Proceedings of the 4th Workshop on Interacting with Smart Objects 2015

co-located with 20th ACM Conference on Intelligent User Interfaces (ACM IUI 2015)

Edited by: Dirk Schnelle-Walka [1], Stefan Radomski [1], Tobias Grosse-Puppendahl [2], Jochen Huber [3], Oliver Brdiczka [4], Kris Luyten [5], Max Mühlhäuser [1]

[1] Technische Universität Darmstadt, Germany, Karolinenplatz 5, 64289 Darmstadt, Germany
[2] Fraunhofer IGD, Fraunhoferstr. 5, 64283 Darmstadt, Germany
[3] Singapore University of Technology and Design & MIT Media Lab,
8 Somapah Road, Singapore 487372, Singapore
[4] Vectra Networks, 550 S Winchester Blvd #200, San Jose, CA 95128, United States
[5] Hasselt University - tUL - iMinds Expertise Centre for Digital Media,
Martelarenlaan 42, BE3500 Hasselt, Belgium

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Telecooperation Lab

Table of Contents

1 Preface

The increasing number of smart objects in our everyday life shapes how we interact beyond the desktop. In this workshop we discussed how the interaction with these smart objects should be designed from various perspectives. This year's workshop put a special focus on affective computing with smart objects, as reflected by the keynote talk.

1.1 Introduction

There is an ongoing trend to put computing capabilities into everyday objects, turning them into smart objects [4]. Well known examples range from smart kitchen appliances (smart coffee machines, smart knifes and cuttings boards) [1, 2] to smart (tangible) objects [3, 5] and even urban infrastructures [8].

While other venues have focused on the many technical challenges of implementing smart objects, far less research has been done on how the intelligence situated in these smart objects can be applied to improve their interaction with the users. This field of research poses unique challenges and opportunities for designing smart interaction. Smart objects typically have only very limited interaction capabilities. Yet, their behaviour exhibits an amazing amount of intelligence. More information about the previous workshops can be found on our website at http://www.smart-objects.org/.

Extending the topics of our previous workshops, this year's workshop emphasized affective computing with smart objects with a keynote talk by Jean-Claude Martin (https://perso.limsi.fr/wiki/doku.php/martin/accueil). Enabling objects to sense and react on human emotions broadens the acceptance and usefulness of such technologies. However, the physical restrictions in smart objects are very high, which sparked interesting discussions among all participants. Furthermore, the workshop focused on topics like user experience, sensing and actuation technologies, psychological aspects, and application scenarios with respect to smart objects.

1.2 Participants and Workshop Publicity

The workshop had an interdisciplinary appeal. Our participants originated from the areas of IUI, HCI, UbiComp, IoT and related areas like psychology and product design. The program committee comprised researchers who are active in these research areas and who, moreover, encouraged researchers to submit to this workshop.

Thereby, we ensured active participation in preparation and execution of the workshop. We especially encouraged young scientists and Ph.D. students to submit papers to explore their research topics with domain experts. The call for papers and participation was distributed through well-established mailing lists and websites in various research communities, including IUI, CHI, UIST, UbiComp, ITS and TEI. We also promoted the workshop through our website and OSNs.

Always held in conjunction with IUI, the first workshop took place in 2011 (Palo Alto) with following workshops in 2013 (Santa Monica) and 2014 (Haifa). These previous workshops were very successful, which helped us to attract other new participants to this workshop as well. The results of the workshop are made available on the workshop website as well as in the joint TU Prints proceedings.

1.3 Format

Our full-day workshop accepted submissions in the following three categories:

- position papers and posters (2 pages) focusing on novel concepts or works in progress,
- demo submissions (2 pages) and
- full papers (4-6 pages) covering a finished piece of novel research.

Our goal was to attract high-quality submissions from several research disciplines to encourage and shape the discussion, thus, advancing the research of interacting with smart objects. To stimulate discussion between the workshop participants we conducted a poster and demo session to spark further in-depth discussions on selected topics. We also collected topics during the workshop whereby we focused on combining complementary topics. As in previous workshops, this strategy lead to a lively and productive discussion during the remainder of the conference. We also summarized the outcome and published it on the workshops website in addition to the joint TU prints proceedings. This publication strategy attracted higher quality submissions, and increased the exposure of the workshop before and after the event.

1.4 Organizers and Program Committee

Most of the organizers were already members of the first three workshops on interacting with smart objects, held in conjunction with IUI 2011 [2], 2013 [6] and 2014 [7].

- **Dirk Schnelle-Walka** leads the "Talk&Touch" group at the Telecooperation Lab at TU Darmstadt. His main research interest is on multimodal interaction in smart spaces.
- **Jochen Huber** is an SUTD-MIT Postdoctoral Fellow at the MIT Media Lab, focusing oninteraction design for smart mobile projections and wearable technology.
- **Tobias Grosse-Puppendahl** is a PhD candidate at Fraunhofer IGD in Darmstadt. His research focuses on new ways of perceiving the environment with unobtrusive modalities like capacitive sensing.
- **Stefan Radomski** is a PhD candidate at the Telecooperation Lab at TU Darmstadt. His main research interest is about multimodal dialog managment in pervasive environments.
- **Oliver Brdiczka** is the area manager of Contextual Intelligence at Palo Alto Research Center (PARC). His group focuses on constructing models for human activity and intent from various sensors–ranging from PC desktop events to physical activity sensors–by employing machine learning methods.n
- **Kris Luyten** is associate professor at the Expertise Centre for Digital Media - iMinds, Hasselt University. His research focuses on engineering interactive systems, ubicomp, multitouch interfaces and HCI in general.
- **Max Mühlhäuser** is full professor and heads the Telecooperation Lab at TU Darmstadt. He has over 300 publications on ubicomp, HCI, IUI, e-learning and multimedia.

The list of program committee members is as follows:

- Bo Begole (Samsung, USA),
- Marco Blumendorf (DAI Laboratory, Germany),
- Aba-Sah Dadzie (University of Birmingham, United Kingdom),
- Fahim Kawsar (Bell Labs, Belgium),
- Alexander Kröner (Technische Hochschule Nürnberg, Germany),
- Germán Montoro (UAM, Spain),
- Patrick Reignier (Inria, France),
- Boris de Ruyter (Philips, Netherlands),
- Geert Vanderhulst (Alcatel-Lucent Bell Laboratories, Belgium) and
- Raphael Wimmer (Universität Regensburg, Germany).

PC members helped the organizers to publicize the event in more scientific communities and allow for a competent peer-review process. All submissions were peer-reviewed by at least two reviewers.

References

1. Filipponi, L., Vitaletti, A., Landi, G., Memeo, V., Laura, G., and Pucci, P. Smart city: An event driven architecture for monitoring public spaces with heterogeneous sensors. In Sensor Technologies and Applications (SENSORCOMM), 2010 Fourth International Conference on, IEEE (2010), 281–286.
2. Hartmann, M., Schreiber, D., Luyten, K., Brdiczka, O., and Mühlhäuser, M. Workshop on interacting with smart objects. In Proceedings of the 16th international conference on Intelligent user interfaces, ACM (2011), 481–482.
3. Kortuem, G., Kawsar, F., Fitton, D., and Sundramoorthy, V. Smart objects as building blocks for the internet of things. Internet Computing, IEEE 14, 1 (2010), 44–51.
4. Molyneaux, D., and Gellersen, H. Projected interfaces: enabling serendipitous interaction with smart tangible objects. In Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, ACM (2009), 385–392.
5. Molyneaux, D., Izadi, S., Kim, D., Hilliges, O., Hodges, S., Cao, X., Butler, A., and Gellersen, H. Interactive environment-aware handheld projectors for pervasive computing spaces. In Pervasive Computing. Springer, 2012, 197–215.
6. Schnelle-Walka, D., Huber, J., Lissermann, R., Brdiczka, O., Luyten, K., and Mühlhäuser, M. SmartObjects: Second IUI Workshop on Interacting with Smart Objects. In Proceedings of the 2013 ACM international conference on Intelligent User Interfaces, ACM (Santa Monica, CA, USA, Mar. 2013).
7. Schnelle-Walka, D., Huber, J., Radomski, S., Brdiczka, O., Luyten, K., and Mühlhäuser, M. Smartobjects: third workshop on interacting with smart objects. In Proceedings of the companion publication of the 19th international conference on Intelligent User Interfaces, ACM (2014), 45–46.
8. Shepard, M. Sentient City: Ubiquitous Computing, Architecture, and the Future of Urban Space. The MIT Press, 2011.

AmbLEDs: Implicit I/O for AAL Systems

Marcio Cunha
Department of Informatics
Pontifical Catholic University of
Rio de Janeiro
Rio de Janeiro, Brazil
mcunha@inf.puc-rio.br

Hugo Fuks
Department of Informatics
Pontifical Catholic University of
Rio de Janeiro
Rio de Janeiro, Brazil
hugo@inf.puc-rio.br

ABSTRACT

Ambient Assisted Living (AAL) applications aim to allow elderly, sick and disabled people to stay safely at home while collaboratively assisted by their family, friends and medical staff. In principle, AAL amalgamated with Internet of Things introduces a new healthcare connectivity paradigm that interconnects mobile apps and sensors allowing constant monitoring of the patient. By hiding technology into light fixtures, in this paper we present AmbLEDs, a ambient light sensing system, as an alternative to spreading sensors that are perceived as invasive, such as cameras, microphones, microcontrollers, tags or wearables, in order to create a crowdware ubiquitous context-aware implicit interface for recognizing, informing and alerting home environmental changes and human activities to support continuous proactive care.

Author Keywords

Intelligent Interface; Crowdware; Ambient Assisted Living; Smart Light; Internet of Things; Collaborative Systems, Collective Intelligence.

ACM Classification Keywords

H.5.2 User Interfaces (D.2.2, H.1.2, I.3.6).

INTRODUCTION

Driven by an aging population, rising health care costs, lack of professional staff and remote support in most developed countries, there is a growing demand to provide a better delivery of health and social care services for elderly, sick, convalescent and disabled people. Ambient Assisted Living (AAL) is a field of research focusing on IT support for healthcare, comfort and control applications for home environments. AAL facilities often require sensors, actuators and wearable devices, and generally require easy installation and low energy consumption. Current developments in wireless and mobile communications integrated with advances in pervasive and wearable technologies have a radical impact on healthcare delivery systems. Currently, the patients' continuous monitoring is considered the most relevant aspect in healthcare.

