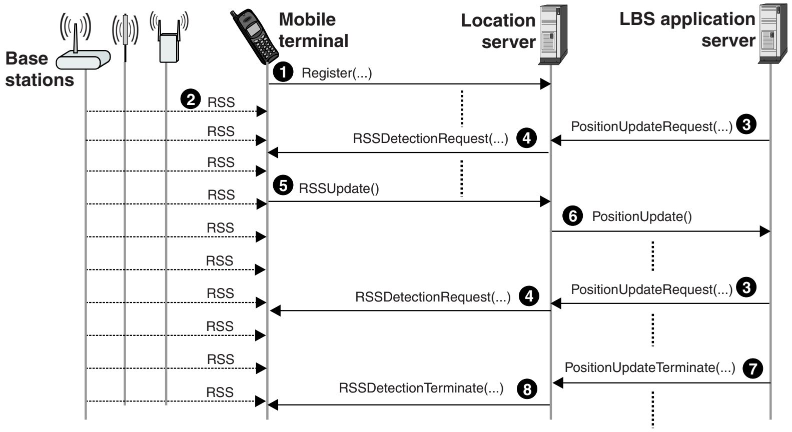


Fig. 1. Proposed Tracking Protocol

privacy due to the reduced amount of collected data. However, the problem of carrying and synchronizing the database is avoided. The main challenge associated with the new approach is to translate geographical zones into RSS-based representations. The next section explores several methods for that.

3 Detection Methods

This section presents several methods for implementing the proposed procedure. In order to be executable on resource-constrained terminals, space and computational requirements are kept as low as possible. Therefore, the methods mainly constitute simplifications of classical LF techniques. They are defined in terms of cell-based localization, i.e. locations are represented as cells. A cell may correspond to a room or a part of it, or a section of a hallway.

The following definitions are needed:

- $C = \{c_1, ..., c_n\}$ is a finite set of *cells* covered by the location system.
- $Z = \{c_a, ..., c_b\}$ is a subset of C that corresponds to an *update zone*.
- A finite observation space $O = \{o_1, ..., o_m\}$ is assumed, with each *observation* o_i being a pair of a *base station* b and a *measured RSS value* $v \in V = \{v_{min}, ..., v_{max}\}$ according to a discrete value range.
- A *sample* s is a set of same-time same-place observations, one for each visible base station.
- A *fingerprint* f is a set of samples collected within the same cell.

3.1 Common Base Stations

A simple detection method, which does not even consider RSS values, is to inspect the base stations occurring in the samples taken by the terminal and compare them with those found in the fingerprints for the cells of the update zone Z. If the number of common base stations $n_\cap$ exceeds a certain threshold, the terminal is assumed to be within Z.

3.2 Ranking

A possible improvement can be achieved by ranking common base stations according to their RSS values. Instead of considering the whole update zone at once, for each fingerprint within Z, the common base stations' ranking is compared to their ranking in the terminal's samples. The comparison is done using the spearman rank-order correlation coefficient as proposed by [7]. If for any of the fingerprints a certain threshold is exceeded, the mobile terminal is assumed to be within the zone.

3.3 Manhattan Distance

A common *deterministic* method in LF systems calculates the Euclidian distance in RSS space between a terminal's measured samples and the fingerprints in the database [8]. A simplified version can be applied for the envisioned zone detection: First, instead of the Euclidian distance, using the Manhattan distance as proposed by [9] comes with less computational overhead. Second, current LF systems compare the distances of a measured sample to all collected fingerprints and yield as a result the location associated with the minimum distance. However, in our approach this would require the whole fingerprinting database to be available at the terminal. As an alternative fixed distance thresholds are proposed, one associated with each fingerprint of Z. The thresholds are independent of the remaining fingerprints in the database and are based merely on the experienced deviations in a cell. The standard deviations σ_{c_i,b_j} of the RSS values experienced in cell c_i regarding all visible base stations $b_j \in B_{c_i}$ can be easily derived from a cell's fingerprint. Upon the deviations, for each cell contained in the update zone a distance threshold T_{c_i} is calculated as follows:

$$T_{c_i} = \sum_{b_j \in B_{c_i}} \sigma_{c_i,b_j} \tag{1}$$

T_{c_i} is computed for each cell c_i of Z. Also for each cell, the means μ_{c_i,b_j} of the base station's RSS values are provided. Thus, at the terminal for each cell $c_i \in Z$ the Manhattan distance $manDist(c_i)$ is calculated based on the means of the measured RSS values m_{b_j}, with b_j being in the set of base stations B_o observed by the terminal, as follows:

$$manDist(c_i) = \sum_{b_j \in B_o \cap B_{c_i}} |m_{b_j} - \mu_{c_i,b_j}| \tag{2}$$

A mobile terminal is estimated to be within Z, if and only if at least one of the cells $c_i \in Z$ satisfies the Manhattan distance: $manDist(c_i) < T_{c_i}$.

A problem of the ranking method and the one based on Manhattan distance is that often the terminal's samples and the fingerprints only have a few base stations in common. As a possible solution, both methods detect a terminal to be out of a cell, if there are less than three base stations in common.

3.4 Bayes Estimator

Several LF systems use Bayesian estimation [10,11,12], which represents a *probabilistic* method. In simple terms, for each cell in the system a probability is calculated based on the current samples taken by the terminal. The cell associated with the highest probability is picked to be the current one of the terminal. In the following the method is adapted for zone detection by collapsing the underlying probabilistic model to a simpler one:

Instead of testing one hypothesis for each cell in the system, only two hypothesis are tested: H_0 states that the terminal is located within the zone, while hypothesis H_1 states that it is located out of it[1]. The probability vector $\vec{\pi}$ describes the probabilities of these two hypotheses being true, defined as follows:

$$\vec{\pi} = \begin{bmatrix} P(H_0) \\ P(H_1) \end{bmatrix} \tag{3}$$

To estimate the probabilities of the two hypotheses, a Bayes estimator is used. The estimator calculates a probability vector $\vec{\pi}$ based on a previous probability vector $\vec{\pi}^{'}$ and a measurement which corresponds to an element o_j in the finite observation space. Initially, both entries of $\vec{\pi}^{'}$ have the same probability. Then, $\vec{\pi}$ is continuously updated by the following equation, where $P(o_j|H_i)$ is looked up in the simple model provided by the location server:

$$\vec{\pi}_i = \frac{P(o_j|H_i)\vec{\pi}_i^{'}}{P(o_j|H_0)\vec{\pi}_0^{'} + P(o_j|H_1)\vec{\pi}_1^{'}} \tag{4}$$

The simple model is created as follows: The probabilities $P(o_j|H_0)$ are calculated based on a set of fingerprints taken from cells in the zone. In turn, the probabilities $P(o_j|H_1)$ are calculated based on a set of fingerprints of cells not in the zone. For that the histogram method [10] is used.

In addition to the Bayes estimator, a simple Markov model is used to guard the transitions of the detector over different time steps. Thus, in a new time step $t+1$, $\vec{\pi}^{t+1}$ is calculated based on the previous estimate $\vec{\pi}^t$ at time t as follows:

$$\vec{\pi}^{t+1} = A\vec{\pi}^t \tag{5}$$

where the Markov model A is defined as follows:

$$A = \begin{bmatrix} P_s & P_{ch} \\ P_{ch} & P_s \end{bmatrix} \tag{6}$$

---

[1] Two hypotheses are used to ease notion instead of one hypothesis and the negation.

P_s is the probability of sustaining the same hypothesis and P_{ch} is the probability of changing to another hypothesis. The probabilities could be defined based on the sizes of the zones or the expected movement behavior of the mobile terminals.

4 Evaluation

In this section evaluation results are presented for the proposed detection methods concerning their accuracy and efficiency. The results have been achieved based on collected IEEE 802.11 RSS measurements. Two scenarios are considered. One concerns the accuracy of the methods and is based on correctly recognizing the entering and exiting of single update zones randomly placed in an indoor environment. The methods' efficiency is evaluated in the second scenario, where a terminal is continuously tracked while moving around in the same indoor environment, i.e., whenever the terminal notifies the server about leaving an update zone, it is configured with a neighboring one. In addition to these simulative evaluations, an analysis of the computational and space requirements for each of the proposed methods is given. As a benchmark for comparison, a reference strategy based on terminal-assisted LF with periodic RSS reporting according to [12] was used.

All observations used in the evaluation were collected in an 802.11 infrastructure with 22 reachable base stations by a laptop with an Orinoco Silver 802.11 card. The evaluation does not address the issue that different 802.11 cards may measure RSS values differently. However, a possible solution that could be applied for the Manhattan and the Bayes detector is proposed in [13]. The Common Base Station and the Ranking detector are already designed to overcome the problem, compare [7]. Samples underlying the fingerprints as well as those for the terminal's localization were taken at 1 Hz. The set of fingerprints covers 63 cells in an office building, compare Figure 2. The building was broken up into cells with an average size of 16 m^2 matching rooms or parts of hallways. Each fingerprint consists of 60 seconds of samples collected by a person walking around in the fingerprinted cell. The observations taken for the localization were collected during 5 walks, totaling 34 minutes. They were taken on different days along different routes as shown in Figure 2. The framework for taking the samples is partly based on software by the Placelab project [3].

4.1 Accuracy

To assess the detectors' accuracy, each of them was tested by 50 different circular zones placed randomly in each of the 5 walks, yielding a total of $5 \times 50 = 250$ tested zones per detector. The circle radii were randomly selected between 4-10 meters.

The parameters used by the detectors in the evaluation were chosen based on the results of a number of initial experiments. For the common base station detector the threshold for being in a zone was set to 70% overlap. For the ranking

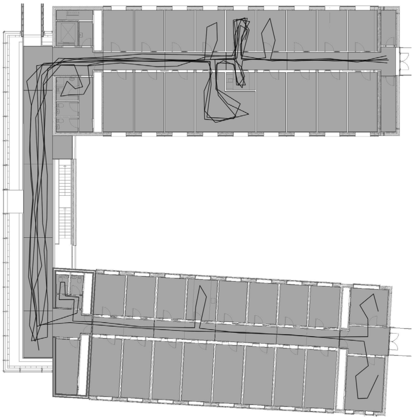


Fig. 2. Layout of sampled area, covered by 63 cells

detector a threshold of 0.9 for the spearman rank-order correlation coefficient was used. For the Bayes estimator detector the probabilities for the Markov model were set to $P_s = 99\%$ and $P_{ch} = 1\%$.

The detectors' accuracies are compared at a time frame level, with each frame being one second long. Therefore, the three measures: *sensitivity*, *specificity* and *global accuracy* are calculated as described below. The calculations are based on the following metrics: TP (true positives) equals the number of time frames the terminal stays in a zone and correctly detects to do so. FP (false positives) is the number of time frames the terminal does not stay in a zone, yet wrongly a zone-containment is detected, TN (true negatives) is the number of frames out of the zone correctly documented by a detector. Finally, FN (false negatives) equals the number of frames spent within the zone, but falsely assumed to be out of the zone. The *sensitivity* is then defined as $Sn = TP/(TP + FN)$. The *specificity* is defined as $Sp = TN/(TN+FP)$. Neither Sn nor Sp alone constitute a good measure of global accuracy. For calculating *global accuracy* the *correlation coefficient (CC)* is used, a well-known mathematical concept which is normally used for mapping two random variables onto one and which has been applied in gene prediction [14] for combining specificity and sensitivity. This application of the CC is adopted in this work and thus the global accuracy quantifies how much the sensitivity and the specificity agree about a detector's performance:

$$CC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TN + FN) \cdot (TP + FN) \cdot (TN + FP)}} \tag{7}$$

All three measures take their values between 0 and 100 percent, where values close to 100 indicate good detection accuracy.

The first evaluation assumes that the terminal provides the detector with single samples as an input value, corresponding to a sampling time of one second, compare Figure 3. The results show that the common base station detector and the ranking detector are the least accurate detectors with a global accuracy of 24.55% and 56.54% respectively. The ranking detector performs better than the common base station detector, which indicates that taking the ranking of the RSS measurements into account gives a gain in accuracy. The low sensitivity of the common base station detector shows that the low global accuracy is caused by a tendency to not detect zone presence. The Manhattan distance detector yields a global accuracy of 60.73%. The most accurate of the detectors is the Bayes estimator detector with a global accuracy of 85.96%. The reason may be its detailed model for representing RSS values. In comparison, the reference strategy yields a global accuracy of 90.12%, which is only slightly better than the Bayes detector.

Evaluations were also run based on longer sampling times at the terminal-side, compare Figure 4. For the ranking and the Manhattan distance detector multiple samples taken for each base station were aggregated to their mean value. The evaluation shows that the accuracy of the common base stations and Manhattan distance detectors increases to respectively 41.35% and 68.96% with five samples. The accuracy of the ranking detector, the Bayes estimator detector and reference system only increase with a small gain to respectively 57.06%, 86.55%, and 92.28% with five samples. Again, the Bayes estimator is the best of the detectors, even when using single samples. Such short sampling times are desirable in order to increase the responsiveness of the system.

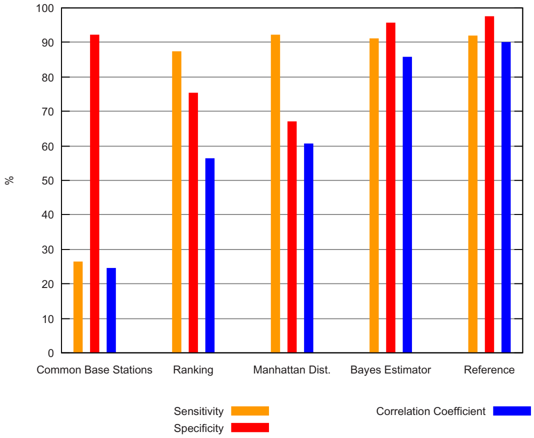


Fig. 3. Results for a single sample

It was also important to evaluate whether the proposed detectors could handle zones of different shapes and sizes. Therefore, simulations based on five different shapes of approximately equal sizes were conducted. The evaluated shapes were circles, squares, annuli, holed-squares and polygons with between 4 to 8 edges. Figure 5 shows the obtained results, which indicate that all detectors

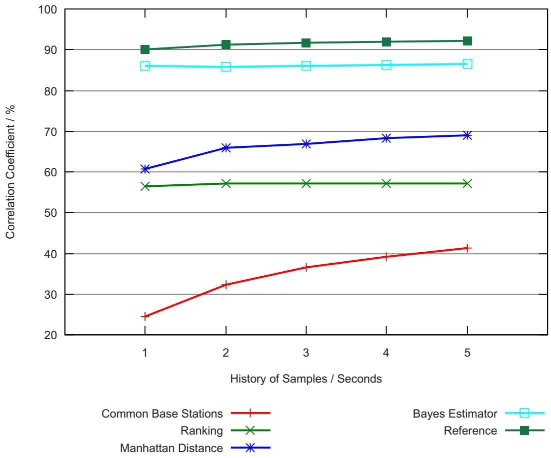


Fig. 4. Results for increasing sampling times

perform best with closed shapes, however, with little accuracy losses for the more irregular-shaped polygons. For both of the holed shapes there is about a 10% decrease, showing that the detectors are still able to handle such complex zones. The results of the ranking detector differ from these trends as they indicate a better support for polygon-shaped zones.

To evaluate the impact of the size of the shapes, evaluations were run with circle-shaped zones of different radii. The results are shown in Figure 6. It can be seen that all the detectors' accuracy drops for very small zones, primarily because the detectors have very little fingerprinting data to base their estimates on. One can also see that the threshold selected for the ranking detector is not optimal for larger zones. All detectors, however, experience a decrease in accuracy for radii above 20 meters. This fact can be attributed to the detectors being pessimistic, that is, they prefer estimating a terminal to be out of a zone over being contained in it. The pessimism shows up as an increase in errors when more and more space of the evaluated walks is covered by a zone. The collected data did not enable us to correctly evaluate circle-shaped zones with radii above 24 meters, because in this case more than 70% of the time frames of the walks would be contained by the zone. Based on the accuracy evaluations it can be concluded that the Bayes estimator detector is the most accurate and robust of the proposed detectors.

4.2 Efficiency

To evaluate the efficiency of the proposed protocols and detectors, another evaluation simulating the continuous tracking of a terminal has been carried out. The evaluation is based on the same collected walks as before and a simple tracking protocol: First, a circle-shaped zone detector of 10 meter radius is set up with its center located at the starting cell of the walk. When the detector reports

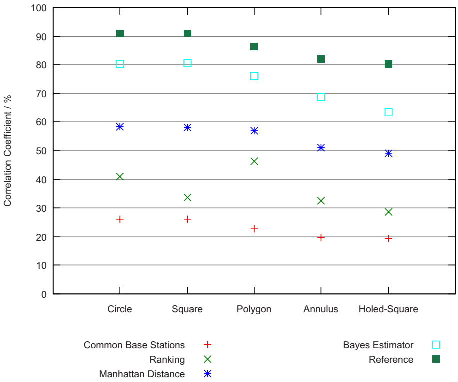


Fig. 5. Results for different zone shapes

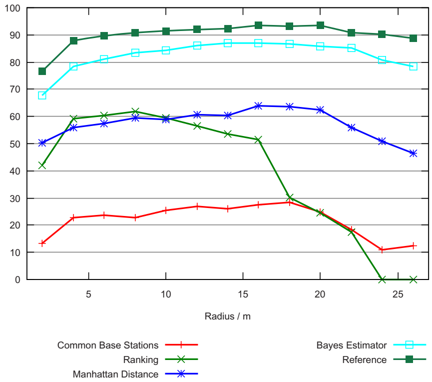


Fig. 6. Results for different zone sizes

that the terminal has moved out of the zone, a second detector is set up with a new zone, now with the just-estimated location being its center. This process is repeated until the end of the walk. To be able to use the same collected walk data several times each evaluation is run several times with the first five different locations in the collected walks as starting points. During the evaluation the following statistics are collected: the *correctly saved updates*, which count the time frames when the detector correctly estimates that it is in a zone and therefore an RSS update is avoided; the *wrongly saved updates*, which count the frames where the detector wrongly estimates that it is in a zone and therefore does not send an RSS update; and the *RSS updates*, which are actually sent when the detector has estimated that the terminal may have moved out of the current zone. The used walks in the evaluation actually represent a worse-than-average scenario, because the terminal is moving most of the time. In a scenario with a more static movement pattern a larger number of RSS updates would be saved.

The results show that for all of the detectors the number of RSS updates is considerably lowered in comparison to the 9572 RSS updates produced by secondwise RSS reporting, which was assumed for the reference system, compare Figure 7. The common base stations (CBS) detector, the ranking detector, and the Manhattan distance (MD) produce the most updates with respectively 2721, 693, and 803 RSS updates. The RSS updates produced by the Bayes estimator (BE) detector is 192 which is close to the efficiency of a perfect detector, which would produce 114 RSS updates. The Bayes estimator shows the fewest RSS updates but generates more wrongly saved updates than the Manhattan distance detector respectively 423 and 89. However, the detectors' performance can be fine tuned by changing some of the parameters. For instance, wrongly saved updates can be traded for generating a few excessive RSS updates, which in turn can be filtered out at the location server, thus ensuring better overall accuracy. In summary, considering all three metrics the Bayes estimator detector is the best choice.

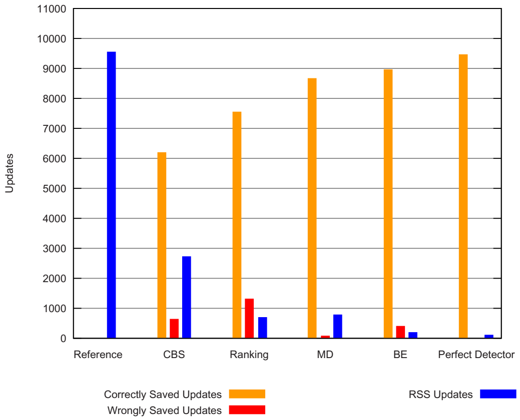


Fig. 7. Efficiency evaluation results

4.3 Space and Computation Analysis

In this section the space and computation requirements of the different detectors are analyzed. The analysis is based on the following parameters: M is the number of observations provided by the terminal to the detector; B_{zone} is the number of base stations visible from cells in the zone; B_{all} is the number of all base station covered by the system; Z is the number of cells in the zone; V is the number of possible RSS values. For each of the detectors the results of the analysis are given in Table 1.

Table 1. Space and computational requirements on mobile terminals

Detector	Computations	Space
Common Base Stations	$O(M)$	$O(B_{zone})$
Ranking	$O(M + Z \times B_{zone} \times \log(B_{zone}))$	$O(B_{zone} \times Z)$
Manhattan Distance	$O(M + B_{zone} \times Z)$	$O(B_{zone} \times Z)$
Bayes Estimator	$O(M)$	$O(B_{all} \times V)$
Reference System	$O(1)$	$O(1)$

The computation and space requirements are low for both the common base stations detector and the reference system, the latter because it does not perform any extra calculations or use any additional space on the mobile terminal. The ranking detector has higher space requirements and computation requirements, because it needs to sort the measurements and also store the calculated rankings for each cell in the zone. The Manhattan distance detector has lower computation but the same space requirements. Computations are needed for calculating the Manhattan distances to all cells in the zone and each distance computation considers all base stations visible in the zone. Its space use is attributed to

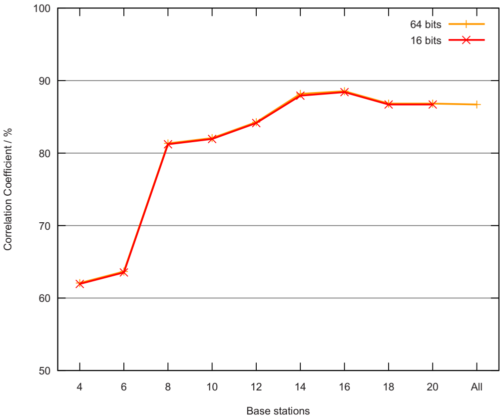


Fig. 8. Effect of the number of base stations on the accuracy of the Bayes estimator

storing mean values for all cells in the zone. The Bayes estimator detector has low computation requirements, but the highest space requirements because it needs to store the simple probabilistic model.

To further reduce the space consumption of the Bayes estimator three techniques are proposed. First, a lossless compression technique for representing repeated entries is applied, which just counts repetitions of the same values. Because 802.11 RSS measurements in practice only span a small range of V and because the entries are generated using the histogram method, the entries contain a lot of repetitions. Second, the representation of the entries is constrained to only 16 bits. Third, the number of base stations used for the entries can be reduced.

For example, without these techniques the space consumption of the detector on the collected data, with $V = 255$, $B_{all} = 47$, two hypotheses, and a 64 bit representation of probabilities, the memory needed for representing one zone would be $2 \times 47 \times 255 \times 8b = 95,9Kb$. However, when the first two techniques are applied and all base stations are kept, the data can be compressed to $1Kb$. If the number of base stations is also reduced to a maximum of 12, even $0.5Kb$ are possible. Both values seem fairly acceptable.

To learn whether the reduction of base stations and bit representation negatively affects the accuracy of the Bayes detector, an extra accuracy evaluation was run and the results are shown in Figure 8. They indicate that the reductions do not have a major impact on the accuracy, as long as the maximum number of base stations is not limited to fewer than 8. However, this number is only valid for the zone sizes used in the evaluation because for larger zones more base stations might be needed for a whole zone to be covered.

To subsume, the Bayes estimator turns out as the best of the presented methods for all considered aspects: accuracy, responsiveness, support for different sizes and shapes, as well as efficiency. With respect to the reference system, it yields a comparable accuracy, while the number of exchanged messages is strongly reduced. As discussed, the little lack of accuracy can be counterbalanced by slightly reducing the number of saved update messages.

5 Related Work

Related work presented in the following is restricted to RSS fingerprinting and divided into infrastructure-based and infrastructure-less systems.

5.1 Infrastructure-Based

One of the first infrastructure-based systems was RADAR [8], that applied different deterministic mathematical models to calculate the position (in coordinates) of a terminal based on IEEE 802.11 measurements. Similar methods have also been applied to GSM [15]. The mathematical models used had to be calibrated for each site where the systems had to be used. In comparison to RADAR, later systems have used probabilistic models instead of deterministic models. This is because a good deterministic model for the volatile radio environment has not been found. As in the case of the deterministic models in RADAR, the probabilistic models are calibrated for each site. Examples of systems, which determine the coordinates of a terminal, are published in [10,11,16]. Systems determining the logical position or cell of a terminal are published in [12,17]. From a perspective of resource-constrained terminals, existing systems are not optimal with respect to the overhead induced by using poll or periodic update protocols only, as discussed in Section 2. However, from an accuracy perspective the proposed zone updating protocol has the drawback that history tracking algorithms cannot be applied to improve LF accuracy. A possible solution is to report RSS values sampled over the last n seconds whenever a zone update is due. This way, a possible historical analysis and the decision whether the update is really in the zone or not could still be done at the server-side.

In addition to the above systems, which estimate the location of terminals, a number of systems, such as [7], have been studied where the calibration step is only carried out by users for tagging relevant places. The systems propose simple metrics based on signal strength measurements to quantify when terminals are in proximity of calibrated places.One of the strengths of these simple metrics is that they overcome the problem of 802.11 cards returning different RSS values. Such systems are relevant to this work with respect to the methods they propose for proximity detection. However, such systems can only detect presence at a single point and not within zones with specific shapes and sizes, as addressed in this paper.

A system which has addressed, by using additional sensors, the needs of resource-constrained terminals when used with fingerprinting-based indoor location systems is [18]. They propose a communication protocol between the location server and the terminal, which dynamically adapts the RSS update rate of the terminal based on the distance to the last reported update using measurements from an accelerometer. In comparison, the methods proposed in this paper do not require any extra sensors and are therefore usable for a broader range of terminals where such extra sensors are not present or too expensive to include. In addition to this, the proposed methods in this paper can also be used with arbitrary shaped zones and not just zones defined by a distance to a specific point.

Thus, in comparison to existing infrastructure-based solutions the proposed approach represents an improvement, because it enables efficient tracking and accurate zone detection based on RSS measurements only.

5.2 Infrastructure-Less

Most infrastructure-less systems are based on protocols which are more energy-efficient than for instance IEEE 802.11, such as IEEE 802.15.4 or communication over the 433/916 MHz bands reserved for telemetry. In [19] a system is presented which senses the proximity of a mobile node to static beacon nodes which output their id and position. The position of the mobile node is then estimated by finding the centroid of the positions of the proximate beacon nodes. A system that proposes methods for infrastructure-less localization inspired by infrastructure-based techniques is MoteTrack [9]. The system consists of a number of wireless sensor network nodes where some have the role as static beacon nodes and other are mobile nodes which the system should locate. The system is based on location fingerprinting using RSS to the static beacon nodes. The fingerprints are stored distributely over the static beacon nodes and provided to the mobile nodes when in proximity. The system's method for location estimation is based on weighted nearest fingerprints based on the Manhattan distance instead of the Euclidian distance to lower computation needs. The computing of the location estimates can be carried out by either the mobile nodes or the beacon nodes, depending on which of the proposed sharing techniques is used. These systems are related to the proposed methods in terms of how they achieve energy-efficiency and do decentralized estimation. However, because all such systems assume that there is no infrastructure, they do not address how to combine decentralized estimation with the capabilities of infrastructure-based solutions.

6 Conclusion and Further Work

The paper proposed the novel approach of zone-based RSS reporting for location fingerprinting, where the terminal is dynamically configured with RSS-based representations of geographical update zones. Only when the terminal detects a match to the RSS patterns, it reports its measurements to the server. Several methods for realizing zone-based RSS reporting were proposed and profoundly compared. As it turned out, an adaption of classical Bayes estimation is a promising approach, which, in comparison to the assumed reference system, strongly reduces message overhead while yielding a high accuracy and responsiveness. Given the mechanisms described in this paper, existing approaches for efficiently realizing proactive LBSs – which, so far, assume terminal-based positioning like GPS – can be easily applied to LF systems. This concerns not only single-target LBSs, but also proactive multi-target LBSs, compare [4]. Two further issues subject to future work are discussed in the following.

First, with some technologies, such as IEEE 802.11, already the RSS scanning is rather resource consuming, which makes it desirable to minimize the needed scans. One possible method, which, however, only applies to big zones, is to subdivide a zone in a way that in the central part of it a long scanning interval is used, while short intervals are applied at the borders of the zone. Another method is using an moving-versus-still estimator based on RSS measurements, such as the one proposed in [16], to estimate whether the terminal is moving or not, and then adapt the scanning intervals to this information. However, the proposed estimator is rather expensive in terms of needed samples and computations, so a scaled-down version would have to be developed.

A second issue this work has not addressed is how the building layout in terms of floors affects the detection methods. LF techniques evaluated for both GSM and 802.11 in [15] have shown good performance, at least in office-like buildings, for estimating the floor level. So, at least for the Manhattan distance detector and the Bayes estimator, floor errors should not be a major issue. The presented detectors also allow zones to be defined over several floors.

Acknowledgments. We appreciate the comments, advice, and insights of our reviewers and especially our shepherd John Krumm. We thank Carsten Valdemar Munk for helping collecting signal strength measurements. M. B. Kjærgaard is partially funded by the software part of the ISIS Katrinebjerg competency center http://www.isis.alexandra.dk/software/.

References

1. Küpper, A., Treu, G.: Efficient Proximity and Separation Detection among Mobile Targets for Supporting Location-based Community Services. ACM SIGMOBILE Mobile Computing and Communications Review **10**(3) (2006) 1–12
2. Küpper, A.: Location–based Services — Fundamentals and Operation. John Wiley & Sons (2005)
3. LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, J., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, P., Borriello, G., Schilit, B.: Place Lab: Device Positioning Using Radio Beacons in the Wild. In: Proceedings of the Third International Conference on Pervasive Computing. (2005)
4. Küpper, A., Treu, G., Linnhoff-Popien, C.: TraX: A Device-centric Middleware Framework for Location-based Services. IEEE Communications Magazine **44**(9) (2006) 114–120
5. Wolfson, O., Sistla, A.P., Chamberlain, S., Yesha, Y.: Updating and Querying Databases that Track Mobile Units. Distributed and Parallel Databases **7**(3) (1999) 257–387
6. Leonhardi, A., Nicu, C., Rothermel, K.: A Map-Based Dead-Reckoning Protocol for Updating Location Information. In: Proceedings of the 16th International Parallel and Distributed Processing Symposium. (2002)
7. Krumm, J., Hinckley, K.: The NearMe Wireless Proximity Server. In: Proceedings of the Sixth International Conference on Ubiquitous Computing. (2004)

8. Bahl, P., Padmanabhan, V.N.: RADAR: An In-Building RF-based User Location and Tracking System. In: Proceedings of IEEE INFOCOM. (2000)
9. Lorincz, K., Welsh, M.: MoteTrack: A Robust, Decentralized Approach to RF-Based Location Tracking. In: Proceedings of the First International Workshop on Location and Context-Awareness. (2005)
10. T.Roos, P.Myllymäki, H.Tirri, P.Misikangas, J.Sievänen: A Probabilistic Approach to WLAN User Location Estimation. International Journal of Wireless Information Networks **9**(3) (2002) 155–164
11. Youssef, M., Agrawala, A.: The Horus WLAN Location Determination System. In: Proceedings of the Third International Conference on Mobile Systems, Applications, and Services. (2005)
12. Haeberlen, A., Flannery, E., Ladd, A.M., Rudys, A., Wallach, D.S., Kavraki, L.E.: Practical Robust Localization over Large-Scale 802.11 Wireless Networks. In: Proceedings of the Tenth ACM International Conference on Mobile Computing and Networking. (2004)
13. Kjærgaard, M.B.: Automatic Mitigation of Sensor Variations for Signal Strength Based Location Systems. In: Second International Workshop on Location- and Context-Awareness. (2006)
14. Burset, M., Guigó, R.: Evaluation of Gene Structure Prediction Programs. Genomics **34** (1996) 353–367
15. Otsason, V., Varshavsky, A., Marca, A.L., de Lara, E.: Accurate GSM Indoor Localization. In: Proceedings of the Seventh International Conference on Ubiquitous Computing. (2005)
16. Krumm, J., Horvitz, E.: LOCADIO: Inferring Motion and Location from Wi-Fi Signal Strengths. In: Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services. (2004)
17. Castro, P., Chiu, P., Kremenek, T., Muntz, R.: A Probabilistic Room Location Service for Wireless Networked Environments. In: Proceedings of the Third Internal Conference on Ubiquitous Computing. (2001)
18. You, C.W., Chen, Y.C., Chiang, J.R., Huang, P., Chu, H.H., Lau, S.Y.: Sensor-Enhanced Mobility Prediction for Energy-Efficient Localization. In: Proceedings of Third Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks. (2006)
19. Bulusu, N., Heidemann, J., Estrin, D.: GPS-less Low Cost Outdoor Localization For Very Small Devices. IEEE Personal Communications Magazine **7**(5) (2000) 28–34

TrackSense: Infrastructure Free Precise Indoor Positioning Using Projected Patterns

Moritz Köhler[1], Shwetak N. Patel[2], Jay W. Summet[2], Erich P. Stuntebeck[2],
and Gregory D. Abowd[2]

[1] Institute for Pervasive Computing, Department of Computer Science
ETH Zurich, 8092 Zurich, Switzerland
koehler@inf.ethz.ch
[2] College of Computing & GVU Center
Georgia Institute of Technology
85 5th Street NW, Atlanta GA 30332-0280 USA
{shwetak, summetj, eps, abowd}@cc.gatech.edu

Abstract. While commercial solutions for precise indoor positioning exist, they are costly and require installation of additional infrastructure, which limits opportunities for widespread adoption. Inspired by robotics techniques of Simultaneous Localization and Mapping (SLAM) and computer vision approaches using structured light patterns, we propose a self-contained solution to precise indoor positioning that requires no additional environmental infrastructure. Evaluation of our prototype, called TrackSense, indicates that such a system can deliver up to 4 cm accuracy with 3 cm precision in rooms up to five meters squared, as well as 2 degree accuracy and 1 degree precision on orientation. We explain the design and performance characteristics of our prototype and demonstrate a feasible miniaturization that supports applications that require a single device localizing itself in a space. We also discuss extensions to locate multiple devices and limitations of this approach.