This paper aims to study how the Internet of Things (IoT), Autonomic Computing and Smart Lights may be used to provide a novel interface to provide ubiquitous connectivity with Visible Light Communication (VLC) while collect and analyze data for deciding and acting in AAL. This information is stored in the cloud and is accessed in a mobile collaborative environment used by patients and caregivers, to feed and train the system database and algorithms, to perform as a distributed task service to help divide caring responsibilities and training the system's automation. This new collaborative crowdware environment is called AmbLEDs. It is a new intelligent interface to detect activities of daily living (ADLs) and to trigger implicit interaction in AAL. Its technology is based on sensors and actuators embedded into LEDs fixtures shipped with code and enough processing power to make them autonomic based on situational context and connected to a collaborative system.

RELATED WORK

Several articles [1][2][3] show that healthcare professionals understand that the best way for detecting emerging medical conditions before they become critical is to look for changes in activities of daily living (ADLs). These routine activities comprise eating, getting in and out of the house, getting in and out of bed, using the toilet, bathing, dressing, using the phone, shopping, preparing meals, housekeeping, washing clothes and administering proper medications. For tracking the ADLs a distributed mobile infrastructure composed of sensors, actuators, microcontrollers, communication networks must be installed in the patients' homes.

A number of approaches to recognize ADLs in AAL have been considered in several papers [4][5][6]. One is the setup of a large and invisible infrastructure of sensors such as cameras and hidden microphones, presence sensors embedded into walls and ceilings, water pipes sensors and strain sensors under floorboards. Although this approach provides access to a wide variety of information, the cost of installing and maintaining it is usually very high.

Another approach is to use multiple low-cost sensors that cheapen the implementation and facilitate the setup throughout the home [3][7][8]. The disadvantage of this approach is that these sensors are obtrusive and ask for regular maintenance, like battery changes or corrections in their positions (e.g., sensors fixed on the doors of medicine cabinets, kitchen, refrigerator, walls, doors, etc.). According to Fogarty et al. [9], the elderly reject such sensors because they interfere with the look of their homes or create feelings of embarrassment or loss of privacy related to a need for assistance. A third approach is to use wearable devices [10], taking into account that the elderly, sick or convalescent may opt to avoid using such devices, by forgetting to use them every day or being unable to use them due to their health condition or disability.

Although others have written about the potential of sensor networks [11], we are unaware of work where the focus was on answering whether it is possible to recognize activities in diverse home settings using sensors embedded in light fixtures to be ubiquitous and pervasive to detect activities of daily living supported by a crowdware platform for system setup, configuration, and as an intelligent interface for the exchange and analysis of data in a collaborative fashion enabled by IoT, Autonomic Computing and VLC.

Autonomic Computing

To leverage the selective collection of information, AmbLEDs appropriates IoT technologies to provide data to the collaborative system in order to make possible the semiautomatic decision-making and information delivery anytime and anywhere. Caregivers and medical staff use collaborative data analysis to help the machine learning algorithms classify and recognize ADLs in the AAL. The idea of using the concepts of IoT is to provide relevant information in the correct format when and where needed, to establish communication between lights and to bridge the gap between the web and the real world.

However, to gather and access these data require different properties depending on their nature or even the role of the actor who is accessing it. Therefore, AAL may be viewed as a set of environments: hospitals, family homes, etc., each one containing different characteristics and requirements (emergency, security, monitoring, etc.). These characteristics make it necessary to build AmbLEDs applications as autonomic [12], with self-configuration, self-management, self-organization, self-healing and self-protection, to be flexible and adaptable to different environments and needs for users with different expertise and health condition.

Each element in autonomic computing must include sensors and actuators. The sensors responsibilities are to monitor the behavior of the system, while the actuators are used to enable any actions that may be necessary [13]. The process begins with the system collecting data from the sensors and comparing the observed situation in the environment with what it is expected. Then, the system analyzes the data and makes decisions on how to act, apart from medicine prescription. If an action is required, it is performed and its effects are monitored, creating an autonomic feedback control loop (Figure 1).

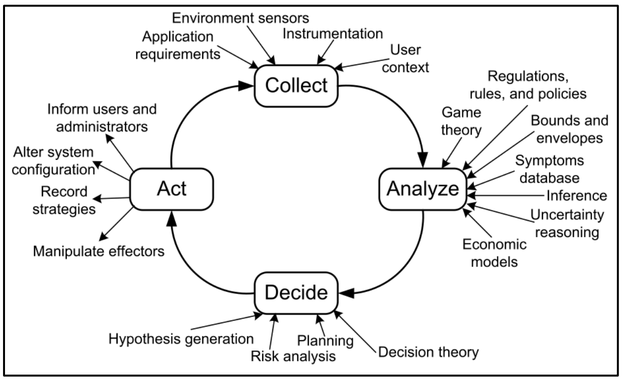


Figure 1 – Autonomic Feedback Loop

Autonomic computing also provides a reference knowledge base containing the system states, symptoms, references, rules and models to compare with the system observed behavior. In AmbLEDs this base is built and enhanced collaboratively by medical staff, families and caregivers, to describe variations and unique circumstances of each patient's condition or environment particularities, in order to build a collective intelligence with which to classify activities and routines [14].

SMART LIGHTS

Smart Lighting comprises a heterogeneous and multi-disciplinary area within illumination management, allowing integrating a wide set of sensor and control technologies, together with information and communication technologies. Its goal is to achieve higher efficiency and lower negative impact derived from the use of energy for illumination, in combination with enhanced intelligent functionalities and interfaces of lighting in the environment [20]. One of the principal Smart Lighting enablers has been the introduction and emergence of semiconductor based digital light sources such as LED (Light Emitting Diode) and next generation LED technologies such as Organic Light Emitting Diodes, also known as OLEDs or Solid State Light (SLL) sources [15].

Besides the advantage of low consumption (range 3-12 volts), LEDs do not depend on the lamp/socket paradigm, are smaller, resistant, and are able to emit different light spectrums to suit the user and lit environments needs, directly affecting the health, humor and productivity [15]. LEDs can also deliver optical and data communications (LiFi) or Visible Light Communication (VLC), and are becoming a new option to scalable and secure wireless communication [16].

LEDs and Sensors

Lights may be configurable in arrays containing many sensors, actuators and microcontrollers at their side, transforming them into a network of ubiquitous and pervasive sensors. For example, lights with moisture, temperature, infrared, noise, and gas sensors (carbon monoxide, butane and propane) enable AmbLEDs to capture useful data ensuring the safety and welfare of the elderly, sick, convalescent and disabled people. Temperature sensors on all light fixtures allow to assemble a thermal map for the whole house, enabling caregivers to remotely monitor the ideal temperature according to each patient's health, and to detect possible problems with the heating or cooling systems. AmbLEDs also come with an embedded speaker and a scent diffuser to give audio and olfactive feedback in order to play an ambient music with a specific scent paired with color changes in the light for therapeutic purposes or to trigger implicit interaction based on situational context.

Visible Light Communication

LEDs provide an almost ideal platform for VLC. An LED can emit and receive light at the same time (with multiplexing) [17]. In this research we propose to use the AmbLEDs as a LED-to-LED communication system for VLC. Such system can modulate light intensity with high frequencies so that the human eye is not affected by the light communication [17]. Light communication has several advantages: it is visible (in contrast to invisible radio communication), so it is easy to determine who can listen to (or receive) a message and will be used as a communication means between the lights themselves. In the midst of an emergency, if wireless communication is interrupted, lights with gas sensors will use the VLC, passing the command to the other lights that do not have such sensors to also blink red, and like a swarm, the information will pass on until all are flashing with the same color to warn the dweller.

Since AmbLEDs can operate as a virtual swarm, we can fragment the idea of a single light source. New services and APIs can use VLC to allow other devices receive the same lighting commands to overcome configuration overload and multi-device interactions: TVs, furniture, digital picture frames, refrigerators, etc. If someone gets into a room at night, not only will the secondary lights illuminate in the wall footers, but the TV could environmentally glow as well. Moreover, lights can ripple or flash in series across the room, when necessary to convey an idea of conduction to somewhere, for example, an individual route towards the kitchen to remind you to drink water or towards the apartment door at exercise time.

COLLABORATION IN AMBLEDS SYSTEMS

According to Chen et al. [18], we should consider the impact on patients and caregivers as part of AAL systems. By studying ADLs, we must not only address the physical, social and emotional needs of patients but also of their caregivers. Considering the caregivers' needs is especially important, since the burden of care may negatively impact their health and well being, leading to anxiety, stress or even death [18]. This same reasoning applies to the family, medical, social service, etc. Hence the collaborative environment is not only for the patient but also for the network that surrounds him.

The mobile collaborative environment serves as a repository of real-time information collected from AmbLEDs to provide data and information to feed the symptoms and ADLs classification databases of patients in the autonomic layer. The autonomic system, fed with the data captured by the sensors, supports activities in the collaborative environment, such as automatic alerts (with several risk levels) promoting communication, task distribution and its coordination, thus dividing the burden on all stakeholders involved in the process. The system also enables the exchange of experience among the community, providing psychological support among individuals who are experiencing the same difficulty, comparison of treatments, symptoms and experiences. This data exchange records the collaboration group's collective intelligence to feed the autonomic system database. This enables the algorithms training and fine-tuning for analysis and decision-making, based on the experiences and activities of hundreds or thousands of AmbLEDs, hence decreasing the chance of overtraining algorithms.

The collaborative environment is also used to investigate how the information captured by AmbLEDs can be worked in to provide the elderly, sick and disabled people, to be in touch with their families, relatives, and neighbors and meet some of their basic needs while respecting heir privacy and wishes more generally to be respected and not overtaken by well-meaning family members, social services or medical teams. The collaborative environment should provide some sort of self-help and a more formal external support, given that the system can also inform patients where their caregivers and family members are. The environment should also provide integration with the neighborhood and the local community to promote digital and social integration.

CONCLUSION

AmbLEDs provide a realistic solution to the problems expected as a result of the increase and population aging in all developed countries. At the center of these environments, the IoT is the layer that supports sensors' and objects' connectivity to the Internet, in order to monitor patient's daily lives activities. Autonomic computing offers intermediation for environments with self-management and self-adaptation to provide trust and security through the Autonomic Feedback Loop; and the mobile collaborative environment brings the collective intelligence of medical staff, family members and caregivers to the system algorithms, to support tasks distribution and provide awareness and context to the Autonomic layer.

Currently we are prototyping AmbLEDs with sensor integration and communication between devices with VLC and Internet connectivity. The second phase is the agent modeling and the Feedback Loop for the autonomic computing. The third phase is the modeling of the collaborative environment that will manage and store the data from the first and second phases, supporting collaboration, tasks distribution and building the caregiver's community collective intelligence. The fourth phase is the machine learning algorithms and classification tasks in the knowledge base from second and third phases. Finally, the fifth phase is the evaluation of the impacts of this new approach on real environments for the patients and caregivers.