1 Introduction and Motivation

We introduce a solution to indoor localization, TrackSense, that requires no additional infrastructure in the environment and provides 3D positioning and orientation data that performs well against existing research and commercial solutions. Although we have seen great progress toward the goal of indoor localization, almost all of the solutions that offer precise (few centimeter) indoor localization have been limited to techniques that require the introduction of new infrastructure to the physical space (*e.g.* cameras or beacons). These solutions are often costly and typically require time-consuming installations, and it is not easy to move the instrumentation from one space to another. Although existing commercial positioning systems are adequate for prototyping user experiences, their ultimate success relies on a localization approach that is inexpensive and easily deployed. TrackSense is appropriate for situations where the localized device has a clear view of the walls and ceilings. By centralizing all computation to a single, small device, we reduce cost and substantially increase the number of places the localized device can be used.

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 334 – 350, 2007.

In addition to the inherent technical challenges, there are several motivating applications in which a single computational device benefits from precise location. Patel *et al.* demonstrate a see-through augmented reality, handheld device capable of performing precise at-a-distance interaction [20]. Their iCam device provided simple authoring and retrieval of digital content attached to physical objects, as well as manipulation of digital content in an augmented reality game. The iCam relied on a commercial ultra-wideband positioning system for localizing the handheld. Cao and Balakrishnan demonstrated the use of a handheld projector for viewing and interacting with multiple dynamically defined information spaces projected in the physical space [5]. Their application used a commercial camera-based motion capture system to determine the pose and position of the projector. In addition to these research prototypes, many examples of augmented reality rely on precise tracking of an object (such as an individual's head) and these applications would be improved by any solution that would speed deployment in multiple spaces.

TrackSense determines its distance and orientation to fixed large planes in a space (*i.e.*, walls and ceilings) and uses that information to calculate its 3D position and pose in the room. Inspired by robotics localization and camera-projector research, our solution uses a camera to locate and track a grid pattern projected onto surfaces in the camera's field of view. This solution is more accurate and reliable than standard computer vision feature extraction techniques, because the exact feature (the grid pattern) is known and ever-present in the camera's view. It also provides a useful complement to traditional stereo vision, which does not perform well on plain surfaces. In addition, our technique provides information regarding its pose that is not available with standard ultrasonic or laser range finding solutions. Combining our solution with less precise room-level positioning systems we can provide localization within an entire world coordinate frame. Our current prototype is bulky and only demonstrates localization of a subsection of any given room. However, we also describe a miniaturized system that can be extended to an entire room.

2 Related Work

Indoor location technologies have been a long-studied topic in pervasive and ubiquitous computing. Hightower and Borriello provide an overview of the various location technologies and techniques [14]. The two basic approaches are to build the entire infrastructure from the ground up (*e.g.*, Ultra-wideband [35], ActiveBadge [37], Cricket [25], Vicon [36], NorthStar [21] and Active Bat[1]) or to leverage existing infrastructure that can yield localization, either through triangulation or fingerprinting (*e.g.*, 802.11 work such as RADAR [2] and Place Lab[15], GSM Cell Towers [22], Bluetooth [19], and powerlines [24]). Typically, solutions that offer precise indoor localization of a few centimeters use the first approach of installing new environmental infrastructure that is both expensive and hard to move, thus limiting location-based applications to a few highly specialized environments. Although researchers are exploring ways to leverage existing public infrastructure, the solutions are currently limited to resolutions of a few meters (room-level).

The robotics community has a long history of exploring ways to localize autonomous robots without having to install custom infrastructure or gather *a priori* topological knowledge of the environment. Researchers have extensively studied the use of highly precise laser or ultrasonic range finders to automatically construct a feature map of the environment and then later consult it for localization [9, 18, 34]. This class of techniques is called Simultaneous Localization and Mapping (SLAM). A visual variant of SLAM, visual SLAM (vSLAM), builds a map entirely using vision [7, 30]. SLAM solutions typically employ various statistical and probabilistic models for localization. In addition, SLAM is a recursive process that evolves over time to improve accuracy and address changes in the environment. Although inspired by robotics, our solution does not rely on a statistical model or the construction of a complete map of the environment.

Vision-based techniques extract features from the physical environment, such as detecting planar surfaces for 3D model extraction [3, 6, 8, 17, 27, 33]. One limitation of purely vision-based techniques is the requirement of easily discernible and static features in the environment. Features many not always be available, such as on single-colored or plain walls. Additionally, lighting conditions may change the way features appear at different times. In our solution, the features (projected grid) are placed artificially in the environment to ease feature extraction. Our solution works best on the plain surfaces on which other computer vision approaches such as stereo vision fail. However, stereo vision techniques would be complimentary when the device is used in more textured spaces.

Our approach with TrackSense is similar to previous vision work using structured light to extract physical feature information from an object, which use projected coded patterns of light at an object to extract the 3D features of that object [3, 29]. Other research has used the detection of structured light on a planar surface to automate projector calibration [31, 32]. These solutions temporally encode different structured light patterns; we focus on a static pattern produced by a laser to ensure a small, low-cost solid-state solution.

Finally, augmented reality researchers have explored using fiducials (such as barcodes or 2-dimensional glyphs) to determine distance and pose to labeled objects and surfaces [12, 28]. However, large glyphs are needed for long distances and a number of them must be placed in the environment to cover a large space. In addition, glyphs are not always aesthetically pleasing unless they are blended with the décor of the environment. Our approach can be made invisible to the user by using infrared lasers.

3 System and Implementation Details

TrackSense projects a grid pattern into the environment to locate planes (walls) and intersections (corners). By detecting three orthogonal planes (two walls and a ceiling or floor), the system can recover its position and orientation with respect to that corner. By using a 3-axis accelerometer and magnetometer (compass), the unit can determine which corner of a room it is looking at, and hence, its position and orientation with respect to the room's coordinate frame. Practically, several TrackSense units (2-5) angled in different directions can cooperatively identify three planes within their combined views. This section discusses the implementation of a single TrackSense unit and how it obtains distance and orientation measurements from one or more planes it observes in the environment.

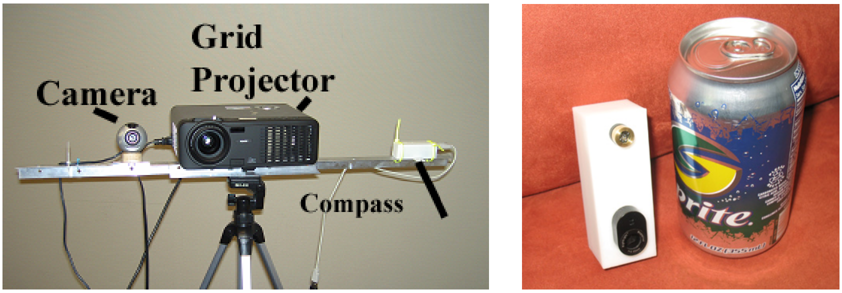


Fig. 1. Left: Operating TrackSense prototype and its components. Right: A miniaturized design prototype with laser diode, camera, and 400 MHz GumStix computer.

3.1 Hardware

TrackSense has a grid projector and a camera (see Figure 1). We used a 2000 Lumen DLP projector, which simulates a grid projecting laser diode. We used a projector to allow for easy prototyping of different sizes and shapes of projected grids. In an actual engineered solution, the relatively large desktop projector would be replaced with a single laser diode and grid diffraction lens, possibly projecting infrared light to make the system's operation imperceptible. For our prototype camera, we used a Logitech QuickCam Pro 4000 USB webcam with VGA (640x480) resolution. Our prototype system also had a custom-built magnetometer for a cost of $50 USD with a resolution of about 2°. In an actual system, several TrackSense units would share a 3-axis accelerometer and magnetometer. For prototyping, these components were connected to a desktop PC running our software, which was written in C++ using Intel's OpenCV and the VXL computer vision libraries.

Projected Grid & Camera Calibration

As with all stereo vision devices, for a TrackSense unit to operate correctly, the grid projector and camera must be calibrated, both with respect to each other and with respect to a ground truth or world coordinate system. By using point correspondences between the grid projector and the camera using a known calibration rig, we can find the Fundamental matrix, F that encodes the relationship between the grid projector and the camera. The Fundamental matrix is defined by the equation $x'^{T} Fx = 0$ for any pair of matching points $x \leftrightarrow x'$ in two images. In other words, if two points x and x' correspond, the equation described above evaluates to 0. Note that mathematically, the grid projector is assumed to be a second camera, with grid line intersections at specific points in the virtual grid "image." We use F to help determine point correspondences as described in Section 3.2.3. By using known world coordinate points on our calibration rig, we can also calculate the projective matrices, P and P`, between the world coordinate system and the camera and virtual grid "camera" that is used to determine the location of detected points with respect to the TrackSense unit. We computed this standard multiple-view calibration with a custom rig using the calibration routines from OpenCV [13]. In actual operation, this calibration would be performed a single time at the time of manufacture.

3.2 System Operation

As the grid is projected onto and reflected from objects in the environment such as the ceiling and walls, the camera detects the lines using a custom edge detection algorithm, Using these detected lines, we can find the location of each grid intersection point. Because we are integrating data from hundreds of pixels for each line, we can develop a mathematical model of the line that is more accurate than any single pixel. Hence, we can measure locations of intersection points with sub-pixel accuracy. By triangulation (using the same math as standard stereo vision), we can find the distance and orientation to each point relative to the camera. Using multiple points, we can detect planes and corners where multiple planes meet. From this, we can recover the orientation and position of the TrackSense unit with respect to the corner. If we use a 3-axis accelerometer and magnetometer to determine which corner we are observing, we can locate the TrackSense unit within the room. Furthermore, if we already can identify the room using a less accurate positioning system, such as GSM fingerprinting [22], the within room position translates to an accurate world position. In this section, we describe elements of this procedure in more detail.

3.2.1 Line Detection

To determine where the grid intersection points occur in the camera image, our system must first detect the projected lines from a potentially noisy image. Using a standard Canny edge detection algorithm would detect the grid, but it would also detect many other lines in the image (*i.e.,* edges of windows, picture frames, desks, pencils, *etc*). One way to enhance the detection of the projected grid would be to take pairs of images, one with the grid turned off, and one with it turned on, and then subtract them to obtain the location of only the grid. However, this reduces the frame rate of the system by one-half, requires precise synchronization between the grid projector and camera, and assumes that the system is not in motion between subsequent frames. Custom hardware operating at extremely high frame rates where these assumptions may be valid could make use of this subtraction technique to greatly simplify the line detection algorithm.

However, our prototype uses a web cam with limited frame rate and no synchronization to the grid projector. We also wanted our system to operate while in motion and be able to provide a new orientation and position with every camera image. To enable this, we developed an enhanced edge-finding algorithm that detects projected grid lines while ignoring many environmental lines. Figure 2 (center) shows results obtained using a standard implementation of the Canny edge detection algorithm [4]. The Canny algorithm looks for gradients in the image, detecting lines for both low-to-high and high-to-low transitions. Naturally occurring lines in the environment (*i.e.,* from a corner or edge) typically only have one of these gradients, either increasing or decreasing, as the edge typically separates objects of different reflectance levels. However, projected lines are typically brighter than the objects they fall upon, leading to both a rising and falling gradient on either side of the line. As shown in Figure 2:

1. For our single luminous projected line, two edges are detected: One for the increasing gradient and one for the decreasing gradient.
2. Each edge of the black square (upper left) results in exactly one detected edge.

We can obtain better results by modifying a gradient-based edge detection algorithm so that a positive gradient followed by a negative gradient of similar magnitude is used to detect a line. This leads to a zero-crossing edge detection algorithm that uses the 1st derivative instead of the 2nd derivative. Figure 2 shows the result of such an algorithm. This helps reduce both the ambiguity problem and the false positives.

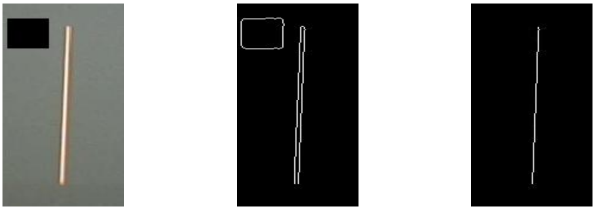


Fig. 2. Left: A line projected onto a wall and a black square representing an object in the environment. Middle: Results obtained applying Canny edge detection to the image on the left. Right: Result obtained applying the gradient based edge detection algorithm we developed.

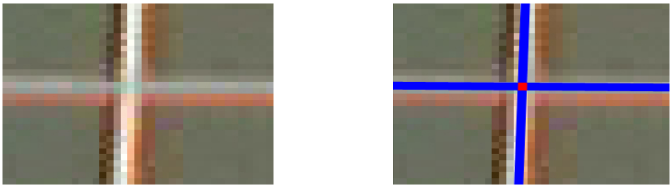


Fig. 3. Left: Zoomed in image of two intersecting lines. Right: The same lines from the left superimposed with lines detected by our line finding algorithm.

3.2.2 Intersection Points

Using the edges detected from the projected grid, a Hough transform can determine the parameters of each line [11]. By mathematically determining the intersection point of each pair of lines, we obtain the position of these points with sub-pixel accuracy. Figure 3 shows an intersection of two projected lines in our camera view, and an overlay of the detected lines (blue) and intersection point (center red dot). Our prototype used a grid of 9 vertical lines and 7 horizontal lines, which corresponded to our 4:3 aspect ratio camera, giving a maximum of 63 detectable points.

3.2.3 Point Correspondences and 3D Reconstruction

An important step of reconstruction is the correct identification of point correspondences between two views. To determine the orientation and distance to any point in the environment, a stereo rig must identify where that point appears in both camera views. Traditional stereo vision algorithms [11] rely on distinctive textures in the pair of images to determine which points from the left camera image corresponds to a particular point in the right camera image. However, we are unable to use a similar method for two reasons. First, our system needs to be able to work on plain walls without features, which lack the texture that traditional stereo vision algorithms rely upon. Second, we are using a projector as a virtual "camera". The advantage of the projector is that our system will work on walls without texture by projecting its own features, but the disadvantage is that the grid is regular and each intersection point looks very much like all the others.[1]

Given a grid intersection point in the projector "view", the correct corresponding point in the camera view must to be found. In order to reduce the search space for this point the epipolar constraint from the Fundamental matrix is applied [13]. Figure 4 shows epipolar lines for our prototype in which the camera and grid projector were mounted horizontally.

To determine point correspondences we use a cost function that is the sum of the squared distance to the epipolar line and the position of that point in the previous image (timestep t-1). Using these cost functions, the Hungarian algorithm [16, 20] minimizes the total cost and produces the best match in correspondences between the grid intersections and the detected line intersections in the camera image. In some cases, intersection points may be missing from the camera image. For example, lying on a dark or textured object or falling outside the view of the camera would prevent detection by the edge and line finding algorithms. To prevent errors in these cases, the cost for questionable points are set to infinite cost (allowing the Hungarian algorithm to skip that point) if the distance to the epipolar line was greater than a pre-set threshold.

Once point correspondences are known, the projective matrices, P and P` obtained in the initial system calibration are used to calculate the 3D position of the point. Linear triangulation is used to obtain the desired 3D position of a point. More details of the linear triangulation approach, along with methods for improving its accuracy can be found in [13].

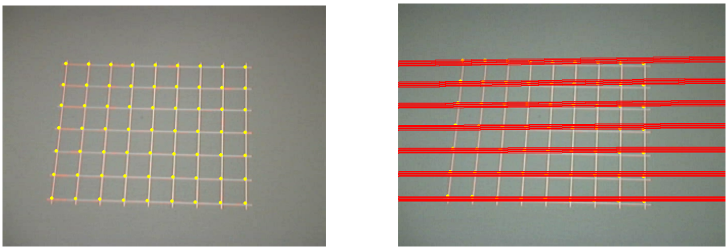


Fig. 4. Left: A projected grid. The yellow dots show points of intersecting lines. Right: Left image superimposed with epipolar lines.

3.2.4 Identifying Planes

A TrackSense unit models walls, ceilings or floors as large planes. After triangulation, we have data points that may represent points on the surface of planes, or may be noise, either from random objects in the environment or measurement errors. In order to develop a robust algorithm that can detect planes correctly, several issues arise.

---

[1] As our prototype uses a data-projector to simulate a laser-grid, we could have used textured patterns (as used in 3D reconstruction using structured light [3, 29]). To work with very small and inexpensive laser diodes, we limited the output of the projector to a uniform static image. With custom diffraction lenses on a laser diode it would be possible to produce a laser pattern that, while static, would not be regular, allowing for easier calculation of point correspondences.

- The exact number of planes in each frame is unknown. Because each TrackSense unit has a finite field of view and operating range, we do not expect to detect more than a maximum of 3 planes of usable size (a corner of two walls and a ceiling or floor).
- It is not known which points lie on the surface of the same plane and form a group.
- A significant amount of points represent noise and have to be classified as outliers so they do not affect the correct computation of a plane.

Our approach uses the RANSAC (RANdom Sample Consensus) algorithm [10]. First, three points which specify a plane are randomly selected, and every remaining point is tested to see if it is close to the candidate plane. We have found that a threshold of 3 cm includes most valid points while eliminating most outliers. After selecting many possible random planes, the one with the largest group of supporting points is chosen. The valid points are then used to compute a least mean square solution for the actual position and orientation of the plane, resulting in better accuracy than any of our single point measurements. Planes without enough supporting points are discarded. The algorithm terminates with failure to detect a plane. Otherwise the previous steps are repeated in order to find the next plane. We have found that with a 7 by 9 grid (63 total points) a threshold of 18 points generally indicates that a valid wall, floor, or ceiling plane has been found. Figure 5 shows a point cloud representing points on the surface of two walls of a corner (left), and the planes that have been fitted to the points using the approach described above (right). Once the four parameters characterizing each plane have been determined, the distance from the TrackSense unit to the plane can be determined geometrically.

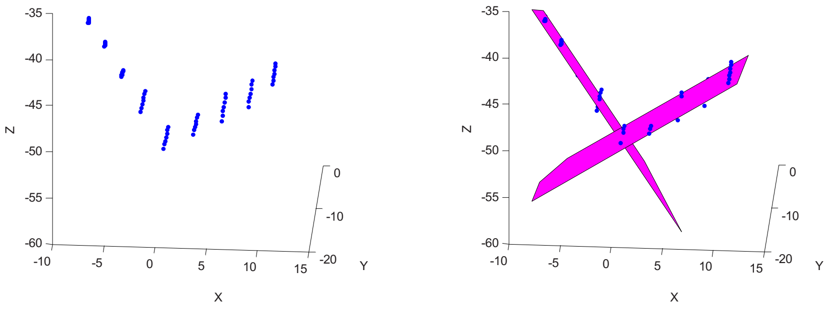


Fig. 5. Left: Point cloud from a two wall corner. Right: Point cloud plus the fitted planes.

3.2.5 Determining Position and Orientation

Depending on the number of walls, we discuss the different strategies for determining the position and orientation of the device.

One Wall: A TrackSense unit that can observe a wall can determine its orientation with respect to that wall, and distance from the wall. With a single wall in view it can act as an enhanced ultrasonic tape measure, being able to calculate the direct distance from the unit to the closest point on the wall. Unlike an ultrasonic tape measure, the

TrackSense unit does *not* have to be pointed directly at the closest point on the wall to make this measurement, as it also knows the orientation of the wall.

Two Walls: By observing two orthogonal walls, a TrackSense unit is able to determine its (X,Y) position with respect to the corner. If it makes the assumption that the two walls are vertical (at 90 degrees to the ground plane) the TrackSense unit can determine its own orientation with respect to the ground plane. Note that the two walls do NOT have to form a 90 degree angle with each other, only with the ground plane.

By making use of a magnetometer a TrackSense unit can determine its global bearing, and determine which corner of the room it is observing, which leads to a global (X,Y) position within the room. Also note that if the room has a different angle at each corner (as opposed to the standard 90 degree corner) the TrackSense unit can measure the angle between each set of two walls and use that to “fingerprint” the corner that it is looking at. Although two walls do not provide a Z (or height) measurement, if the unit is held at a consistent height, or mounted on a mobile base the Z component may be stable. While an ultrasonic transducer could be used pointing towards the ground or ceiling to provide an estimate of Z, we recommend the use of a second TrackSense unit for redundancy.

Three Walls / Corner: By observing three orthogonal planes (such as the intersection of two walls and the ceiling or floor) a TrackSense unit can determine its full 6 degree of freedom position and orientation with respect to the corner. If it can identify the corner, it can also obtain its global position and orientation within the room. Note that our prototype only has enough resolution and field of view to accurately detect two planes simultaneously. Hence, the analysis in Section 4.2.2 assumes a constant Z value (the unit sat on a wheeled platform). We expect in actual operation, 2-5 TrackSense units would operate on the same rigid body. Even if three TrackSense units could each only detect a single unique plane, the combination of distance and orientation data from each TrackSense unit would be equivalent to a single “super TrackSense” observing the three planes directly.

3.2.6 Adding a Magnetometer and Accelerometers

The previous analysis assumes that the TrackSense unit is somewhat vertical. If it is held upside down, its yaw, pitch, and roll measurements will be incorrect by 180 degrees, and if held sideways it will mistake the floor and ceiling for walls and visa versa. If the TrackSense unit will be held generally level (within 30 degrees), the use of accelerometer data is not strictly necessary for satisfactory operation. However, most solid state magnetometers integrate a 3 axis accelerometer (and 2 orthogonal magnetometers) to ensure that the compass bearing is accurate even if the unit is tilted. By using data from a 3 axis accelerometers we can enhance the TrackSense unit in two ways:

- By detecting the 1G acceleration of gravity and magnetic north the unit can operate in any orientation and provide correct yaw, pitch, and roll data (Excepting zero-gravity environments).
- The data from inexpensive accelerometers (with fast update rates, but moderate drift) can be used to provide updated position and orientation data between camera frames or while the cameras do not have a view of enough planes to obtain full 6 DOF data.

4 Performance Evaluation

We conducted four experiments to determine the accuracy and precision of the system's position and orientation measurements. The first two involved measuring the distance and orientation with respect to a single wall or plane. The third measured the ability of a single TrackSense unit to measure the angle between two walls or planes. The fourth experiment used the distance and orientation to the intersection of two walls to measure the (2D) location of the TrackSense unit in a room. These tests allow us to report on the overall positioning and orientation accuracy of our prototype and predict the accuracy and precision of a system using multiple TrackSense units to position and orient itself within a room. For accuracy, we report the difference between the system's determination of its location and the measured ground truth. For precision, we show the smallest discernable position unit by observing the variations of the system's reported position at specific locations.

4.1 Distance and Orientation to a Single Wall

Figure 6 (left) shows the single wall experimental setup. At each test point 300 consecutive samples were taken. In the two experiments, we varied the prototype's distance and angle to the wall while keeping lighting constant at normal office illumination levels.

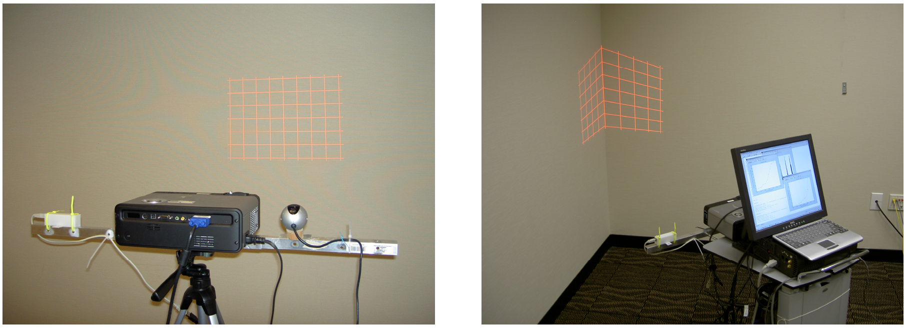


Fig. 6. Left: Experiment set-up for one wall experiment with projected grid. Right: Two wall experiment setup with projected grid. In both experiments, the apparatus was at a fixed height.

4.1.1 Distance

In this experiment, the apparatus was pointed straight towards one wall and the perpendicular distance was measured as the ground truth. We took measurements at nine points ranging from 75cm to 325cm from the wall. The TrackSense prototype has a minimum range slightly under 75cm (due to geometric constraints), and the optimal working range extends to 275cm, although less accurate results can be obtained up to 350cm.

The accuracy of the system is shown in Figure 7. The straight blue line represents the result of a least mean square linear regression for the sample points. The distance between the measured and the actual data closely follows a linear function ($y=ax+b$). This systematic error comes from the fact the lines from the projector increase in thickness as the apparatus is moved farther back, thus shifting the detected lines farther to the left. Using a true laser grid would mitigate this problem substantially, although there would still be a slight systematic error. However, because the system error is linear, a correction factor can be applied at the factory to improve the overall accuracy. This correction factor is a simple offset value learned through experimentation. When we applied the linear correction factor, our corrected accuracy was between 3-4 cm.

Within the working range, the precision is on average 2 cm (see Figure 7). The results indicate that the standard deviation increases quadratically with increasing distance from the wall. As the area imaged by our camera increases with the square of the distance, this curve corresponds to the expected reduction in sensor resolution with respect to wall area.

Precision decreases drastically near 3 meters. Several factors cause the system performance to begin degrading at this point. As the distance from the TrackSense unit to the wall is increased, the intensity and size of the projected lines in the camera image is reduced until the edge detection algorithm can no longer successfully identify all lines. With fewer lines, fewer points are detected, and more incorrect point correspondences are made, leading to more outliers for the RANSAC algorithm. Increasing the resolution of the camera and using a brighter laser grid projector would increase the effective working range.

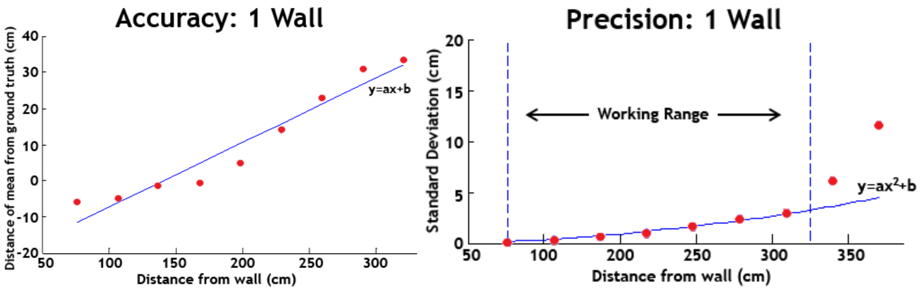


Fig. 7. Accuracy (left) and precision (right) of distance facing a single wall

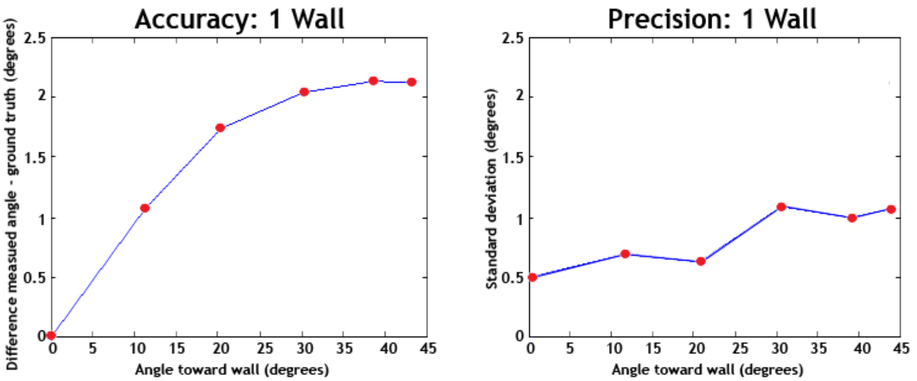


Fig. 8. Accuracy (left) and precision (right) for angle measurement towards one wall

4.1.2 Orientation

We also characterized the system's ability to determine its orientation with respect to a single wall or plane. We measured the angle between the prototype and a wall at six different angles: 0,10,20,30,40 & 45 degrees. At each position, 300 samples were collected to calculate the device's precision. The TrackSense prototype was located 120 cm from the wall, and swiveled from 0 degrees (directly facing the wall) to 40 degrees in 10 degree increments, and a final measurement was taken at 45 degrees. Beyond the 45 degree angle we expect the TrackSense unit would have a less acute angle to an adjacent wall.

Over the tested 0 – 45° range, the TrackSense prototype has an accuracy of 2° or better and has a precision of 1° or better (see Figure 8). We attribute the slight decrease in precision and accuracy as the angle increases to the smaller surface area that is visible to the camera as the incident angle is increased. We did not measure angles beyond 45 degrees because in standard operation we expect the TrackSense unit to have a view to an adjacent wall with a less-acute angle.

4.2 Two Wall Experiments

We investigated using our prototype to measure the angle between two walls, as well as using the distance and orientation from a known corner to measure the (X,Y) location of our prototype within a room (see Figure 6). All measurements were taken with the grid centered horizontally in the corner and the TrackSense unit directly facing the corner.

4.2.1 Angle Between Two Walls

The ability to calculate the angle between intersecting planes is important in recognizing unique corners (*e.g.,* an odd shaped room where each corner has a different angle). For the two wall angle experiment, a movable surface against the corner of a room approximated a second wall, and measurements were taken at three different angles at 90°, 67.5° and 45° as measured with a protractor. The accuracy of angle measurements is shown numerically in Table 1. TrackSense provides accuracy of better than 2°. The accuracy degrades as the angle gets narrower. As the angle between the walls decrease, the angle from each wall to the TrackSense unit increases causing an increase in error similar to that seen in Section 4.1.2. The precision of the system remains relatively constant (around 1°) despite the angle of the walls.

Table 1. Accuracy of angle measurements between two walls

Ground truth	Measured angle (Mean)	Difference (Error)	Standard Deviation
90.00°	90.08°	0.08°	1.70
67.50°	69.09°	1.59°	1.83
45.00°	43.06°	1.94°	1.25

4.2.2 Location Within a Room

By observing a corner, our prototype measures the distance to two walls and can produce an (X,Y) location within a room. Using a fixed height, and keeping the TrackSense grid projector and camera pointing towards a known corner, measurements were taken from a total of 25 positions equally spaced on a grid covering an area of approximately 2.0m x 3.0m. We took 300 samples at each location with standard office lighting conditions.

Figure 9 shows the raw data results of the two wall position experiment. Throughout the experiment, we pointed the apparatus toward the lower left corner of the room (the corner at the origin of the coordinate frame). This raw data has an average accuracy of only 29.0cm. When we apply the linear correction factor discussed in Section 4.1.1, the average accuracy of all 25 data points is increased to 17.3cm. If we look at only the 9 points closest to the corner (grid size of 1.2m x 1.2m), our corrected accuracy is 9.53cm.