The contribution of this work is to show how it is possible to assemble assisted environments that support not only the safety and independence of elderly, sick, convalescent and disabled people as well as relieve caregivers of stress and work overload. This work can be replicated to other areas that require monitoring and distribution of tasks, such as smart cities and factories, which also make intensive use of lights that can be used to control other activities, as well as a user intelligent interface and data collector.

REFERENCES

1. Lawton, M.P., E.M. Brody, 1989. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist, 9:179–186.
2. Rogers, W.A., Meyer, B., Walker N., Fisk, A.D., 1998. Functional limitations to daily living tasks in the aged: a focus groups analysis. Human Factors, 40:111–125.
3. Tapia, E.M., Intille, S.S., Larson, K., 2004. Activity Recognition in the Home Using Simple and Ubiquitous Sensors. In Proc. of PERVASIVE 2004, PP. 3001:158-174, Vienna Austria.
4. Abowd, G., Mynatt, E.D., 2000. Charting Past, Present, and Future Research in Ubiquitous Computing. ACM Transactions on Computer-Human Interaction (TOCHI): 29-58.
5. Chen, J., Kam, A.H., Zhang, J., Liu, N.,Shue, L., 2005. Bathroom Activity Monitoring Based on Sound. Proceedings of the International Conference on Pervasive Computing (Pervasive 2005): 47-61.
6. Rowan, J., Mynatt, E.D., 2005. Digital Family Portrait Field Trial: Support for Aging in Place. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2005): 521-530.
7. Beckmann, C., Consolvo, S., LaMarca ,A., 2004. Some Assembly Required: Supporting End-User Sensor Installation in Domestic Ubiquitous Computing Environments. Proceedings of the International Conference on Ubiquitous Computing (UbiComp 2004): 107-124.
8. Wilson, D.H. and Atkeson, C.G., 2005. Simultaneous Tracking and Activity Recognition (STAR) Using Many Anonymous, Binary Sensors. Proceedings of the International Conference on Pervasive Computing (Pervasive 2005): 62-79.
9. Fogarty, J., Au, C., Hudson, S.E., 2006. Sensing from the Basement: A feasibility Study of Unobstrusive and Low-Cost Home Activity Recognition. In: Proc. Of UIST: 91-100.
10. Ugulino, W.; Cardador, D.; Vega, K.; Velloso, E.; Milidiú, R.; Fuks, H. Wearable Computing: Accelerometers' Data Classification of Body Postures and Movements. Proc. of 21st Brazilian Symposium on Artificial Intelligence. Advances in Artificial Intelligence - SBIA 2012: 52-61.
11. Kim , E., Helal, D. C., 2013. Fuzzy Logic Based Activity Life Cycle Tracking and Recognition. ICOST 2013: 252-258.
12. Kephart JO, Chess DM., 2003. The vision of autonomic computing. IEEE Computer 2003: 41–50.
13. Horn, P., 2001. Autonomic computing: IBM perspective on the state of information technology. IBM T.J. Watson Labs, NY, 15thOctober 2001.Presented at AGENDA 2001.
14. Oliveira, A.I., Ferrada, F., Camarinha-Matos, L.M., 2013. An approach for the management of an AAL ecosystem. Healthcom, 2013 IEEE 15th International Conference o Digital Object Identifier: 601 – 605.
15. Karlicek, R.F., 2012. Smart lighting - Beyond simple illumination. Photonics Society Summer Topical Meeting Series, 2012 IEEE: 147 – 148.
16. Deicke, F.; Fisher, W.; Faulwasser, M.,2012. Optical wireless communication to eco-system. Future Network & Mobile Summit (FutureNetw):1 – 8.
17. Schmid S., Corbellini G., Mangold S., and Gross T., “LED-to-LED Visible Light Communication Networks,” in *MobiHoc, 2013 ACM*, Aug. 2013.
18. Chen, Y., Ngo, V., Park, S. Y., 2013. Caring for Caregivers: Designing for Integrality. Information and Communication in Medical Contexts February, San Antonio, TX, USA: 23–27.

Personalized interactive public screens

Paolo Cremonesi
Politecnico di Milano, DEIB
P.zza Leonardo da Vinci 32
Milano, Italy
paolo.cremonesi@polimi.it

Antonella Di Rienzo
Politecnico di Milano, DEIB
P.zza Leonardo da Vinci 32
Milano, Italy
antonella.dirienzo@polimi.it

Franca Garzotto
Politecnico di Milano, DEIB
P.zza Leonardo da Vinci 32
Milano, Italy
franca.garzotto@polimi.it

ABSTRACT

Crowded public indoor events such as expositions or fairs are nowadays common in large cities; a significant example of such kind of events is Expo 2015 that will take place in Milan during the current year. A lot of people usually crowd these shows and within this context experimental interactive media installations are gaining recognition as new art form. Considering the emerging need to support masses, matching offers to users and personalizing recommendations, there are other interesting possibilities of using the same digital infrastractures for contributing to a lively urban society, improving visitors experience. In order to achieve this, we developed an information service that integrates multiple (touch and touchless) interaction paradigms on personal devices and large public displays. It exploits personalization techniques in order to offer new engaging user experiences involving large amounts of multimedia contents.

Author Keywords

Motion-based touchless interaction, large screen, mobile devices, personalization, recommender system, mobile interaction.

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: Multimedia Systems, User Interfaces

INTRODUCTION

The themes of EXPO Milano 2015 spreading the sustainable food culture, shared with resources and tools, are having the admirable outcome of focusing international political debate on the world food problem and food security, and are also serving as a strategic litmus test for the city of Milan and its ambition to become an "ever-smarter city". Proposing new methodologies to suggest interactive and collective activities in this context is a consistent effect. Big electronic screens are a consolidated technology available in indoor and outdoor exhibitions for showing information to users. Their introduction into this kind of public context has become one of the most visible tendencies of contemporary urbanism and it is a great challenge to broaden their use by enriching their well-known advertising component and make them suitable for new challenging purposes. Their digital and networked nature makes these screening platforms an experimental visualization zone that can improve already existing application domains, like tourism and fashion domains.

Touch solutions enable direct inputs onto a display screen [8] [18]. An advantage of these types of interaction is that there is a direct relationship between what the eyes see and what the hands do [5], which has been shown to increase user satisfaction and initial acceptance [14]. The use of human touch, allowing the screen to function as an input pointing device [7], is also more intuitive and therefore easier for novice users to learn [15][18]. On the other hand, touch systems suffer from many limitations:

- Sterilizing before and after use. Usually, surfaces that are touched have to be made aseptic after being used.
- Maintenance costs resulting from wear. Surface treatment needs to be done to withstand a reasonable amount of the wear and tear that comes with the day-to-day life of a touchscreen device.
- Vulnerable to vandalism.

To address these issues and to protect the public screen from being dirty or destroyed, a new solution is the touchless interaction. Free-form [16] interactions have advantages as compared with touch based interfaces under the following conditions:

- Sterile environments. When using touchless interaction technologies, effort, time and money can be saved.
- Maintenance costs resulting from wear are greatly reduced.
- Vandalism-prone environments. When using cameras or similar devices for sensing fingers, hands, eyes etc. from a certain distance, it is possible to place the display device and the sensor at vandalism-proof places, e.g. behind glass walls. Among other things, this can be a useful solution for information systems at public places or for interactive store window displays.

However, the touchless interface is a barrier to users if they have to deal with huge amount of content. This issue is comparable to the TV remote control when you are browsing and you are forced to scroll through one laborious letter at a time just to get a title. It results inefficient and time consuming.

IUI 2015 Workshop: Interacting with Smart Objects, March 30th, 2015, Atlanta, USA.

Likewise, in these touchless solutions finding content relevant to users' interests is still an open issue, due to limited research effort spent in creating personalized catalogs which fit users' needs and filter the huge amount of data. This is the reason why we decided to apply personalization to touchless interaction.

After performing the above analyses, we ended up with our approach, which includes a set of core features. First, our User Experiences (UXs) integrate large public displays with personal devices (tablets or smartphones) and combine multiple interaction paradigms to explore the information through the screens, individually or in group. We exploit the interpretation of body movements (touchless gestural interaction) with Microsoft Kinect depth sensor as well as touch gestures on personal devices as control mechanisms (multi-touch remote interaction). In addition, we *personalize* contents on the large displays to increase users' engagement, to assist information finding, to facilitate contents exploration and to reduce information overload satisfying decision among the large number of points of interest (POIs), "recommending" the items that are likely to fit users' interests and characteristics. Personalization techniques are applied to create a user profile of the current user or group in front of the screen; based on the profile a recommendation engine suggest the most relevant items for the user/group. [1]

STATE OF THE ART

To the best of our knowledge, there are no works that describe touchless interactive screens with personalization.

With regard to *personalization*, the most similar application area is in E-Tourism domain. A large number of recommendation systems for e-tourism have been developed over the last few years [9] [22] and they target to provide personalized service recommendations to the users through their handheld and personal devices. COMPASS [19] is a mobile tourist application by Van Setten et al. which integrates a recommender system and a context-aware system. The Intrigue system [1] is an interactive agenda offered by a tourist information server that assists the user in creating a personalised tour along tourist attractions. This research focuses on planning and scheduling a personalised tour taking into account the location of each tourist attraction and the interests of the user. Console et al. [4] created a prototype system called MastroCARonte, which provides personalised services that adapt to the user and his context onboard cars. This research focuses on the effects of having such adaptive systems onboard cars. Several techniques have been exploited for POI recommendations. Ye et al. [21] tailor the collaborative filtering (CF) model for POI recommendations, aiming at providing a POI recommendation service based on a collaborative recommendation algorithm which fuses user preference to a POI with social influence and geographical influence.

With regard to *touchless interaction*, emerging technologies enable users to benefit content by allowing remote interaction with large displays situated in public areas. The PointScreen project [20] from the Fraunhofer IAIS (Institute for Intelligent Analysis and Information Systems) employs hand gesture recognition. PointScreen is a novel interface to manipulate digital artifacts touchlessly: the user navigates by pointing toward the screen. Instead of fiducial markers, PointScreen uses electric field proximity sensors. CyberGlove II [3] from Imation is an instrumented glove system that provides up to 22 high-accuracy joint-angle measurements. The gloves use proprietary resistive bend-sensing technology that transforms hand and finger motions into real-time digital data. Pfeiffer et al. [12] employ electrical muscle stimulation (EMS) and vibrotactile feedback to extend free-hand interaction with large displays. In the Communiplay system [10], screens are connected in a public display media space to create a shared touchless interaction. People can play with virtual objects, and people playing at one location can play with people at other locations. SMSlingshot system [6] consists of a portable input device in the form of a slingshot, enabling simultaneous message creation and shooting on a media facade. Moreover several interaction techniques using mobile devices have been proposed [11][2].

The problem with the previous systems is that they don't focus on providing personalized and user-specific information, resulting in a gap between what users require and what is provided to them. We, on the other hand, propose a system that aims to bridge this gap by capturing user personal information and providing him with highly customized content to be consumed in structured and meaningful ways on the public screens.

USAGE SCENARIO

Our system offers users a service that provides them with the content related to POIs in Milano. As a preliminary step, the system detects the presence of the user in front of the screen so that he/she may interact with it. Specifically, presence of the users is detected by employing Microsoft Kinect depth sensor within a range of 4 meters from the screen, which enables to track them as they enter or leave the scene. In this environment, the area in front of the screen can be virtually divided into three distinct areas [13][17], in order to recognize

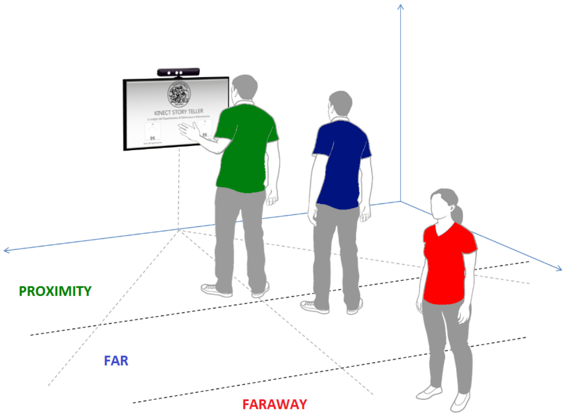


Figure 1. Possible user's positions in the space

[1] A demo video of the application is available online at **https://drive.google.com/file/d/0B47zW2f7aGiOU0dTcXVEdTZHYm8/view?usp=sharing**

the intention of the users: in fact, as you can see in Figure 1, the three areas of Proximity, Far and Faraway, represent areas of Interaction, Attraction and Simple Detection.

In particular,

1. the area with larger distance from the screen, defined *Faraway* area, extends beyond 3.5m away from the screen. When a user is located in this area, the system becomes aware of his presence, but does not react due to the great distance. This situation is associated to the starting or ending interaction state, while the system is in an idle state and it shows a series of video or animation, related to tourism domain, in a continuous cycle.

2. The intermediate zone, called *Far*, instead represents the state of users attraction; in fact, in this area the user is located at a reasonable distance in order to see what is happening on the screen, but still too far away to interact through gestures. This area extends between 1.5m and 3.5m and it is used to invite the user to come closer to the sceen.

3. Finally, the area closest to the screen, called *Proximity*, is the area in which the user has a distance such as to interact with the application through predefined gestures. When a user is located in this area, the system is found in an interactive state and shows the custom content for the user.

SYSTEM ARCHITECTURE

The general software architecture of our framework is depicted in Figure 2. Applications modules can be hosted both on users personal devices and on public large screens. The application on public screens exploits Microsoft's Kinect motion sensing technology to detect presence and implement touchless gestural interactions.

The **sensing and App Module** available in our system is a distributed module that includes both the sensing and the core modules of the Screen Application: the sensing module interacts with the depth image sensor in order to obtain the data of the users presence in front of the screen, while the core module deals with the page navigation and contents visualization. The Screen Application (SApp) is a .NET C# WPF Application that has been developed using a specific design pattern: the Model View View-Model (MVVM), which is the evolution of the traditional Model View Controller (MVC) pattern. The choice of developing with this specific programming language was dictated from the constraint of using the specific Microsoft Kinect depth image sensor.

The **Screen Data Controller** (SDC) is a centralized module that answers to the requests of the various Sensing and App Modules, i.e. it manages different SApp, one for each available screen; the communication can be done by two different mechanisms:

1. RESTful requests to specific endpoints, available from the SDC;

2. Bidirectional communication through a dedicated webSocket channel, opened during the SApps start.

In the first case the SApp always indicates its ScreenID, in order to be identified, while in the second case the SDC associates that ID to a specific channel, previously initialized. An NFC tag is present on every screen, in order to be read from the Mobile Application; this uses its own mobile connection (3G or WiFi) in order to invoke the SDC RESTful APIs. In addition, the SDC can interrogate an Image Analysis Web Service (IAWS), a Cloud Content Platform (ARTES) and a Cloud Recommender System (RecS) in order to ask respectively for the User Information, the available Contents and the Recommendations.

The **image Analysis Web Service** extracts the relevant information of the users, recognized into a specific photo, previously uploaded.

ARTES is the Content platform that collects all the information of a particular POI, merging all the available information from different social media sources.

The **Recommender System** is the module that incorporates the available recommendation algorithms.

Finally the **Social Advanced User Profile** (SocialAUP) manages the OAuth2 and allows the authentication and the users profiling.

PERSONALIZATION

To provide personalized recommendations about POIs, our system makes inferences about the users preferences, relying on the assumption that people having similar characteristics would prefer to choose the same POI. In order to achieve this purpose, our application performs implicit elicitation of users' information while they enter in the proximity area. Before providing access to the data (restaurants, POIs, reviews, photos, and other useful things we retrieved from TripAdvisor), the system captures users' images with the Kinect, then it invokes Betaface, an image processing web service that offers biometric measurements, analysis and extraction of facial features, such as age, gender, ethnicity and emotions. After collecting the demographic information of the participants (i.e. age and gender in Figure 3) it personalizes its recommendations applying the *collaborative filtering* algorithm, i.e. recommends POIs to users based on POI ratings given by people with similar profiles in the db. Finally the

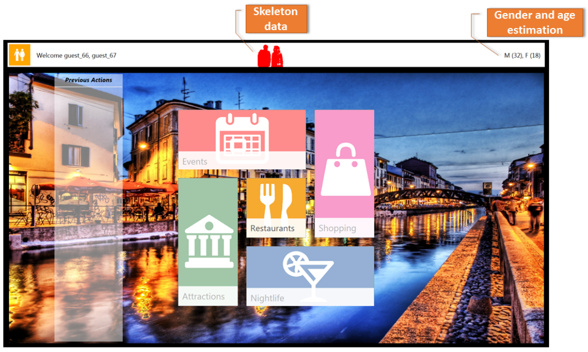


Figure 3. Estimation of demographic data

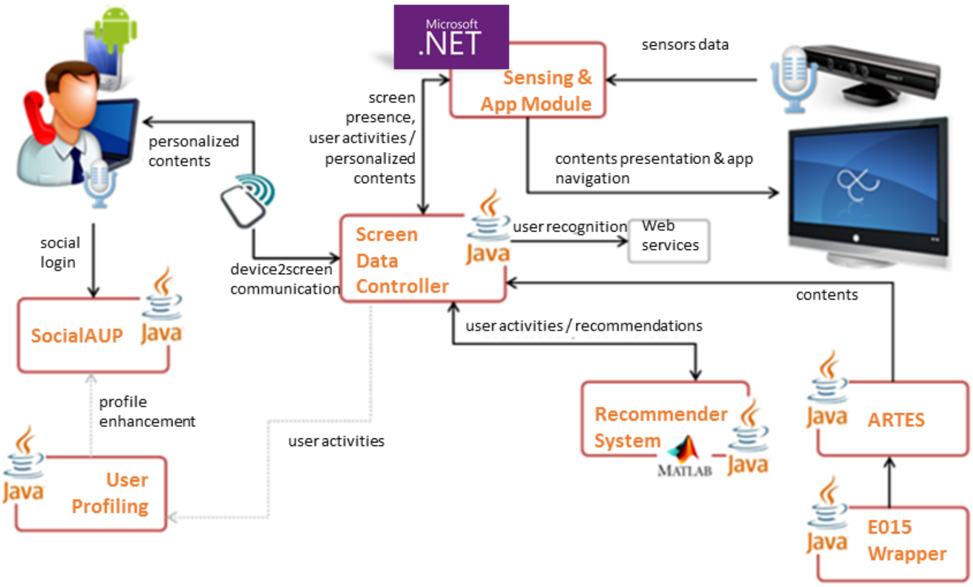


Figure 2. System Architecture

user/group can explore the personalized content with predefined gestures.

In case of several users in front of the screen, it's important to underline that the system is also capable to distinguish if the group is a *Family*, a *Couple* or a *Generic Group*. In particular, a Family is a group of two or more people where at least one user is a child (male or female under 16 years old); a Couple is a group of two people of similar age and different gender. First of all the recommender system selects those POIs that have been rated by users with the same age of those who are recognized in front of the screen. Then if the recognized group of users is a Family, the system takes into account those POIs with the word Child in the description. Instead, the system focus on POIs with the word Romantic for Couples and with the word Groups for Generic Groups.

When selecting a specific POI, the user can see all its details and in addition receives a content based recommendation (see Figure 4), i.e. items with similar characteristic to the chosen one, that is particularly valuable when a user encounters new content that has not been rated before.

There is also another option which allows a better personalization when the user identifies himself with his personal device via NFC, through a specific App developed for the purpose, providing his precise profile data, so that the Betaface phase is skipped. In this case, an *hybrid and dynamic recommendation* is provided which considers the social context as collaborative filtering (as it was explained before) plus a content-based filtering which generates recommendations based on what user has preferred in the past, from the attributes of the recommended items.