To calculate the precision, we first computed the variances of the detected *X* and *Y* values. *var(X)* and *var(Y)* are the squared mean distances of *X* and *Y* from the mean. Because the data is in a Cartesian coordinate frame, the Euclidean distance can be applied, and the average distance of all samples from their mean is

$$std(X,Y) = \sqrt{var(X) + var(Y)} \,.$$

Within a working range of 2m x 3m the precision of the system is at most 15.8cm (approximately 3 cm – 4cm in each direction) for 90% of the readings (see Figure 9). The device performs the best when it is close to at least one wall when pointed at a corner. The reason for the less accurate results when compared to the single wall experiment is that fewer points were being used to define each plane, as the grid pattern is distributed across two different walls. An accuracy of 10 to 17cm with 15cm of precision is still significantly better than other indoor location systems that

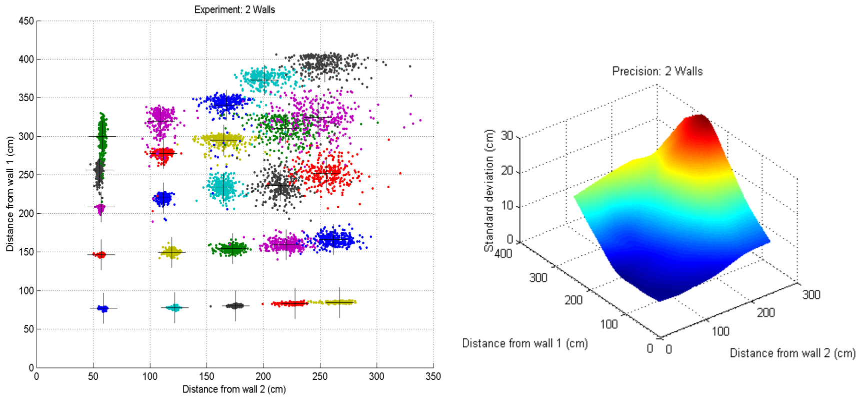


Fig. 9. Two wall experiment. Left: Each dot represents a single data sample. Dots of the same color belong to the same position data set and black crosshairs show mean values. Right: Up front view of interpolated standard deviation of the two wall readings.

do not require the deployment of infrastructure. By using multiple TrackSense units, each plane in the room would be illuminated by more feature points. This would increase the total accuracy and precision, approaching the performance of the single-wall experiments.

5 Miniaturization and Addressing Limitations

In this section, we discuss miniaturizing this system and present two realistic prototypes: a handheld device and a head-mounted unit. We also discuss the limitations of the current prototype and show how we would extend the system to cover an entire room. Finally, we consider some limitations that are beyond simple engineering considerations.

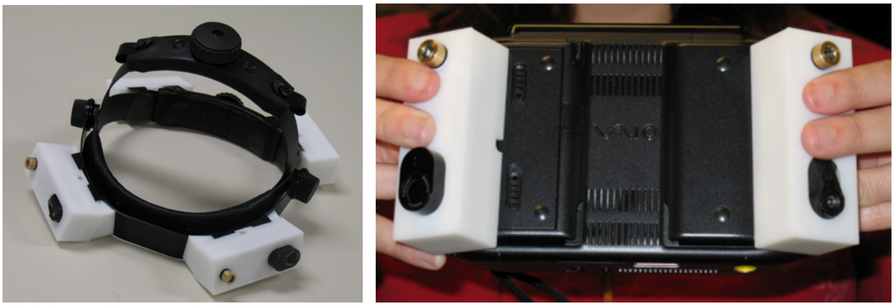


Fig. 10. Headset and handheld with active components on miniaturized TrackSense units

5.1 Miniaturization and Improvements

Our system is composed of relatively simple parts. We chose to use a projector in our prototype because it gave us the flexibility to experiment with various patterns quickly. Because the projected pattern is static, we can replace the projector with an infrared grid laser diode. The infrared diode would eliminate the visible patterns and increase the range because it is brighter. We would replace the camera with a smaller, black-and-white camera with an infrared pass filter and could place multiple laser/camera units on the localized device as a result of the miniaturization (see Figures 1 and 10). Theoretically, a single TrackSense unit with a wide enough field of view could always have at least one corner in its field of view, permanently maintaining a position and orientation fix. Practically, we expect several TrackSense units angled in different directions to cooperatively identify three planes within their combined views, interpolate the location of the (possibly non-observed) corner, and determine their location and orientation regardless of their platform's motion. For the handheld unit (see Figure 10), we placed two TrackSense units facing forwards and angled 45 degrees away from center. This configuration ensures that two walls are detected at any given time for proper localization within the entire room. A third camera facing up or down enables full 3D positioning. Another strategy is to slightly angle the two front facing units up to capture the wall and ceiling corners, eliminating the need for the third. However, this solution would also limit how much the user can tilt the handheld forward.

On a head-mounted device (see Figure 10), we can place four units looking at 90 degree intervals. The units could be angled slightly upwards for full 3D positioning or a fifth could be added pointing vertically. The advantage of the head mounted unit is that more units can be installed facing in opposite directions, which would result in better precision in a larger room. As we saw in the results, the farther the device is from a wall, the lower the precision and accuracy. Since the head mounted device has a full view in all directions, the system can select the closest walls to offer the best results.

5.2 Limitations

There are still several limitations of our approach worthy of further examination. The current solution only supports one device in a room at a time. This might be acceptable for some applications, but not for multiplayer games or collaborative applications. One solution is to synchronize the devices and have them alternately flash their patterns. Since the devices know their position within the room, the devices can turn off certain parts of the grid to avoid interfering with another device.

An important limitation to our approach is the need for walls in the space. The wall has to be free of major obstructions and large windows. In our experience, posters and other flat objects do not cause major problems, and our implementation can detect outliers. However, many raised objects on the wall cause the system to incorrectly identify the plane. The current technique also assumes a flat surface with little to no curvature, limiting the types of rooms that are appropriate. However, our intent with this device is to enable applications where a user would already want to interact with multiple large surfaces that are relatively plain in the first place [5]. If the intent is to extend our solution to more complex spaces we can incorporate stereo vision techniques that work well in cluttered environments and use both approaches in a complementary fashion.

Some dark wall colors cause problems for detecting the grid. Very bright lighting conditions (*i.e.*, near a window during the day) can make the projected lines too faint. However, most artificial lighting from standard fluorescent and incandescent lights does not cause major problems with detection. Another concern is very tall ceilings, in which case the units would have to be oriented to detect walls and floor corners. To obtain three dimensions, we must select a unit to face downwards. The detector camera resolution also limits how far a unit can be from a wall, thus limiting the working range in a room. Our experimental prototype was limited to a room approximately 5m x 5m in size. A higher resolution camera and the use of a laser grid would improve those limits.

6 Conclusion

We presented TrackSense, a localization system that requires no additional infrastructure in the environment and provides 3D positioning and orientation data that compares favorably against existing solutions. Inspired by robotics localization and camera-projector calibration techniques, our solution uses a camera to locate and track a grid pattern projected onto surfaces in the camera's field of view to determine its distance and orientation to multiple fixed large planes in a space (*i.e.*, walls and ceilings). A system of TrackSense units can obtain up to 4 cm accuracy with 3 cm precision in rooms up to 5 square meters, as well as 2 degree accuracy and 1 degree precision on orientation. The relatively simple hardware used in its implementation makes miniaturization possible.

TrackSense provides localization within a room, but combining it with a room-level localization system, such as WiFi or GSM fingerprinting, can provide localization within a global coordinate frame. In addition, our solution provides a useful complement to traditional stereo vision techniques, which do not perform well on plain surfaces. The addition of another camera would provide localization within both a cluttered and uncluttered environment, thus extending the capabilities of the device further.

References

1. Active Bat. The BAT Ultrasonic Location System. http://www.uk.research.att.com/bat/. 2006.
2. Bahl, P. and Padmanabhan, V. RADAR: An In-Building RF-Based User Location and Tracking System. In the proc of *IEEE Infocom 2000*. Los Alamitos. pp. 775-784. 2000.
3. Bouguet, J.Y., and Perona, P. 3D photography on your desk. ICCV'98, pp. 43–50, 1998.
4. Canny, J. A Computational Approach to Edge Detection. In IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 6, Nov. 1986.
5. Cao, X. and Balakrishnan R. Interacting with dynamically defined information spaces using a handheld projector and a pen. In the proc. of UIST 2006. pp. 225-234. 2006.
6. Cobzas, D. and Sturm, P. 3D SSD Tracking with Estimated 3D Planes, In the proc. of *2nd Canadian Conference on Computer and Robot Vision (CRV) 2005*, pp. 129-134. 2005.
7. Davison, A., Cid, Y., Kita, N. Real-Time 3D SLAM with Wide-Angle Vision. In proc. of IFAC Symposium on Intelligent Autonomous Vehicles. 2004.
8. Dellaert, F., and Tariq, S. A Multi-Camera Pose Tracker for Assisting the Visually Impaired In 1st IEEE Workshop on Computer Vision Applications for the Visually Impaired. 2005.
9. Dissanayake, M. W. M. G., Newman, P., Clark, S., Durrant-Whyte, H. F., and Csorba, M. A solution to the simultaneous localization and map building (slam) problem. IEEE Transactions on Robotics and Automation, 17, 3 (June 2001), pp. 229-241.
10. Fischler, M. and Bolles, R. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. In Communications of the ACM. Volume 24, Number 6, pp. 381-395. 1981.
11. Forsyth, D.A., Ponce, J. Computer Vision: A Modern Approach. Prentice Hall, U.S. 2002.
12. Harle, R.K., A. Hopper. Cluster Tagging: Robust Fiducial Tracking for Smart Environments. 2nd International Workshop on Location- and Context-Awareness 2006. pp. 14-29. May 2006.
13. Hartley, R., Zisserman, A. Multiple View Geometry in Computer Vision. Cambridge Press. 2003.
14. Hightower, J. and Borriello, G. A Survey and Taxonomy of Location Systems for Ubiquitous Computing, University of Washington Tech Report CSC-01-08-03. 2001.
15. LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, I., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, R., Borriello, G., and Schilit, B. Place Lab: Device Positioning Using Radio Beacons in the Wild. In the proc. of *Pervasive 2005*, pp.. 116-133. 2005.

16. Kuhn, H.W. The Hungarian Method for the assignment problem. Naval Research Logistic Quarterly, Issue 2. pp. 83-97. 1955.
17. Lourakis, M.A., Argyros, A.A. Vision-Based Camera Motion Recovery for Augmented Reality. In the proc of the Computer Graphics International (CGI 2004). pp. 569-576. 2004.
18. Liu, Y. and Thrun, S. Results for outdoor-SLAM using sparse extended information filters. In Proc of the IEEE International Conference on Robotics and Automation. pp. 1227-1233. 2003.
19. Madhavapeddy, A. and Tse, T. Study of Bluetooth Propagation Using Accurate Indoor Location Mapping. *UbiComp 2005.* Tokyo, Japan. pp. 105-122. September 2005.
20. Munkres, J. Algorithms for the Assignment and Transportation Problems. Journal of the Society of Industrial and Applied Mathematics, 5(1). pp. 32-38. 1957.
21. NorthStar. Evolution Robitcs. http://www.evolution.com/products/northstar/. 2007.
22. Otsason, V., Varshavsky, A., LaMarca A., and de Lara, E. Accurate GSM Indoor Localization. In the proc. of *UbiComp 2005*. Tokyo, Japan. pp. 141-158. September 2005.
23. Patel, S.N., Rekimoto, J., Abowd, G.D. iCam: Precise at-a-Distance Interaction in the Physical Environment. In the proc. of Pervasive 2006, pp. 272-287, May 2006.
24. Patel, S.N., Truong, K.N., and Abowd, G.D. PowerLine Positioning: A Practical Sub-Room-Level Indoor Location System for Domestic Use. In the proc. of *Ubicomp 2006*. pp. 441-458. 2006.
25. Priyantha, N. B., Chakraborty, A., and Balakrishnan, H. The Cricket Location-Support System. In the proc. of *Mobicom 2000*. Boston, MA. August 2000.
26. Rekimoto J. and Ayatsuka Y. CyberCode: Designing Augmented Reality Environments with Visual Tags. In the proc. of *DARE 2000*. Elsinore, Denmark. pp. 1 – 10. 2000.
27. Rachmielowski, A., Cobzas, D., Jagersand, M. Robust SSD tracking with incremental 3D structure estimation. In the proc. of *Canadian Conference on Computer and Robot Vision (CRV)*. 2006.
28. Rekimoto, J and Katashi, N. The World through the Computer: Computer Augmented Interaction with Real World Environments. In the proc. of *UIST 1995*. Pittsburgh, PA. pp. 29-36. 1995.
29. Scharstein, D and Szeliski, R. High-accuracy stereo depth maps using structured light. In CVPR 2003, volume 1, pp. 195-202, Madison, WI, June 2003.
30. Se, S., Lowe, D., Little, J. Vision-Based Mobile Robot Localization and Mapping Using Scale-Invariant Features. In the Proc of ICRA 2001. 2001.
31. Sukthankar, R., Stockton, R., Mullin, M. Smarter Presentations: Exploiting Homography in Camera-Projector Systems. In the proc. of ICCV 2001.
32. Sukthankar, R., Stockton, R., Mullin, M. Automatic Keystone Correction for Camera-assisted Presentation Interfaces. In the proc. of International Conference on Multimodal Interfaces. 2000.
33. Tariq, S. and Dellaert, F. A Multi-Camera 6-DOF Pose Tracker. In IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR). pp. 296-297. 2004.
34. Thrun, S., Burgard, W., and Fox, D. Probabilistic Robotics. MIT Press, Cambridge, MA, 2005.
35. Ubisense. http://www.ubisense.net. 2006.
36. Vicon MX. http://www.vicon.com/products/systems.html. 2006.
37. Want, R., Hopper, A., Falcao, V., and Gibbons, J. The active badge location system. *ACM Transactions on Information Systems*. Volume 10. pp. 91-102. January 1992.

Objects Calling Home: Locating Objects Using Mobile Phones

Christian Frank[1], Philipp Bolliger[1], Christof Roduner[1], and Wolfgang Kellerer[2]

[1] Institute for Pervasive Computing, ETH Zurich, 8092 Zurich, Switzerland
{chfrank,bolligph,roduner}@inf.ethz.ch
[2] DoCoMo Communications Laboratories Europe, Munich, Germany
kellerer@docomolab-euro.com

Abstract. Locating physical items is a highly relevant application addressed by numerous systems. Many of these systems share the drawback that costly infrastructure must be installed before a significant physical area can be covered, that is, before these systems may be used in practice. In this paper, we build on the ubiquitous infrastructure provided by the mobile phone network to design a wide-area system for locating objects. Sensor-equipped mobile phones, naturally omnipresent in populated environments, are the main elements of our system. They are used to distribute search queries and to report an object's location. We present the design of our object search system together with a set of simple heuristics which can be used for efficient object search. Moreover, such a system can only be successfully deployed if environment conditions (such as the participant density and their mobility) and system settings (such as number of queried sensors) allow to find an object quickly and efficiently. We therefore demonstrate the practicability of our system and obtain suitable system parameters for its execution in a series of simulations. Further, we use a real-world experiment to validate the obtained simulation results.

1 Introduction

To be able to locate everyday objects at the touch of a button is a promising application of ubiquitous computing. However, most systems which can be used for this purpose require an expensive infrastructure for sensing objects, for example, using Radio Frequency Identification (RFID) readers installed in the environment [1, 2, 3]. Their dependency on infrastructure precludes such systems from being used for locating objects in areas larger than confined indoor environments, particularly because of the costs involved in adding infrastructure to a significant fraction of the world, i.e., the space in which objects can be placed.

In this paper, we describe an object localization system based on mobile phones. Mobile phones combine two very useful features: They are omnipresent in environments in which users live and are at the same time inter-connected by a homogeneous world-wide infrastructure. Given that important objects can be augmented with an electronic tag such that they can be detected when brought into the vicinity of a mobile phone, many new applications become possible revolving around managing, monitoring, or locating one's everyday items.

Various technologies could be employed for object tagging and sensing objects within a short range of the mobile phone. For example, RFID tags are expected to be

A. LaMarca et al. (Eds.): Pervasive 2007, LNCS 4480, pp. 351–368, 2007.

Your object 'Umbrella' has been found at 'Office' by 'Bob'.

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

Fig. 1. User issues an object search query

attached to various consumer products in the near future as they may realize significant cost savings in stock and supply chain management. In particular, passive UHF RFID technology or active tags with a small autonomous power source [4] are expected to provide reading ranges of a couple of meters even with small reader modules. If improved variants of today's handheld RFID readers were integrated into mobile phones, a ubiquitous system could be deployed within a few years using the short innovation cycle established through mobile phone sales. In addition to RFID, other upcoming radio communication technologies, some even compatible with the phone's Bluetooth capability, could be used to identify objects in the phone's physical proximity in a similar way. If small, inexpensive Bluetooth-discoverable tags [5] can be built (in fact our prototype relies on battery-powered BTnodes [6]), a ubiquitous object sensing infrastructure is already in place today.

Note that each tagging technology defines a certain trade-off between tag costs, achievable identification range, and costs of reader hardware. Irrespective of the employed technology, we assume that *object sensors* can be integrated into mobile phones, as it has already been done with Bluetooth or NFC today. On a campus, in an office building, or, generally, any relatively dense urban environment, it would be possible to task mobile phones carried by other users with searching for an item one is interested in. In an office environment, any employee's mobile device could participate in sensing such tagged items and "call home", that is, notify the owner once the item is found using the short-range object sensor. If the notification includes some indicator on the location of the device which found the object, the owner may already have enough information to retrieve the item. For example, even a simple confirmation that some item has been left at work, as shown in Figure 1, can be very useful to users as they may stop searching somewhere else (e.g., at home) or just feel re-assured in case a valuable item is missing.

In this paper, we describe the design of our object sensing system based on mobile phones. A particular challenge of our application is to distribute an object search query to a subset of users (more generally, to object sensors) that are likely to find a given item. For this, we describe a set of heuristics which our system uses to define the scope of a query. Furthermore, by means of an extensive evaluation of the system's behavior, we demonstrate the practicability of the presented system and the used query scoping heuristics under varying operational conditions and with different system parameters. As a result of our evaluation, we obtain adequate system-parameter settings for a range of usage scenarios.

The remainder of the paper is organized as follows. We begin by surveying related work in Section 2 and describe the service architecture involved in providing our prototype's functionality in Section 3. In Section 4, we then discuss heuristics used for query scoping and detail the query dissemination protocol in Section 5. In Section 6, we provide initial evaluations obtained using our prototype in our own office environment. Further, in Section 7 we describe the simulation model we use to study the system's behavior in a wide-area environment and present the obtained results in Section 8. We summarize the results and present our conclusions in Section 9.

2 Related Work

Various related work has argued for the relevance of locating everyday objects, monitoring the presence of items, or avoiding their loss. Many such systems employ a pre-installed object sensing infrastructure [1, 2, 3]. Compared to these systems, we do not rely on a pre-installed infrastructure which is costly to deploy and to maintain, but focus on the use of mobile phones as hubs to a ubiquitous infrastructure. In the Smart Watch prototype [7], RFID readers transmit their current readings to the passing user's personal device, thus enhancing it with an object sensor. The user is then notified if objects are missing compared to readings which were collected earlier. Another prototype [8] allows train passengers to register their luggage with pre-installed object sensors, which then protect against theft and remind passengers of their items before unboarding the train. Both systems focus on reminding users before a loss takes place. As we assume that tagged objects will be numerous and often intentionally left behind, we aim to avoid immediate notification. Instead, we study how the location of an item can be determined on a user's request.

Note that an object search system could also be implemented by proactively sending all sensor readings to a centralized service which could be queried when an object needs to be located. Such a system would face the challenge of a global data collection system, such as IrisNet [9] or Hourglass [10]. In this paper, we use a reactive, that is, query-based approach, as the number of sensor readings (e.g., *object X seen by object sensor A*) is expected to be much larger than the number of queries. Moreover, our system avoids aggregation of all readings in a centralized database as this would have severe implications for the user's personal privacy. In particular, it is incompatible with a privacy enhancing feature of our system: Objects which have previously been associated to their owner, will only be detected by a sensor after this sensor has received an explicit query for the given object. Search queries for such objects contain the owner's key obscured by a random session identifier thereby implementing a lightweight authentication protocol [11] between owners and their objects.

Finally in [12], the authors use the mobile phone as a gateway to access heterogeneous health-related sensors which have been pre-installed in the environment of the user. Based on a different application, their system does not deal with two aspects which are central in this paper: Query scoping (determining which sensors should be queried from a large and homogeneous sensor array) and the obtained sensor coverage (identifying the numbers of users and the mobility patterns which allow for reliable detection of objects).

3 Use Case and Service Architecture

The main use case of our system, outlined in Figure 1, includes various aspects which we describe in this section together with the respective services implementing them.

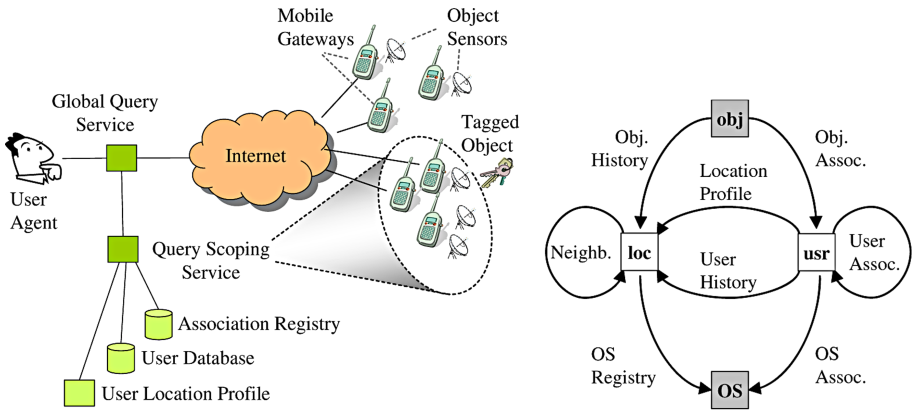
Association. The association service serves three main purposes. First, it keeps track of associations between users and objects (Figure 1(b)). Objects are visible to everybody in their initial state, but with association become visible only for queries initiated by their owner's device. Such association, accompanied by the exchange of a shared key [11], prevents other users from using the system to perform a wide-area search for the associated object. In this context, the mobile network operator may act as a gatekeeper and only distribute queries issued by object owners. Second, user to object sensor association maintains a set of object sensors which are particularly relevant to the user, for example, object sensors which have been installed at a remote holiday home. In the scenario of Figure 1, Bob's mobile device has been previously associated as a favorite object sensor. Third, similar user to user associations can be used to grant group access rights to certain objects, e.g., for families or groups of colleagues, or to determine in which circumstances users' identities are revealed.

Storage. However, our system does not depend on associated object sensors or users being near the tagged object at the time the query is issued. Instead, it stores certain context information when an associated object leaves the range of the local object sensor. Specifically, the mobile device stores a location trace of the user around the "loss" event which can help finding it later on. In a different usage scenario, which is not the focus of this paper, user-installed object sensors may simply report a stream of sensed objects, for example, carried by users passing by. For these cases, the service infrastructure provides users with a *user database* service that may be used as a sink for events generated by object sensors and mobile devices. Moreover, all association relationships are kept in a storage component called *association registry*. Storage services are available both on the user device and in the back-end infrastructure [13].

Localization. For remembering the location of an object when it goes out of range of the local object sensor and to provide location information for found objects, some location information must be available on the mobile device. While the prototype implementation is based on UMTS cell information, better localization functionality can be added in a future system if increased accuracy is desired. We will quantify the effect of varying positioning accuracy in Section 8.

Location Profile. In our prototype, we adapted [14] to perform statistics on the UMTS cells in which users spend most of their time. This will allow us to implement a search strategy which mainly considers locations where the user spends much time. Our prototype includes functionality for naming these locations [15], such as 'Office' in Figure 1(d).

Figure 2(a) gives an overview of the system architecture: As mentioned, the mobile phones are used to link *object sensors* to the back-end infrastructure. The back-end infrastructure hosts the *global query service*, which provides adequate dissemination support, cost control, and validity management for user queries, as we will describe in Section 5. The global query service in turn is based on the *query scoping service*, which implements application specific heuristics for retrieving the most appropriate subset of object sensors and is described in the following Section 4.



(a) System architecture (b) Data model

Fig. 2. System architecture and data model

4 Query Scoping Service

There are various heuristics for distributing a query to a relevant subset of object sensors. For example, one may distribute it to sensors near the location where the object was last in range of the local object sensor. Similarly, all conceivable heuristics will be based on some kind of history data available in the system. To elaborate on this, we show a simple data model of our application in Figure 2(b): *Objects* are associated with *users* (object owners) by the association service and also with *locations* (e.g., cells) in which an *object* has been observed in the past. Users may choose to record a history of their location on their mobile device (*user history*) or to enable the *location profile* service, which computes the *locations* that are most relevant for a given user (e.g., their home or office). In this simple model, locations are related to other locations via the *neighborhood* relation. Moreover, *users* can be associated with certain *object sensors* which they often use (e.g., which they have installed in their office or car) and with other *users* which are family, friends, or colleagues. Finally, the mobile network operator keeps a database (*OS registry*) which stores the current location (e.g., the current network cell) of certain mobile phones which can be used as *object sensors*.

Note that we omitted some details in the data model (most prominently a more refined location model). However, we can use the data model to show how many conceivable heuristics correspond to paths from an entity of type *obj* at the top to an entity of type *object sensor (OS)* at the bottom of Figure 2(b).

For example, we can query object sensors which:

I) Are near the location where the object was last seen.
II) Are near locations recently visited by the user.
III) Are near locations where the user spends a large amount of her/his time.
IV) Are associated with the object owner (as in Figure 1).
V) Match the above strategies III and IV for a different (associated) user, such as a family member, or even for a friend of a friend, etc.

While, intuitively, none of these heuristics can guarantee success, they all incorporate particular application-level assumptions on where users keep personal belongings and where these are generally left. Note how each heuristics represents a path in the data model of Figure 2(b): Heuristics I corresponds to the path (*obj-loc-OS*) on the left, while heuristics V corresponds to the path *obj-usr-usr-loc-OS*.

The application programmer may now define which relation types to use in the search by assigning weights to each one. The search algorithm, described in detail in complementary work [16], is then started with a source entity (i.e., the sought object) and a *destination type* (i.e., object sensors) as parameters. It will then "unfold" this annotated data model into a search tree of entities related to the sought object. Each edge in the tree, during algorithm execution, will be annotated with a relevance measure, derived from the user-defined relevance of the respective relation. The entities in the tree are visited in order of decreasing relatedness to the source entity. If a visited entity is of type *destination type*, it will be added to the algorithm's result list (in which contained entities are also ordered by their relatedness to the sought object). The algorithm stops once *entity limit* entities have been returned or relatedness falls below a given *relatedness threshold*.

The algorithm's advantage is that relations can be encapsulated in distributed components executed on various platforms of the system (e.g., some users would prefer to store their location profile on their mobile device only, while other relations are stored on the back-end server). In this paper, we do not focus on the search algorithm itself (details are described in [16]), but analyze the performance of the contained heuristics.

Please note that heuristics I-III employ the *OS registry* relation, which associates a set of locations L with a set of object sensors near them. Due to user mobility, however, the object sensors near the set L will change with time. To compute a query scope that is independent of user mobility, when heuristics I-III are implemented, *location* will be the *destination type* parameter passed to the search algorithm. The returned set of locations L, for example, a set of cells, is then passed on to the global query service, which will distribute the query to these locations as described in Section 5.

5 Global Query Service

If users are interested in finding an object o, they will issue a find query to the global query service. At this time, they may specify a message cost limit $q_{\max}$ denoting a limit on the messages sent during query dissemination and a time limit $t_{\max}$ after which the query will terminate at last.

If the search strategy IV is chosen, the set of object sensors is determined by the scoping algorithm. Here, a query will be distributed to the first $q_{\max}$ sensors returned by the algorithm and be active for at most $t_{\max}$ time.

If search strategies I-III are chosen, query scoping will not directly return a set of object sensors, but a set of locations. In the basic location model we employ, these can either be a set of cells (the most basic localization already available on the phone) or a set of geographic points (if phone localization is more precise) together with an associated measurement error. Because the set of object sensors associated with these locations may change over time, our system installs (or un-installs) a query at sensors which come close to (or, respectively, depart from) these locations. Whether a sensor s is close to the returned locations is defined by the implementation of a predicate f (which maps s to either *true* or *false*).

Depending on the way *locations* are modeled, we use two different implementations of $f(s)$. Given a set of cells C, $f(s)$ will be true if the mobile phone (with its object sensor s) is currently served by any of the cells in C. Note that this information is already available at the mobile network operator, that is, it can be accessed without additional costs on the server side of our infrastructure.

In case the mobile devices are equipped with more accurate positioning means, the locations returned by query scoping will instead be a set of geographic points P. Here, $f(s)$ will be true if the current position measured by a mobile phone's object sensor s is within a certain range r away from the points P. This range r will depend on the error incurred at the positioning sensor when the points in P were measured (we will discuss a concrete implementation in our evaluation section). Note that such additional positioning information will only improve the efficiency of query dissemination if positioning information of all object sensors is already known at a database on the server. Otherwise, it would be inefficient to propagate all object sensor positions to the server before query dissemination, and therefore a different approach is chosen: The query is distributed to object sensors in a set of cells C which "cover" the whole area surrounding the points P (the actual object sensor will be turned on only later, once the predicate $f(s)$ evaluates to true). Note that the total number of distributed queries is now the same as if locations were a set of cells C.

When installing queries for such "location-based" strategies I-III, the total number of object sensors at which a query will be installed (q_{total}) is made up of two parts, $q_{\text{total}} = q_{\text{init}} + q_{\text{mob}}$. Here, q_{init}, denotes the number of users queried initially at the time the query is issued. At this time, q_{init} users/sensors are randomly chosen out of the initial query scope $S_{\text{init}} = \{s|f(s) = \text{true}\}$. The number q_{init} is set as

$$q_{\text{init}} = \min(q_{\max}, |S_{\text{init}}|, A/o_A) \tag{1}$$

where o_A represents the area which can be covered by a typical object sensor (for example a disk around the sensor with a given radius) while A denotes an estimate of the total area in which f would return true. Here, we assume that due to user mobility, A/o_A object sensors should be enough to cover the total area involved, although if all users were stationary, sensing areas o_A may overlap and more queries would be required.

In addition to q_{init}, the query will be installed at a second set of sensors q_{mob} for which $f(s)$ becomes true while the query is active. Given a query duration $t = \min(t_{\text{reply}}, t_{\max})$, where t_{reply} denotes the time at which a sensor has reported having found the object o, this effort q_{mob} can be modeled as

$$q_{\text{mob}} \sim A^{\circ}mt \tag{2}$$

where A° denotes the circumference of the scope area A and m denotes a factor representing the mobility of users. Therefore, both query success and communication effort are expected to rise with t and with m.

After a query installation at an object sensor s, object sensing will be performed continuously until $t_{\max}$ expires. The mobile device associated with s un-installs the query autonomously either when $f(s)$ becomes false or $t_{\max}$ is reached.

A query is declared successful if some object sensor s reports having found o at time t_{reply} with $t_{\text{reply}} \leq t_{\max}$. The current position of s represents the location at which the object was found and will be included in the reply issued to the user (if a user-defined name is associated to the location of s, a reply will look as in Figure 1(d)). Once the first report is received, it could be useful to uninstall the query at all participating sensors (sending an additional $q_{\text{init}} + q_{\text{mob}}$ messages) in order to avoid useless "object-found" reports if sensors come in range of the object at a later time. However, as the number of such useless reports is usually much smaller than $q_{\text{init}} + q_{\text{mob}}$, we do not explicitly uninstall the query, but instead let each sensor s autonomously remove the query on the timeout $t_{\max}$ or when $f(s)$ turns false. Finally, a query is terminated without success, once the query timeout $t_{\max}$ is reached (in contrast, reaching $q_{\max}$ does not terminate the query – in this case, a reply may still be received by queried sensors due to user mobility).

In the remainder of the paper we will evaluate the presented query scoping service based on the search heuristics of Section 4. In particular, Sections 7 and 8 focus on evaluating the practicability of strategy I where we search an area close to the location where the object was last seen. Strategies II and III, similarly based on locations, depend on real-world data which is hard to model accurately. Nevertheless, one can use the considerations of Eq. (1-2) to relate the results we will present for strategy I to different search areas with different sizes. Finally, in the following Section 6, we evaluate the simple heuristics IV by means of a small user study performed in our office environment – the obtained results will then serve as a validation of the results obtained through simulation.