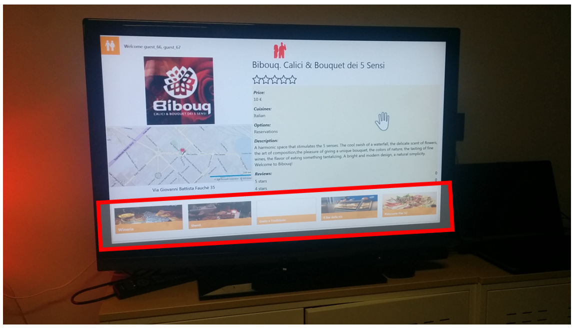


Figure 4. Content-based recommendation

INTERACTION

The content navigation is allowed through the following gestures:

- *Push to select*, to select a particular item (see Figure 7.a);
- *Swipe left / right* to browse through the items on the left / right (see Figure 7.b);
- *Grab and move* to ”grab” a particular content and scroll horizontally / vertically (see Figure 7.c);
- *Hands on head*, to level up in the hierarchy of the data structure as you can see in Figure 5 and in Figure 7.d;

The map showing the location of POIs is also interactive and it's possible to have a gesture-based interaction with it:

- *Zoom In / Out*, to zoom into the map as you can see in Figure 6 and in Figure 7.e;

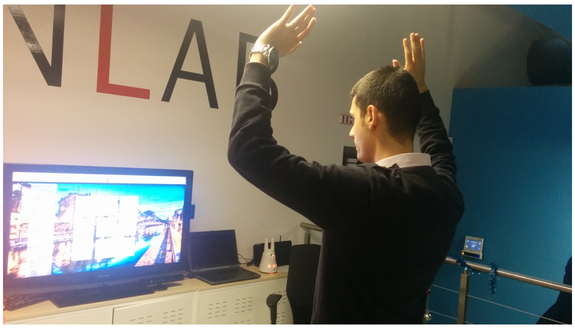


Figure 5. Gesture-based interaction

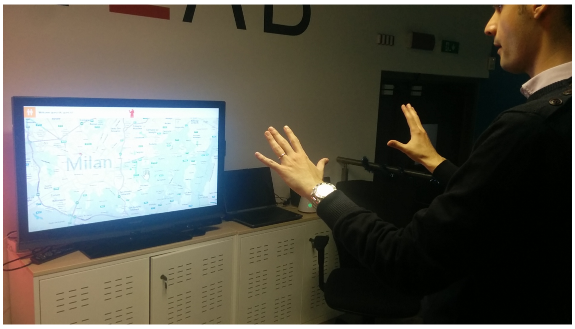


Figure 6. Interactive Map

- *Grab and move*, to scroll the map in all directions (see Figure 7.c);

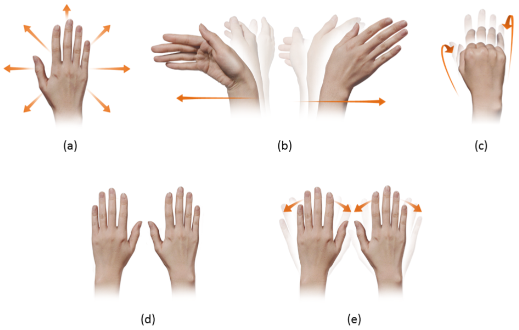


Figure 7. Gestures

CONCLUSIONS AND FUTURE WORKS

Our framework enables a personalized exploration of restaurants and points of interest in Milano. Some sets of items and metadata have been retrieved from public sites to validate the approach. The application is delivered in two versions, each one supporting a different interaction paradigm: touch-less gestural interaction or touch interaction using personal devices. After a set of incremental prototypes and iterative evaluations done in our labs, the first beta version of the application has been tested in a public space. In the cafeteria of our university, 12 users have been involved in assisted test sessions. Data gathered from observations and semi-structured interviews show a good degree of usability of both versions, and comparable levels of user satisfaction and engagement. Still, a number of aspects need to be improved: the aesthetics of the visual interface on the large display (e.g. layout and visual quality of multimedia contents) and the performance and precision of both the body movements processing and the age and gender interpretation component. From the experience gained so far, the research challenge of our application relies upon the intrinsic complexity of

- enabling touchless interaction with large amount of multimedia contents;
- supporting the interaction of single and group of users;
- combining all these aspects with personalization features.

A key issue of touchless interaction is how to support movement and mid-air gestures that are intuitive and natural. This problem is exacerbated when interacting with large amounts of multimedia contents, which need to be organized in non-trivial information architectures. In these contexts, the amount of interaction tasks is higher compared to those involved in the interaction with simple, linear information structures, and tasks are semantically more complex. In addition, supporting the interaction of both a single person and a group raises the issue of discriminating between individuals and groups movements and of interpreting them. Finally, an open problem is how to identify the characteristics of both individuals and groups that are appropriate for profiling purposes, to meet the algorithmic requirements of recommendation engines and to build effective recommendations. On the other hand, since the application screen proved to be robust, modular and customizable with its centralized management screen navigation, with its Kinect sensor management module and the total separation between application logic and user interface, we are working to revive the Screen App for another application domain, ie the sector of Fashion and Design.

ACKNOWLEDGMENTS

This work is supported by the EC commission through EIT ICT Labs Project 14183/2014 StreetSmart.

REFERENCES

1. Ardissono, L., Goy, A., Petrone, G., Segnan, M., and Torasso, P. Ubiquitous user assistance in a tourist information server. In *AH*, P. D. Bra, P. Brusilovsky, and R. Conejo, Eds., vol. 2347 of *Lecture Notes in Computer Science*, Springer (2002), 14–23.

2. Ballendat, T., Marquardt, N., and Greenberg, S. Proxemic interaction: designing for a proximity and orientation-aware environment. In *ACM International Conference on Interactive Tabletops and Surfaces*, ACM (2010), 121–130.

3. Bellucci, A., Malizia, A., Daz, P., and Aedo, I. Human-display interaction technology: Emerging remote interfaces for pervasive display environments. *IEEE Pervasive Computing 9*, 2 (2010), 72–76.

4. Console, L., Gioria, S., Lombardi, I., Surano, V., and Torre, I. Adaptation and personalization on board cars:

A framework and its application to tourist services. In *AH*, P. D. Bra, P. Brusilovsky, and R. Conejo, Eds., vol. 2347 of *Lecture Notes in Computer Science*, Springer (2002), 112–121.

5. Dul, J., and Weerdmeester, B. A. *Ergonomics for beginners: a quick reference guide*. CRC, 2008.

6. Fischer, P. T., Zllner, C., Hoffmann, T., Piatza, S., and Hornecker, E. Beyond information and utility: Transforming public spaces with media facades. *IEEE Computer Graphics and Applications 33*, 2 (2013), 38–46.

7. Greenstein, J. Pointing devices. *Handbook of Human-Computer Interaction* (1997), 1317–1348.

8. Harvey, C., Stanton, N. A., Pickering, C. A., McDonald, M., and Zheng, P. To twist or poke? a method for identifying usability issues with the rotary controller and touch screen for control of in-vehicle information systems. *Ergonomics 54*, 7 (2011), 609–625. PMID: 21770749.

9. Kabassi, K. Personalizing recommendations for tourists. *Telematics and Informatics 27*, 1 (2010), 51 – 66.

10. Müller, J., Eberle, D., and Tollmar, K. Communiplay: a field study of a public display mediaspace. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, ACM (2014), 1415–1424.

11. Pears, N., Jackson, D., and Olivier, P. Smart phone interaction with registered displays. *IEEE Pervasive Computing 8*, 2 (2009), 14–21.

12. Pfeiffer, M., Schneegass, S., Alt, F., and Rohs, M. Let me grab this: A comparison of ems and vibration for haptic feedback in free-hand interaction. In *Proceedings of the 5th Augmented Human International Conference*, AH'14, ACM (New York, NY, USA, 2014).

13. Prante, T., Röcker, C., Streitz, N., Stenzel, R., Magerkurth, C., Van Alphen, D., and Plewe, D. Hello. wall–beyond ambient displays. In *Adjunct Proceedings of Ubicomp*, Citeseer (2003), 277–278.

14. Rogers, W. A., Fisk, A. D., McLaughlin, A. C., and Pak, R. Touch a screen or turn a knob: Choosing the best device for the job. *Human Factors 47*, 2 (2005), 271–288.

15. Rydstrom, A., Bengtsson, P., Grane, C., Brostrm, R., Agardh, J., and Nilsson, J. Multifunctional systems in vehicles: a usability evaluation. In *Proceedings of CybErg 2005, the Fourth International Cyberspace Conference on Ergonomics. International Ergonomics Association*, Thatcher, A., James, J., Todd A. (2005).

16. Saffer, D. *Designing Gestural Interfaces: Touchscreens and Interactive Devices*. O'Reilly Media, Inc., 2008.

17. Streitz, N. A., Röcker, C., Prante, T., Stenzel, R., and van Alphen, D. Situated interaction with ambient information: Facilitating awareness and communication in ubiquitous work environments. In *Tenth International Conference on Human-Computer Interaction (HCI International 2003)*, Citeseer (2003).

18. Taveira, A. D., and Choi, S. D. Review study of computer input devices and older users. *Int. J. Hum. Comput. Interaction 25*, 5 (2009), 455–474.

19. van Setten, M., Pokraev, S., and Koolwaaij, J. Context-aware recommendations in the mobile tourist application compass. *Adaptive Hypermedia and Adaptive Web-Based Systems* (2004), 235–244.

20. Wittenberg, T., Münzenmayer, C., Küblbeck, C., and Ernst, A. Method and system for recognising an object, and method and system for generating a marking in a screen representation by means of a non-contact gesture-controlled screen pointer, Jan. 18 2012. EP Patent App. EP20,100,717,087.

21. Ye, M., Yin, P., Lee, W.-C., and Lee, D.-L. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, ACM (New York, NY, USA, 2011), 325–334.

22. Yu, C.-C., and Chang, H.-P. Personalized location-based recommendation services for tour planning in mobile tourism applications. In *Proceedings of the 10th International Conference on E-Commerce and Web Technologies*, EC-Web 2009, Springer-Verlag (Berlin, Heidelberg, 2009), 38–49.

Mining users' online communication for improved interaction with context-aware systems

Preeti Bhargava
University of Maryland
College Park, MD, USA
prbharga@cs.umd.edu

Oliver Brdiczka
Vectra Networks, Inc.
San Jose, CA, USA
brdiczka@acm.org

Michael Roberts
Palo Alto Research Center
Palo Alto, CA, USA
michael.roberts@parc.com

ABSTRACT

With the advent of the internet, online communication media and social networks have become increasingly popular among users for interaction and communication. Integrating these online communications with other sources of a user's context can help improve his interaction with context-aware systems as it enables the systems to provide highly personalized content to both individual and groups of users. To this end, a user's communication context (such as the people he communicates with often, and the topics he discusses frequently) becomes an important aspect of his context model and new frameworks and methodologies are required for extracting and representing it. In this paper, we present a hybrid framework derived from traditional graph based and object oriented models that employs various Natural Language Processing techniques for extracting and representing users' communication context from their aggregated online communications. We also evaluate the framework using the email communication log of a user.

ACM Classification Keywords

C.3.2 [Special-Purpose and Application-based Systems]: Real-time and embedded systems

General Terms

Algorithms; Design; Experimentation; Performance

Author Keywords

Context-awareness; Context modeling and representation; Communication media; Social networks

INTRODUCTION

With the advent of the internet, online communication media and social networks have become increasingly popular among users for interaction and communication. Users often have several email accounts and profiles on various social networking sites. Scoble and Israel [9] claim that integrating social media with other sources of contextual information (such as smartphones) and inferring the user's context (such as preferences) from it can enable the next generation of applications that will provide highly personalized content as *"It is in our online conversations that we make it clear what we like, where we are and what we are looking for."*

Thus, a user's *communication context* (such as the people he communicates with often, and the topics he discusses frequently on various online communication and social media) becomes an important aspect of his context model. This context can be employed to improve the interaction between him and a smart object such as a context-aware system and can be mined from his online and social network communications. It enables the system to infer his interests and provide him with personalized content in several domains. Moreover, the system can incorporate additional cues or context such as location to further enhance its capabilities. For instance, it can determine other users with whom the current user interacts more often than others via online media. Whenever the system determines that they are co-located, it can recommend content that will be of common interest to all of them based on their mutual interaction. Ultimately, this enables the context-aware system to provide targeted content to both individual and groups of users.

Oftentimes, a user expresses a subset of his interests on each social medium or network, that he participates in, and with each of his friends. Moreover, these subsets can be mutually exclusive. For instance, consider a user who has been involved in email chains and Facebook discussions with several friends, where each discussion is on a different topic - sports, music, food and movies etc. To infer a wide spectrum of his interests, these interactions have to be aggregated. However, a comprehensive framework to extract and represent this context from his aggregated online communications has not been developed so far. In this paper, we present such a hybrid framework which is derived from traditional graph based and object oriented models and employs various Natural Language Processing techniques. We also evaluate the framework using the email communication log of a user.

RELATED WORK

Since our work spans several ideas, we have organized the related work into three subsections. We highlight their shortcomings as well as differences with our approach.

Context Modeling and Representation

A significant amount of research has been carried out in context modeling and representation. Several surveys such as those by Bettini et al. [1] and Bolchini et al. [3] have summarized the different types of context models (key-value, mark-up schema based, logical, object-oriented, graphical, ontological) in general use today. Of these, ontological models such as SOUPA [4] are considered most expressive as they promote knowledge sharing and reuse across different applications and thus enhance

IUI 2015 Workshop: Interacting with Smart Objects, March 30th, 2015, Atlanta, USA

Node Type	Attributes	Information
User	id, name	Current user in session
UserProfile	id, name, type	User's profile on online media or social network
Entity	name, type	Any other user or person
Time	value, type	Time of any granularity
Communication	Ranked list of topics, Source, Raw contents, Timestamp of last update	User's communique with other users
ElectronicInfo	value, type	Email address, URL

Table 1. Node Types, their attributes and the information they represent

Edge Type	Information	Weighting criteria	Direction
Connection	Connection(s) such as friend/contact etc.	Number of connections	User ⟶ User/Entity
Contribution	Contribution to a communication	Number of contributions	User ⟶ Communication

Table 2. Directed Weighted Edge Types and the information they represent

Edge Type	Information it represents
Is-a	Superclass/subclass
Has-a	Ownership
Has-member/Member-Of	Membership
Relationship	Any other relationship

Table 3. Directed Unweighted Edge Types and the information they represent

their interoperability. However, the major disadvantage of ontologies is that the standard languages used to express them can be slow and intractable. Hence, a hybrid approach combining one or more of the context model types is often considered more practical for general use.

PersonisJ [6] is an example of such a hybrid framework which models users based on an ontology encoded in a JSON (key-value) format. However, each ontology has to be defined on a per application basis and the only pre-defined context that exists is 'location'. On similar lines, we use a hybrid model which combines the resuability and extensibility of object oriented models with the flexibility and adaptability of graphical approaches.

Modeling and representing user interests

The Friend-of-a-Friend (FOAF)[1] ontology supports limited modeling of interests by representing them as pages on topics of interest. The Semantically Interlinked Online Communities (SIOC)[2] ontology too has limited support for representing user interests through the sioc:topic property, which has a URI as a value. However, the specifications of these ontologies do not mention how the topics are extracted or inferred.

Modeling communities and roles

Modeling communities from communication media is a research area that has received attention in recent years. MacLean et al. [7] developed a system for constructing and visualizing a user's social topology from his personal email but did not discover conversation topics. Mccallum et al. [8] demonstrated that email conversation topics can be discovered based on social network structure and can be used to predict people's roles. Dev [5] presents a user interaction based community detection algorithm for online social networks.

Even though our framework shares the same theoretical foundation with the aforementioned works, we focus on extraction and representation of communication context to augment a user's context model rather than community detection and role prediction.

FRAMEWORK AND ALGORITHM FOR EXTRACTION AND REPRESENTATION OF COMMUNICATION CONTEXT

We extract and represent a user's communication context as a part of a graphical user model called the *Semantic Graph*. The Semantic Graph is composed of instances of different primitive node and edge types that higher-level application specific classes can inherit from. A user's context can be completely represented in a context-aware system by means of the graph's nodes and edges. Each graph is formed by the union of several different sub-graphs, where each sub-graph represents a different aspect of the user's context model such as the user's physical context or activities, spatial context and location, communication, interests, etc.

In this paper, we focus on generating the communication sub-graph of a user's Semantic Graph which represents his extracted communication context. Tables 1, 2 and 3 show the different node and edge types in this sub-graph which have been derived from the primitive nodes and edges of the Semantic Graph.

We employ Algorithm 1 to generate this sub-graph. The input to the algorithm is the user's communication log or "journal" over a window of time. The log consists of one or more journal entries, which are essentially *communication threads*, in chronological order. These threads may be sourced from multiple online communication and social media feeds such as Gmail, Facebook etc. Each log entry represents a thread and has the following components: participants (users and other entities), contents, topics, timestamp of the last updates made by the participants, and number of messages/words contributed by the participants. Each thread can represent an email exchange, post/status update, or message chain and contains only the aggregated information formed by processing the exchange. Table 4 shows a sample communication log between a user and her friends on Gmail[3].

As explained in Algorithm 1, we first generate a set of topics for each individual thread using topic modeling based on Latent Dirichlet Allocation [2]. We then cluster similar threads together by determining the nearest neighbor node, from among the existing communica-

[1] http://xmlns.com/foaf/0.1/ [2] http://rdfs.org/sioc/spec/

[3] Due to space and privacy constraints, only the thread topics and not the contents are shown.

Algorithm 1: Algorithm for extracting and representing user's communication context

```
Input: User's communication journal or log for a window of time
Output: Communication sub-graph of the Semantic Graph for the user
Remove all stop words from the contents of each thread;
foreach thread do
    Generate a set of topics for its contents using topic modeling;
    // Cluster all the similar threads together
    Determine the nearest neighbor for the thread using cosine similarity based on tf-idf;
    if no nearest neighbor node is found then
        • Create a new communication node between the participants and set its contents as the contents of the thread;
        • Add the list of thread topics to the node with frequency = 1 for each topic;
        • Timestamp of last update for the node ⟵ Timestamp of the last update to the new thread;
    else
        • Append the contents of the thread to the existing contents of the nearest neighbor node;
        • Update the list of topics and their frequency of occurrence in each node by adding the topics of the new thread: If
          a topic has occurred previously in the node, increment the occurrence field, otherwise add the new topic to the list;
        • Sort the list of topics based on the frequency of occurrence to generate a ranked list of topics;
        • Timestamp of last update for the node ⟵ Timestamp of the last update to the new thread if it is later;
end
foreach thread that is coalesced into an existing communication node do
    // Update the existing contribution edges from the participants
    • Weight of each edge ⟵ Add the contribution (messages/words etc.) of the participant to the thread;
    • Timestamp of last update ⟵ Timestamp of the last contribution made by the participant to the thread ;
end
foreach thread that is created into a new communication node do
    // Add a contribution edge instance from each participant in the thread to the new node
    • Create a new contribution edge from each thread participant to the newly created communication node;
    • Weight of each edge ⟵ Contribution (messages/words etc.) of the participant to the thread;
    • Timestamp of last update ⟵ Timestamp of the last contribution made by the participant to the thread;
end
// Update communication matrix for the users
$\forall$ Participant$_i$ and Participant$_j$ $\in$ Participants
if a new communication node has been created between Participant$_i$ and Participant$_j$ then
    C[Participant$_i$, Participant$_j$] ⟵ C[Participant$_i$, Participant$_j$] + 1;
return The communication nodes, contribution edges and the communication matrix as the communication
sub-graph of the semantic graph;
```

tion node instances between the thread participants, for each thread. In our current implementation, the nearest neighbor is determined using cosine similarity based on tf-idf (term frequencies and inverse document frequencies) where the threads are treated as documents. A threshold of 0.293 (1 - cos 45 °) is generally considered an appropriate threshold for cosine similarity and we use that. We will explore other alternative techniques, for determining similarity between nodes, in the future.

If no nearest neighbor is found for a thread, we create a new communication node instance, add the thread's contents to it and set the node's topics as the list of thread topics with each topic having frequency 1. The node is timestamped with the last updated timestamp of the thread. We also add weighted contribution edge instances to it from each of the participant node instances and time-stamp them based on the last contribution made to the thread by each participant. The weight of each contribution edge is computed based on the contribution of that participant to the thread. If a nearest neighbor node is found, we append the contents of the new thread to the existing contents of the node, and update the ranked list of topics as well as the timestamp of the node. Additionally, we update the weights and timestamps of the contribution edges.

We also maintain a *communication matrix C* for the users where each entry C[Participant$_i$,Participant$_j$] = Number of communication nodes between Participant$_i$ and Participant$_j$. The communication nodes, contribution edges and communication matrix are returned as the communication sub-graph for the user.

USE CASE FOR EVALUATION

Figure 1 shows the communication sub-graph of a user's semantic graph extracted from the threads in Table 4 using our framework and algorithm. It consists of two communication nodes - node # 1 between participants - current user 'XYZ', friends # 1 and # 2, derived from threads # 1 - 5 and node # 2 between participants - user and friend # 1, derived from thread # 6. The last updated time for each node is also shown and is based on the last updated time of the newest thread coalesced into that node. In the current implementation, the strength

Thread	Participants	Thread Topics	Last Updates	# of messages
1	User, Friends #1, # 2	indian food, food, restaurant, palo alto	June 9, June 10, June 12	21,7, 7
2	User, Friends #1, # 2	food, free, restaurant	June 13, June 13, June 11	20,2, 1
3	User, Friends #1, # 2	food, restaurant, meet, work,	June 17, June 17, June 18	5,2, 1
4	User, Friends #1, # 2	indian food, apartment, car	June 17, June 17, June 18	2,1, 1
5	User, Friend #1	food, apartment, car	June 19, June 19	2, 2
6	User, Friend #1	science fiction movie, friday, theater	June 24, June 25	2, 2