6 Real-World Experiment

The crucial question is whether the object search system can perform well enough to be a useful application. For answering this question, we first come back to the scenario introduced in Figure 1, where the user is at home and tries to verify the whereabouts of a given object which was left at the office. The mobile phones of the user's officemates (e.g., Bob) are registered with the association service and thus are considered relevant object sensors.

Our experiment was performed with four users working on the same floor. The users were given mobile phones running the object search prototype already tasked to perform continuous object sensing for all objects (using repeated Bluetooth discovery) and reported them in regular intervals to the back-end database. Similarly, 10 BTnodes [6] representing tagged objects were distributed in various rooms of the same floor. Figure 3 shows the experiment's setup (tagged objects are shown as numbered circles while the offices of the four participating users are shaded).

Note that while Bluetooth may be too expensive and battery-intensive to be used as an object tagging technology in a future system, it nevertheless allows to test whether, given a future technology with similar radio range, the mobility of a few office colleagues suffices to detect a given object in reasonable time.

Each user's readings were reported to the database as *(user, time, obj_id)* tuples. We considered only readings obtained during core office hours (9 a.m.–5 p.m.) while all others were discarded, resulting in around 30 hours of data. Using the collected data, we could compute the reply time of a query for a given object o issued to the four colleagues at an arbitrary point in time (say at time t_q): The reply time corresponds to the time between t_q and the next database entry on the queried object o.

Based on this consideration, we computed the expected average query reply time for each object, given that queries for this object were distributed uniformly over the

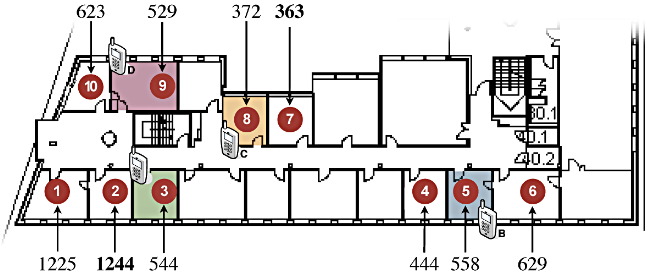


Fig. 3. Experiment setup

experiment time. Note that non-office hours are simply "skipped" in this computation. In order to save messaging costs, user devices cached seen objects and only re-reported them to the database 10 minutes after their last report on the same object. This way, even if an object sensor has seen the object continuously, the resulting reports will yield an average query reply time of 5 minutes instead of zero.

For each object, Figure 3 shows the average reply time in seconds. Intuitively, we obtain better results for objects with a participating user in the same room. Further, note that the best results were obtained for objects close to the printer and the coffee machine (objects 7 and 8), while the worst results are for objects in rooms that were not visited by the participants during the time of the experiment.

We show a cumulative density curve of the observed reply times for object 2 (with worst results), object 7 (with best results), and the average over all objects in Figure 6(a). In all cases, reasonable success rates can be obtained with a maximum query time $t_{\max}$ of 30 minutes.

7 Model

In the last section, we focused on a small and confined search area and query scope. In the remaining sections, we use simulations to investigate the characteristics of an object search system operating in the wide-area with a larger user base, which provides us with a basis for the design of such a system given certain environmental conditions.

Note that adequate models of a future execution environment are hard to obtain, as these must consider many aspects of daily life. To provide an accurate basis for the design of an efficient system, models must define the number of participating users, the frequency at which these users lose or search for certain objects, particular scenarios in which objects are lost, the average number of tagged objects owned by each user, and so forth. Intuitively, such a model contains many parameters which cannot be influenced by the system designer. We call these *environmental parameters*. Our approach to these parameters is to investigate a significant portion of the (unfortunately) vast parameter space.

On the other hand, there are many *design parameters* which determine the system's performance and can more or less directly be set and varied by the system developer. These include the size of the search scope (the number of users that participate in searching for an object), the sensing range of an object sensor (which can be influenced by employing more expensive tag and object sensing hardware), or the timeout used for queries. For these *design parameters* we aim to find the most approriate values, i.e., the parameter settings which can implement object search with the least communication overhead.

Scenario and Metrics. In the evaluated scenario, a user misplaces an object o and later issues a search query to the global query service. We assume that at the time the object left the range of the integrated object sensor, the user's mobile device recorded its location p. This location p will act as a hint for the search (implementing the presented heuristics 1). We will evaluate two versions, a *cell-based* version in which p is a cell, and a *position-based* version in which p is an actual geographic point measured with a certain positioning error. In both versions, query scoping is performed according to Section 5.

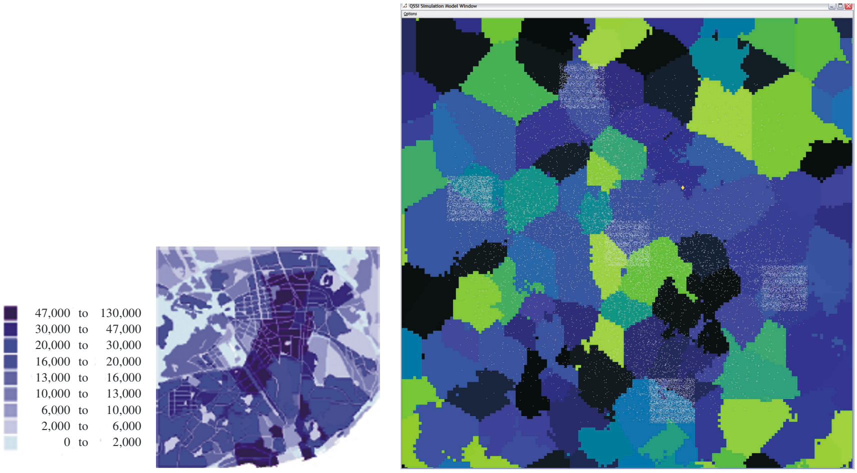
The main metric we observe is the *success rate* of our system. This rate corresponds to the fraction of queries upon which a notification from some object sensor is received within the query timeout $t_{\max}$. Further, we will examine the overhead for query distribution q_{total} and the contained part q_{mob} which is caused by user mobility.

In our experiments, we will not examine object sensing costs explicitly, as we expect wide-area query installation to dominate the total cost due to the object sensor's much shorter wireless range and potential energy efficient optimizations of the object sensing implementation (e.g., object sensing could be performed only after a user has moved).

Environment Model. We assume that the object is left in a densely populated urban environment. In this setting, we will study how an object can be found by users who move according to pedestrian mobility models (see details on the mobility models below) in a square area of 1 km^2. The choice of the user density u_d is derived from the total population during the day in an urban area estimated in the Momentum project [17] (a downtown Lisbon example which we cite from [17] is shown in Figure 4(a)). For urban environments, the authors estimate the fraction of users with "pedestrian" mobility patterns as around 50%-70%[17, p. 37] from the total. The total includes other users who are assumed to be stationary or moving differently, e.g., with higher speeds on streets; we omit these users in our simulations. Moreover, as we are only interested in users associated with *a single* mobile provider, we chose more pessimistic values for the user density u_d: We will vary u_d from 100-2000 users/km^2, values which represent around a hundredth to a tenth of the estimated total daytime population. The default value for u_d is 500 users/km^2.

As mentioned, in some settings we use cell identifiers for positioning. To study such scenarios, we used actual position and orientation data from UMTS antennas together with a detailed model of land use types (e.g., buildings, highways, open, water) provided by the same project to compute the identifier of the strongest-signal cell for each point of the simulation area [17, 18]. In Figure 4(b), we give an example of a resulting cell-coverage map computed for a UMTS network of downtown Lisbon. For cell-based scenarios the simulation area is enlarged to 10 km^2 to avoid border effects with fairly large cells. We are aware that in reality several cells may be observed at a given location at different points in time, depending on dynamic factors such as interference, fading, or user mobility. Because of this effect, in the *cell-based* scenario additional cells would need to be searched to be successful.

In this study, however, we assume that the object can be found in the cell where it was last seen. Results for scenarios in which several cells need to be searched to cover a certain location could be extrapolated from the results we provide.



(a) Lisbon population distribution per km^2 during the day [17]

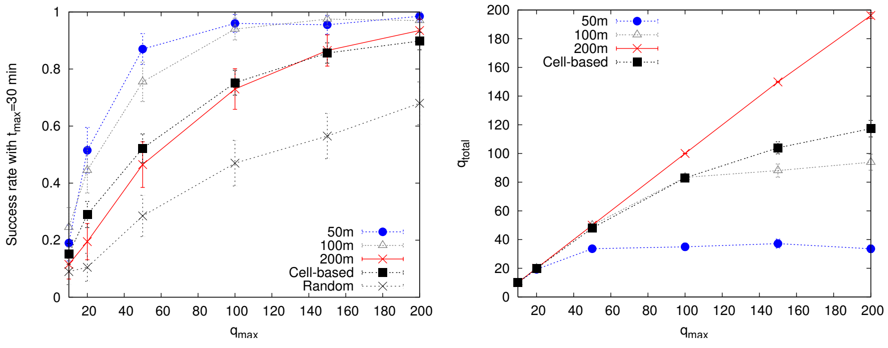
(b) Lisbon cell coverage and campus model

Fig. 4. Environment models

Mobility Models. Generally, the tagged object will not move once its owner left it somewhere. In turn, the users' mobility model is a crucial aspect in our evaluation, as it determines the coverage obtained by object sensors carried by users. Therefore, the user densities we assumed in our simulations are quite small compared to actual densities observed in urban environments, on campus, or on office floors. In this regard, the fraction of simulated users represents the subset of all users who move according to the model we simulate. Additional users, e.g., sitting in their offices or moving differently, may then only improve results.

In the most basic setting, we use a random waypoint mobility model parameterized for pedestrian users. Users pick a random destination and start moving towards it with a speed drawn uniformly from (2,4) km/h. (The average speed of 3 km/h is chosen according to the ETSI guidelines [19].) As our simulation area can be fairly large, a trip's destination is chosen to be within 200 m from the user's current position.

We will also use a second mobility model which was derived from user WLAN traces observed on the Dartmouth campus [20]. The model includes *hotspot* regions which represent central points of the campus (for example, a hotel, a library, or a cafeteria). These hotspots tend to contain many users and also represent popular destinations chosen by the campus population. In our adaptation, we use five hotspot regions out of which one is in the middle and the other four shifted to each side of our simulation area. Each hotspot region's size is one hundredth of the simulation area. Half of the trips of a given user are made inside the current hotspot and half are directed to another (arbitrary uniformly drawn) hotspot on the campus. The remaining (non-hotspot) area is called the *cold* region. In our implementation, users never choose a destination in the cold region but only travel through it. As hotspot regions have a higher density, their positions and sizes are apparent in Figure 4(b), which shows a total density of 1000 users/km^2 in a 10 km^2 area of downtown Lisbon.



(a) Success rate (b) Sent messages

Fig. 5. Success rate and overhead with different positioning technologies

As in [20], the chosen trips include (in our version 2 to 5) waypoints, which are uniformly drawn from the rectangle between the user's source and destination and visited in the order of their distance to the source. The chosen speed and pause times follow log-normal distributions parameterized according to [20, table 3]. Note that the pause time distribution has a mean of 0.71 hours with a high standard deviation of several hours as the original paper found that users tend to stay in a hotspots for longer periods.

To avoid an initial transient period, we used initializations of user trips according to the perfect simulation method [21]. In the campus mobility model, however, some distributions had to be estimated, and thus a transient period of 1000 s remains. Object search queries are issued after this period.

Sensor Model. A mobile device operates its object sensor continuously as long as a query is installed and running. In the default case, we assume that the object sensor has a sensing range of 5 m, that is, the object sensor sends a notification once the user carrying it comes within 5 m of the sought object.

Moreover, in some simulations we assume that the user has a position sensor available (e.g, GPS). To model the sensor's localization error, the position returned by the sensor is drawn uniformly from a disk centered around the actual position of the user. We refer to the radius of this disk as the *positioning error* e_p used in the simulation (e_p is set to 100 m if nothing else is stated). Note that with this error distribution, the density of observing an actual error, say e, is proportional to the circumference of a circle with radius e, and therefore the mean error is $(1/\sqrt{2})e_p$. This positioning error occurs not only when the owner's mobile device records the position p where it has last seen an object, but also when distributing a query to object sensors near p, as these object sensors' positions are measured with an independent positioning error.

Alternatively, we also model a scenario in which the positioning sensor simply returns the identifier of the UMTS cell to which the mobile phone is currently connected. This is a worst case scenario as most cell-based localization approaches combine signal strength information from multiple nearby cells together with antenna positions to obtain more accurate localization results.

Scoping. As mentioned, in the position-based version the scope will consist of a disk with a certain radius r around the position hint p. Because at the time when p was measured the object is out of range of the user's object sensor, we set r as $r = s_r + e_p$ where s_r denotes the range of the object sensor and e_p the positioning error. Similarly, if p is a cell id, the scope will consist of the object sensors served by the cell p.

Note that in this model, the object really lies within the computed scope. While this is not always true in reality, our simulation focuses on evaluating the system's performance for situations in which the search strategy is in fact correct. If in reality, the search strategy should fail in q percent of all cases, the resulting success rates can be extrapolated from the results we provide.

8 Evaluation

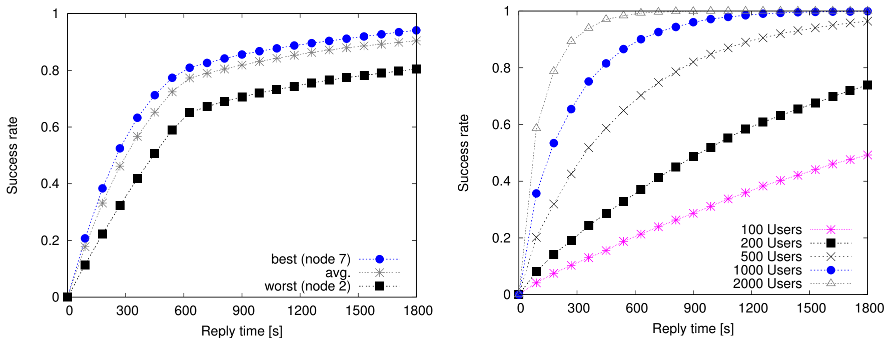
Using the simple scenario and the environment models described above, we aim to investigate several aspects of a future object sensing system. Foremost, given some scope, we want to confirm whether it is possible to find objects with reasonable success rates and small-enough overhead. Further, we aim to investigate how large cell-based scopes compare to position-based scopes and to a random query dissemination strategy which queries a certain fraction of all users. Moreover, we aim to gain insights into the sensitivity of the system's performance with regard to parameters such as user mobility, object sensing range, or chosen query timeouts.

8.1 Success Rate

In the first set of simulation runs, we investigated the query success rate observed with *position-based* scoping and *cell-based* scoping. Figure 5(a) shows the fraction of successful queries (upon which the queried sensors have located the object within 30 minutes), when the user-imposed limit $q_{\max}$ denoting the maximum number of queries is varied. Five different graphs show the results obtained with different positioning errors e_p (from e_p=50 m to e_p=200 m), cell-based scoping, and a random strategy where we distribute the query to a fraction of $q_{\max}/500$ of all users. For all graphs, the obtained success rate can be increased by raising the maximum number of queries $q_{\max}$ and reaches acceptable levels of above 90% with $q_{\max}$=200.

The number of messages q_{total} which were actually sent in the same runs is shown in Figure 5(b). As, by the definition of our protocol, the search area becomes larger with an increased positioning error, the required effort increases as well. Similarly, searching the coverage area of the whole cell where the object was left requires sending more messages before obtaining reasonable success rates. Note, however, how in Figure 5(b) the actual number of sent queries q_{total} at some point stops growing with the user-imposed limit of maximum queries $q_{\max}$. This is because with small enough scopes the object is found before the maximum message limit $q_{\max}$ is reached. Observe also, how the performance of cell-based scoping is comparable to a 200 m positioning error and even outperforms the latter in terms of q_{total}. This is because with position-based scoping and large positioning errors, many ineffective queries are sent to mobile devices which erroneously measured a position which was close to the position hint p.

Finally, as Figure 5(a) shows, any scoping performs better than a random strategy. Even if 40% of all users are queried, the success rate is still only around 60%. Needless to say, the communication effort of the random strategy is worst as it is proportional to the total number of users (not shown).



(a) Real-world measurements (b) Random waypoint simulations

Fig. 6. Cumulative density functions of reply times

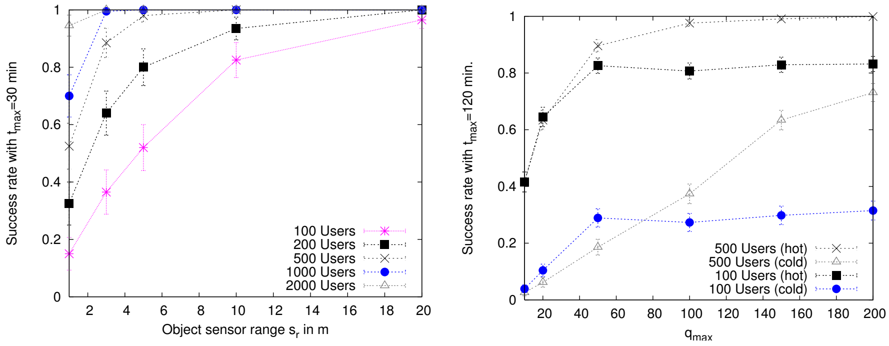
8.2 Timeout, Sensing Range, and Different Mobility Models

Apart from scoping, several other parameters may significantly influence the performance of the system.

The first such parameter is the timeout used for queries. Here, it is not clear, *when* the replies were received and whether a more adequate choice of the timeout (previously set to $t_{\max}$=30 min) can be made. Note that choosing an adequate timeout is particularly relevant when *object sensing itself* is considered a significant cost. Especially because in reality the object might be outside the chosen scope, it is important not to sense in vain for too long, but at the same time to issue a confident “not-found” reply. Further, the system performance is expected to vary with the user density. We show the interplay of these two parameters with position-based scoping in Figure 6(b). Each graph represents the cumulative density function of the reply time obtained after 5000 repeated simulation runs (each data point represents the fraction of requests answered within the given timeout) if no message limit $q_{\max}$ is imposed. As expected, the likeliness of finding the object increases with a longer timeout, but for high user-densities very short timeouts are already sufficient. Moreover, very good success rates can be obtained with $t_{\max}$=30 min, even with user densities down to 500 users/km^2. For lower densities longer timeouts must be used.

Observe that the graphs of Figure 6(a) measured in our office floor experiments, in which the actual user density was greater than 4000 users/km^2, are comparable with user densities of 500 to 200 users/km^2 in Figure 6(b). This is compatible with our earlier conjecture that the random waypoint simulation only models the “pedestrian” fraction out of the total users, and confirms that the approach to look at user densities which are smaller than in reality is valid.

A second important parameter, which is expected to have a large impact on the performance, is the sensing range of the employed object sensors. In the runs shown in Figure 7(a), we demonstrate the impact of the sensing range on the success rate of position-based scoping. With 2000 users/km^2, even a sensing range of 1 m yields acceptable results. As expected, however, the sensing range has a high impact. When designing a future system that shall be robust to small user densities, it seems worthwhile to invest in object sensing technology with a higher range.



(a) Sensing range (b) Campus mobility

Fig. 7. Varying sensing range and mobility

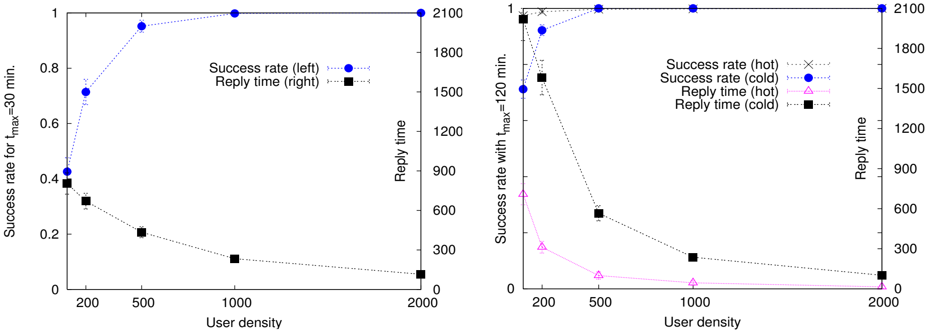
Finally, a third crucial parameter is the mobility of the system's participants. Here it is unclear whether the random waypoint model used is perhaps too optimistic. To analyze this, the simulation results of Figure 7(b) show the success rate with the campus mobility model when raising the message limit $q_{\max}$. We show four graphs for the cases in which the object was left in a hotspot or in the cold region with two different user densities. Because pause times in this model are quite long, we extended the query timeout $t_{\max}$ to 2 hours. Note, however, that the total number of queries remains limited to $q_{\max}$ and therefore the results remain comparable to earlier simulation runs shown in Figure 5(a). Here, for very small user densities, the success rate cannot be improved by raising $q_{\max}$ as the timeout remains the dominating constraint. For 500 users/km^2, however, the object can often be found with at most 200 messages even if it lies in the cold region.

8.3 Effects of Increasing User Density

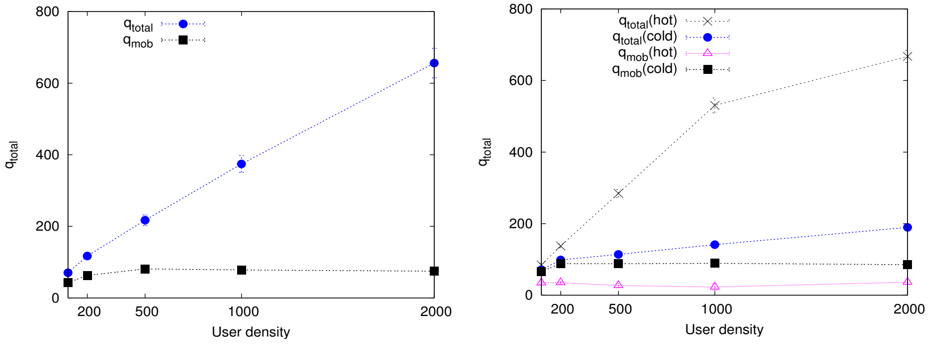
Additional lessons can be learned when observing our metrics' sensitivity to an increasing user density. These experiments were performed with cell-based scoping and are shown in Figure 8. We show the success rate and the query reply time while varying the user density in Figure 8(a), and analogously the results for the campus mobility model in Figure 8(b). The corresponding overheads are shown in Figure 8(c) and 8(d), respectively. Note that for these runs no limit $q_{\max}$ is set.

Both overhead figures include the total overhead q_{total} and the overhead due to user mobility q_{mob} included in the total. It is interesting to observe that q_{mob} does not increase with higher user densities. We explain this by the fact that the query reply time decreases with increased user density and therefore compensates for the expected increase in the mobility-based overhead. Quite differently, q_{init} (equal to $q_{\text{total}} - q_{\text{mob}}$) increases proportionally to the user density as the number of queries is not limited by a certain $q_{\max}$.

The main result here is that once the success rate is good, an increased number of messages is "wasted" towards lowering the reply time. In other words: waiting for users to move is more efficient than simply querying more users. As a consequence, if a higher reply time were acceptable, then the protocol can do with much less queries (e.g., by computing q_{init} as if the density were 500 users/km^2).



(a) Performance (random waypoint) (b) Performance (campus)



(c) Overhead (random waypoint) (d) Overhead (campus)

Fig. 8. Cell-based scoping when the user density is varied

Summary. Summing up, we observed that high success rates can be obtained with a range of different mobility patterns and scoping variants. Cell-based scoping, which is free from additional overhead in propagating object-sensor position information, proved to be particularly valuable. Finally, in certain circumstances the system may even work with very low user densities which represent a hundredth of the expected daytime population in an urban area.

9 Conclusion

In this paper we presented the architecture, design, and evaluation of an object search system relying on mobile phones as omnipresent object-sensing devices. Based on the ubiquitous mobile network infrastructure which is already in place, wide-area search for everyday objects becomes possible without incurring the high costs involved in instrumenting a larger environment with an object-sensing infrastructure.

Our system makes use of an unconventional approach, which relies on the participants' mobility in order to cover an essential portion of the users' space. We therefore spent significant effort on modeling and testing the circumstances in which such an object search system would be used in the large. The results are encouraging. In all our experiments, we could observe a high rate of successful queries, that is, of objects being found. While the time until a reply can be obtained varies with user mobility and density, our conjecture – that most of the time an object found event will be received eventually – was confirmed. Moreover, we could show that even in settings with high positioning errors or which rely solely on the observed cell id for localization, the total overhead for distributing an object search query remains acceptably low. While this does not change the basic fact that objects left in deserted places will not be found, we showed that for objects left within the users' space such a system is feasible.

In a broader context, this paper has analyzed the properties of the coverage obtained from user-carried sensors. By means of the average query reply time we observed with a certain sensing range, participant density, and mobility pattern (e.g., 30 minutes), we have quantified the time which must pass before a point-shaped phenomenon has been sufficiently covered. Therefore, the query reply time can be interpreted as the (reciprocal of the) maximum sampling frequency that our user-centric infrastructure can implement for a certain spot of the area under observation. If the phenomenon changes only insignificantly between samples, then coverage is sufficient. For example, recent work has mentioned measuring air quality or average noise levels [22] in urban areas. In such systems, the examined trade-offs between sensing range, maximum sampling frequency, and participant density are likely to re-appear.

In future work, we aim to collect real-user data using our prototype implementation (e.g., data on the participants' social networks) and use it to test the effectiveness of user-profile based heuristics which were not evaluated so far. Further, current work focuses on extending our model with an aggregated cost measure which integrates the costs of query dissemination with the costs of object sensing.

Acknowledgments. We would like to thank Michael Fahrmair, Friedemann Mattern, and Daisuke Ochi as well as our anonymous reviewers for their thoughtful comments and suggestions on draft versions of this paper, and Chie Noda for contributing to the presented work in earlier phases of our research collaboration. Moreover, we would like to thank Hans-Florian Geerdes for valuable advice on the Momentum dataset [18] and Christian Floerkemeier for pointing out suitable hardware options for the implementation of a future object search system.

References

[1] Yap, K.K., Srinivasan, V., Motani, M.: MAX: Human-centric search of the physical world. In: Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems (SENSYS'05), San Diego, CA, USA (2005)
[2] Decker, C., Kubach, U., Beigl, M.: Revealing the retail black box by interaction sensing. In: Proceedings of ICDCS 2003, Providence, RI, USA (2003)
[3] Want, R., Fishkin, K., Gujar, A., Harrison, B.: Bridging Physical and Virtual Worlds with Electronic Tags. In: ACM Conference on Human Factors in Computing Systems (CHI'99), Pittsburgh, USA (1999)
[4] RF Code, Inc.: Mantis^{TM} active RFID tags 433 MHz data sheet. `www.rfcode.com/data_sheets/433_mantis_tags.pdf` (2006)
[5] Wibree Technology. `www.wibree.com` (2006)

[6] BTnodes. www.btnode.ethz.ch (2006)

[7] Borriello, G., Brunette, W., Hall, M., Hartung, C., Tangney, C.: Reminding about tagged objects using passive RFIDs. In: Ubicomp 2004, Nottingham, England (2004)

[8] Shimizu, H., Hanzawa, O., Kanehana, K., Saito, H., Thepvilojanapong, N., Sezaki, K., Tobe, Y.: Association management between everyday objects and personal devices for passengers in urban areas. Pervasive 2005, Demonstration, Munich, Germany (2005)

[9] Gibbons, P.B., Karp, B., Ke, Y., Nath, S., Seshan, S.: IrisNet: An architecture for a world-wide sensor web. IEEE Pervasive Computing **2**(4) (2003)

[10] Pietzuch, P., Ledlie, J., Shneidman, J., Roussopoulos, M., Welsh, M., Seltzer, M.: Network-aware operator placement for stream-processing systems. In: Proceedings of the 22nd International Conference on Data Engineering (ICDE'06), Atlanta, GA, USA (2006)

[11] Engberg, S.J., Harning, M.B., Jensen, C.D.: Zero-knowledge device authentication: Privacy & security enhanced RFID preserving business value and consumer convenience. In: Proceedings of the 2nd Annual Conference on Privacy, Security and Trust (PST'04). (2004)

[12] Trossen, D., Pavel, D.: Building a ubiquitous platform for remote sensing using smartphones. In: Proceedings of the 2nd Annual International Conference on Mobile and Ubiquitous Systems: Networks and Services (MobiQuitous'05). (2005) 485–489

[13] Bolliger, P., Langheinrich, M.: Distributed persistence for limited devices. System Support for Ubiquitous Computing Workshop at Ubicomp 2006 (Ubisys 2006) (2006)

[14] Laasonen, K., Raento, M., Toivonen., H.: Adaptive on-device location recognition. In: Pervasive 2004, Vienna, Austria (2004) 287–304

[15] Smith, I., Consolvo, S., LaMarca, A., Hightower, J., Scott, J., Sohn, T., Hughes, J., Iachello, G., Abowd., G.D.: Social disclosure of place: From location technology to communication practices. In: Pervasive 2005, Munich, Germany (2005)

[16] Frank, C., Roduner, C., Noda, C., Kellerer, W.: Query scoping for the Sensor Internet. In: Proc. of IEEE International Conference on Pervasive Services (ICPS), Lyon, France (2006)

[17] Ferreira, L., Correia, L.M., Xavier, D., Vasconcelos, A., Fledderus, E.: Deliverable d1.4: Final report on traffic estimation and services characterisation. Technical report, Momentum Project, IST-2000-28088 (2003)

[18] Momentum: Models and simulation for network planning and control of UMTS. momentum.zib.de/data.php (2006)

[19] ETSI: Selection procedures for the choice of radio transmission technologies of the UMTS. Technical Report 3.2.0, European Telecommunications Standards Institute (1998)

[20] Kim, M., Kotz, D., Kim, S.: Extracting a mobility model from real user traces. In: Proceedings of INFOCOM 2006, Barcelona, Spain (2006)

[21] Le Boudec, J.Y., Vojnović, M.: Perfect simulation and stationarity of a class of mobility models. In: Proceedings of INFOCOM 2005, Miami, USA (2005)

[22] Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M.B.: Participatory sensing. Workshop on World-Sensor-Web (WSW'06): Mobile Device Centric Sensor Networks and Applications (2006)

Author Index

Abowd, Gregory D. 334
Aipperspach, Ryan 216
Anderson, Ken 216
Assad, Mark 55

Begole, James “Bo” 73
Bolliger, Philipp 351

Carmichael, David J. 55
Chalmers, Matthew 109

Ellingsen, Gunnar 234
Ellis, Carla 251

Fielding, Jeffrey J. 162
Fleisch, Elgar 268
Floerkemeier, Christian 198, 280
Frank, Christian 351
Fraser, Katie 180

Garzonis, Stavros 19
Gellersen, Hans 144

Hasvold, Per 234
Havinga, Paul 298
Henderson, Tristan 162
Henriksen, Eva 234
Hodges, Steve 280

Jacobson, Linda 73
Jenkins, Jam 251

Kapadia, Apu 162
Kay, Judy 55
Kellerer, Wolfgang 351
Kjærgaard, Mikkel Baun 316
Köhler, Moritz 334
Kotz, David 162
Krumm, John 127
Kummerfeld, Bob 55

Langheinrich, Marc 198
Linnhoff-Popien, Claudia 316

Maes, Pattie 1
Mainwaring, Scott D. 216
Mallinson, Hugo 280
Marin-Perianu, Mihai 298
Marin-Perianu, Raluca 298
Mayrhofer, Rene 144
Merrill, David 1
Metzger, Christian 268
Meyer, Jan 268
Morrison, Alistair 109

Nyholm, Heli 37

O’Malley, Claire 180
O’Neill, Eamonn 19

Page, Mitchell 91
Partridge, Kurt 73
Patel, Shwetak N. 334

Rodden, Tom 180
Roduner, Christof 198, 351

Sala, Matthias C. 73
Scheible, Jürgen 37
Scholl, Jeremiah 234
Scholten, Hans 298
Schwarzentrub, Beat 198
Stuntebeck, Erich P. 334
Summet, Jay W. 334

Tennent, Paul 109
Thompson, Peter 19
Thorne, Alan 280
Treu, Georg 316
Tröster, Gerhard 268
Tuulos, Ville H. 37

Vande Moere, Andrew 91

Warr, Andrew 19
Williamson, John 109
Woodruff, Allison 216