Table 4. Communication log for a user

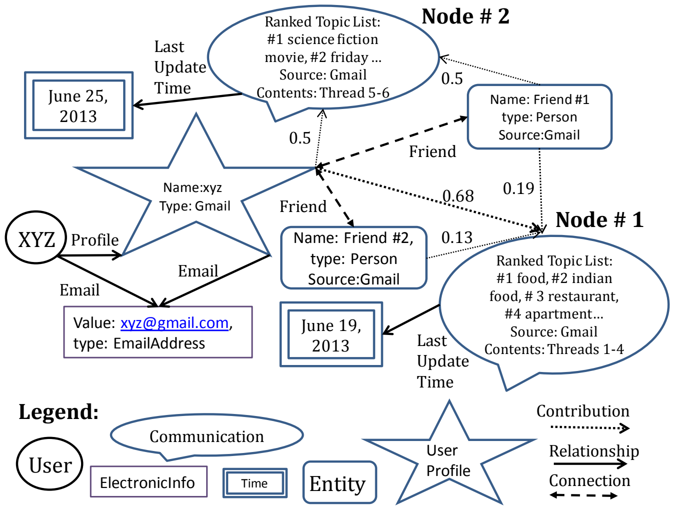

Node # 2
Ranked Topic List: #1 science fiction movie, #2 friday ... Source: Gmail Contents: Thread 5-6
Last Update Time
June 25, 2013
0.5
0.5
Name: Friend #1 type: Person Source:Gmail
Friend
Name:xyz Type: Gmail
Friend
0.68
0.19
Node # 1
XYZ
Profile
Name: Friend #2, type: Person Source:Gmail
0.13
Email
Email
Ranked Topic List: #1 food, #2 indian food, # 3 restaurant, #4 apartment... Source: Gmail Contents: Threads 1-4
Value: xyz@gmail.com, type: EmailAddress
June 19, 2013
Last Update Time
Legend:
Communication
User
ElectronicInfo
Time
Entity
User Profile
Contribution
Relationship
Connection

Figure 1. Communication sub-graph of a user's Semantic Graph extracted using our framework and algorithm

of the contribution edge from each participant is calculated based on the % of all messages contributed by them but other alternative approaches based on the number of words can also be used.

As shown in Figure 1, the top ranked topics between user 'XYZ' and her friends are 'food', 'indian food' etc. 'XYZ' and friend # 1 communicated more often on these topics and hence, their contribution edges have a higher weight (represented by edge width) while the contribution from friend # 2 has a lower weight. Since 'XYZ' and friend #1 have 2 communication nodes between them, the corresponding communication matrix entry, C['XYZ', Friend #1], is 2. The communication matrix entry for 'XYZ' and friend # 2, C['XYZ', Friend # 2], is 1 since they have one communication node between them.

Thus, a user's communication context can be easily represented using this framework. As more communication threads from other social media are integrated, the sub-graph incorporates aggregated information from them and expands dynamically. Since, we also record the source and timestamp of the communique, issues like provenance and timestamping are taken care of. The framework can be easily extended to derive more node and edge types as required, thus, making it expressive and extensible. It does not target any specific domain and is intended to be general and universally applicable. Furthermore, the user's extracted communication context can now enable a context-aware system to infer a user's preferences (such as 'Indian food' and 'Science fiction movies' in this case). This, in turn, helps personalize the information provided by it to the user.

CONCLUSION AND FUTURE DIRECTIONS

In this paper, we demonstrated that a user's communication context can help improve his interaction with a smart object such as a context-aware system as it enables the system to infer his interests and provide highly personalized content to both individuals and groups of users. We presented a hybrid framework and algorithm for extracting and representing a user's communication context from his aggregated online communications. This framework is derived from traditional graph based and object oriented models and employs various Natural Language Processing techniques. It has been incorporated as a part of a graphical user model called the Semantic Graph. We also evaluated the framework using the email communication log of a user. Our next step is to perform its real-time validation with several users to evaluate its scalability, flexibility and universality.

REFERENCES

1. Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., and Riboni, D. A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing 6*, 2 (2010), 161–180.
2. Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent dirichlet allocation. *The Journal of machine Learning research 3* (2003), 993–1022.
3. Bolchini, C., Curino, C., Quintarelli, E., Schreiber, F., and Tanca, L. A data-oriented survey of context models. *ACM SIGMOD Record 36*, 4 (2007), 19–26.
4. Chen, H., Perich, F., Finin, T., and Joshi, A. Soupa: Standard ontology for ubiquitous and pervasive applications. In *The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services* (2004).
5. Dev, H. A user interaction based community detection algorithm for online social networks. In *Proceedings of the ACM SIGMOD international conference on Management of data* (2014).
6. Gerber, S., Fry, M., Kay, J., Kummerfeld, B., Pink, G., and Wasinger, R. *PersonisJ: mobile, client-side user modelling.* Springer, 2010.
7. MacLean, D., Hangal, S., Teh, S. K., Lam, M. S., and Heer, J. Groups without tears: mining social topologies from email. In *Proceedings of the 16th ACM international conference on Intelligent user interfaces* (2011).
8. McCallum, A., Wang, X., and Corrada-Emmanuel, A. Topic and role discovery in social networks with experiments on enron and academic email. *J. Artif. Intell. Res.(JAIR) 30* (2007), 249–272.
9. Scoble, R., and Israel, S. *Age of Context: Mobile, Sensors, Data and the Future of Privacy.* CreateSpace Independent Publishing Platform, 2013.

Work in Progress
Modality Selection Based on Agent Negotiation for Voice Search on Smartphones Paired With Wearable Accessories: A Preliminary Architectural Design

W. L. Yeung
Lingnan University, Hong Kong
wlyeung@ln.edu.hk

ABSTRACT

Intelligent wearable accessories are becoming popular with smartphone users. They bring more modality options and convenience to our interaction with smartphone applications. Real benefits come when applications can adapt intelligently to the new and existing interface options. This paper suggests how such adaptation can be realised in a voice search application using a multi-agent architecture. A preliminary design of the architecture is presented here together with an outline of our future work.

Author Keywords

Smartwatch; multimodal fission; multi-agent architecture; contract net protocol.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

With Apple Watch, Pebble Smartwatch and an increasing number of Android Wear device models becoming available, mobile information retrieval is facing new opportunities and challenges. As extension of smartphone functionality, these wearable accessories also present to us new ways to interact with mobile applications and services. In particular, multimodal interaction is set to gain more traction with wearable accessories [4].

Before Apple introduced Siri in 2011, mobile search was already growing rapidly as the number of mobile phone users increased. According to an European study [1], active mobile search users accounted for only 8% of the mobile user population in the late 2010's. By 2013, 58.7% of smartphone users accessed search while 73.9%of tablet users access search [2]. When searching for location-based information, search engines can filter and rank search results based on the user's location. A study on mobile search for location-based information showed that the average number of clicks decreased and the relevance of the first-ranked result increased with location-based mobile search [9].

In the past, mobile search as a means of Internet information retrieval was hampered by the limited size of both the input keyboard and the display on a typical smartphone. Voice search using Siri or Google Now is becoming increasingly popular as a means of information retrieval on smartphones. The use of wearable accessories such as Apple Watch renders mobile voice search interaction more compelling.

While performing a voice search on a proper smartphone has its pros and cons (e.g. hands-free convenience vs. voice recognition processing delay), it could well be the only option on a smartwatch. Nevertheless, if the smartwatch is equipped with a display, the results of a voice search can be customised for:

- the smartwatch display, or
- the smartphone display, or
- voice response on smartwatch, or
- voice response on smartphone, or
- voice response on earphone or bluetooth-connected headset (if present), or
- any combination of the above.

Smartwatches are generally designed to present information in ways that are different from smartphones, given the smaller displays and possibly different navigation methods. Even the dialogues of voice response can differ between smartphones and smartwatches.

The additional means of presenting search results brought by wearable accessories could bring choice in user interaction; this, however, could translate into either convenience or hassle, depending on how well the smartphone works seemlessly with the wearable accessory in adapting the presentation of voice search results. This is illustrated in the following voice search scenarios:

1. User raises arm and utters to smartwatch, “Next appointment.” and the smartwatch displays the time, subject and location of the next appointment. User swipes smartwatch display to check the location on the map.

2. User raises arm and utters to smartwatch “Next appointment.” and the smartwatch replies, “Your next appointment, dental check up, will be in 1 hour 20 minutes. Would you like to show the map location on the phone?”. User replies, “Yes.” and the phone displays the map.

The above scenarios assume that the map application is available on both the smartphone and the smartwatch. Scenario 1 involves using only the smartwatch whereas scenario 2 requires both the smartwatch and the smartphone. While it would be straightforward to provide a user setting for choosing between the two scenarios (e.g. by enabling/disabling the verbal suggestion in scenario 2), we can regard them as two different multimodal presentations of voice search results; the choice between them could be made by the system depending on the current conditions at the time of usage and also the current user preferences. For instance, scenario 1 could be automatically chosen when:

- either the smartwatch, smartphone, or both are muted,
- the smartphone is playing music, or
- the user is walking.

On the other hand, scenario 2 could be the default option when the user is driving.

Today’s wearable smartphone accessories have built-in sensors such as pedometer, accelerometer, GPS receiver, etc. that help detect the environmental and behavioural conditions such as “the user is walking”, “the user is driving” under which voice search is carried out. The smartphone operating system can record such sensor data together with other relevant data such as user settings and usage of smartphone applications as records of voice search scenarios for adapting the modality of voice search.

Our work is mainly concerned with voice search on smartphones that are paired with smartwatches but it can also be readily generalised to other smartphone applications that work with smartwatches.

RELATED WORK

The work described in this paper is closely related to the previous work on the WWHT (What, Which, How and Then) model for multimodal fusion/fission based on contextualisation [10]. An adaptive multimodal user interface for mobile devices with visual, audio and tactile outputs is presented in [14]. It can replace the normal graphical user interface (such as Android or iOS) when the user is highly mobile or cannot afford full attention to operating the device. The multi-modal interface lightens the user’s cognitive load through simple text messages, blink patterns, voice messages, alarm patterns and vibration patterns.

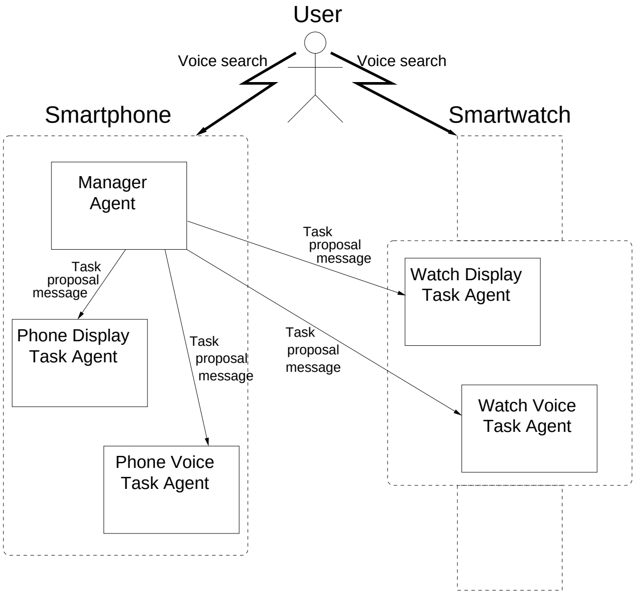


Figure 1. A multi-agent architecture for modality selection

Agents are employed in the JASPIS architecture to manage alternative inputs, dialogues as well as presentations in adaptive speech applications [13]. Agent negotiation is involved in the planning of multimodal presentations using MAGPIE [5]. Rather than dealing with modality selection, agents representing components (e.g. table cells) of a particular modality (e.g. table) are responsible for negotiating over the allocation of limited overall resources (e.g. maximum table size) in the planning of the presentation. The selection of modality is based on heuristics only. Furthermore, unlike the system discussed in this paper, MAGPIE utilises a hierarchical blackboard architecture in which agents are organised into hierarchical groups.

Machine learning has traditionally been important for modality handling (e.g. speech recognition, eye tracking, etc.) and modality fusion [3]. In [6], a multimodal multimedia computing system adapts its user interface based on environmental factors (e.g. location, noise level, etc.) with the help of machine learning. In this system, a single agent with machine learning capability is responsible for modality selection. The collection of touch interaction data on smartphones for mining behavioural patterns of smartphone users has previously been suggested [8]. In [11], machine learning is applied to user-smartphone interaction data to predict the disruptiveness of smartphone notifications. Six different machine learners (SVM, NB, kNN, RUSBoost, Genetic Programming and Association Rule learning) were employed to analyse the interaction dataset.

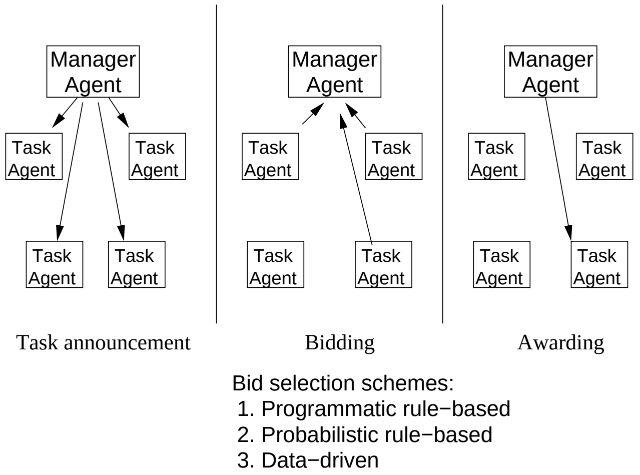


Figure 2. Agent negotiation process

A MULTI-AGENT ARCHITECTURAL DESIGN

This section gives the outline of a multi-agent architecture that supports dynamic and intelligent adaptation of the interface of a smartphone voice search application.

We follow the seminal work on agent negotiation based on the control net protocol [12] and employs two types of agent, namely, manager agent and task agent. Figure 1 depicts the multi-agent architecture for a paired smartphone-smartwatch configuration. A manager agent is responsible for announcing tasks by sending out task proposal messages to task agents. Assuming that only one voice search application is running at any one time, handling voice search requests one at a time, there need be only one manager agent. As mentioned in the Introduction section, voice search results can be customised for different modalities and, for each different modality, a task agent is responsible for negotiating with the manager agent for modality selection.

As soon as a voice search result becomes available, it is directed to the manager agent for selecting the most appropriate way to deliver the result to the user. The manager agent initiates a negotiation process (see Figure 2) which consists of three phases :

1. The manager agent broadcasts a task announcement message to all task agents.
2. For each task announcement message, each task agent checks the current status of its associated device (phone or watch) and preference settings of the user and decide whether to bid for the task. If yes, the agent needs to formulate a bid message and submit it to the manager agent.
3. The manager agent collects all the bids and evaluate them to decide which modality is best suited for presenting the current voice search result.

Note that in case the manager agent does not receive any bids, it simply wait for a random amount of time and then re-announce the task.

Bid Formulation and Selection

Task agents formulate bids based on the characteristics of the voice search results, user profiles/settings and the status of the output channels. A task agent may also choose not to bid under certain circumstances (e.g. when the watch's battery is low).

Various bidding schemes can be devised and it is the aim of our research to experiment with a number of such schemes and evaluate their performance in supporting multimodal mobile voice search. Currently, we are considering the following bidding schemes:

1. **Programmatic rule-based bid selection** Task agents simply provide the status of their output channels in their bids. The manager agent is programmed according to a set of predetermined logical (if..then..else) rules based on the available data to arrive at a decision. e.g.:

   If the surroundings are noisy and the user is walking, output to Watch Display.

   If the surroundings are *not* noisy and the user is walking, output to Watch Voice.

2. **Probabilistic rule-based bid selection** A problem of the above Programmatic rule-based scheme above is that the potential number of conditions and their combinations are potentially very large and tedious to code and there is no provision for uncertainty or ambiguity in the conditions. Following the expert system-based approach in [7], task agents can simply provide the status of their output channels in their bids and the manager agent utilises a set of probabilistic rules for evaluating bids. Each rule captures some knowledge about the user and/or environmental conditions and assign a probabilistic reward/penalty to an output channel (or a combination of output channels), e.g.:

   If the surroundings are noisy, output to Phone/Watch Voice (50%) .

   If the user is walking, output to Watch Display/Voice (50%).

   For each voice search output, the manager agent has to activate the relevant rules, compute the probability of any combined conditions and select the one with the maximum cumulative reward.

3. **Data-driven bid selection** The scheme requires a set of historical multimodal voice search scenario data accumulated by the smartphone and a data mining algorithm to derive rules for associating voice search scenarios with suitable output modality. The level of suitability in each scenario can either be rated subjectively by the user or derived from the user reaction to the voice search result in each case. For instance, if the user ignores an output on the smartwatch and chooses to open the result on the smartphone's display instead, the suitability of the former is considered as

low. For subjective rating, the voice search application can be extended with a rating menu or a rating dialogue.

When considering the computational complexity of the above schemes, the first one is trivial as it is not based on any rules or data. The second scheme has a complexity dependent on the set of rules and it affects the real-time performance of the application since the expert system will run in real-time. Finally, the complexity of the third scheme is based on the amount of data but the data mining algorithm is supposed to be run offline.

FURTHER WORK

The multi-agent architecture and bid selection schemes outlined in the preceding section are being implemented in Android and Android Wear. Student helpers will be invited to participate in the experiments. The experimental design is modelled after that in [11]. An Android voice search application which extends Google Voice Search is being developed for the experiments in this research. The custom application will support web search, contact search and appointment search.

Each helper will carry an Android smartphone and wear an Android smartwatch customised for the experiments. Usage data will be collected internally by the smartphones and each helper will also provide subjective ratings of the voice search application on a daily basis. The ratings will cover both qualitative and quantitative aspects of their experience; these include the suitability of modality selection as well as the relative performance of the different bid selection schemes. Collected data will be analysed for evaluating the relative merits of the different bidding schemes as well as the multi-agent approach on the whole.

CONCLUSION

Smart accessories for smartphones are coming of age and intelligent multimodal user interface is becoming mainstream. A multi-agent architecture lends itself to the problem of modality selection based on distributed sources of data. This research addresses the relative merits of different bid selection schemes and the results are expected to contribute to the design of more efficient and effective user interface for smartphone applications.

REFERENCES

1. Church, K., Smyth, B., Cotter, P., and Bradley, K. Mobile information access: A study of emerging search behavior on the mobile Internet. *ACM Transactions on the Web (TWEB) 1*, 1 (2007), 4.
2. comScore Inc. 2013 Mobile Future in Focus. Tech. rep., 2013.
3. Dumas, B., Lalanne, D., and Oviatt, S. Multimodal interfaces: A survey of principles, models and frameworks. In *Human Machine Interaction.* Springer, 2009, 3 – 26.
4. Feng, J., Johnston, M., and Bangalore, S. Speech and Multimodal Interaction in Mobile Search. In *Proceedings of the 20th International Conference Companion on World Wide Web*, WWW '11, ACM (New York, NY, USA, 2011), 293 – 294.
5. Han, Y., and Zukerman, I. A mechanism for multimodal presentation planning based on agent cooperation and negotiation. *Human–Computer Interaction 12*, 1 - 2 (1997), 187 – 226.
6. Hina, M. D., Ramdane-Cherif, A., and Tadj, C. A context-sensitive incremental learning paradigm of an ubiquitous multimodal multimedia computing system. In *Wireless And Mobile Computing, Networking And Communications, 2005.(WiMob'2005), IEEE International Conference on*, vol. 4, IEEE (2005), 104 – 111.
7. Honold, F., Schüssel, F., and Weber, M. Adaptive probabilistic fission for multimodal systems. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, ACM (2012), 222 – 231.
8. Huang, J., and Diriye, A. Web user interaction mining from touch-enabled mobile devices. *Proceedings of HCIR 2012* (2012).
9. Liu, C., Rau, P.-L. P., and Gao, F. Mobile information search for location-based information. *Computers in industry 61*, 4 (2010), 364 – 371.
10. Rousseau, C., Bellik, Y., Vernier, F., and Bazalgette, D. A framework for the intelligent multimodal presentation of information. *Signal Processing 86*, 12 (2006), 3696 – 3713.
11. Smith, J., Lavygina, A., Ma, J., Russo, A., and Dulay, N. Learning to recognise disruptive smartphone notifications. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, ACM (2014), 121 – 124.
12. Smith, R. G. The Contract Net Protocol: High-Level Communication and Control in a Distributed Problem Solver. *IEEE Transactions on Computer 29* (1980), 1104–1113.
13. Turunen, M., and Hakulinen, J. Jaspis-a framework for multilingual adaptive speech applications. In *Interspeech* (2000), 719 – 722.
14. Yamabe, T., Takahashi, K., and Nakajima, T. Towards mobility oriented interaction design: Experiments in pedestrian navigation on mobile devices. In *Proceedings of the 5th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) (2008), 46.

In-air Eyes-free Text Entry: A work in progress

Ali Alavi
ETH Zurich
Leonhardstrasse 21, 8092
Zurich, Switzerland
alavis@ethz.ch

Andreas Kunz
ETH Zurich
Leonhardstrasse 21, 8092
Zurich, Switzerland
kunz@iwf.mavt.ethz.ch

ABSTRACT

In this paper we provide a system for using in-air interaction techniques for eyes-free text entry. We show that complex tasks as text entry can be performed using a mix of simple pinching gestures, which provides the user with a feedback from his or her own sense of touch, compensating for lack of feedback in many touch and in-air interaction techniques.

Author Keywords
Gesture; Human Computer Interaction; Eyes free interaction; In-air gestures; Text entry

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces

INTRODUCTION AND BACKGROUND

Much research has been performed on different text entry methods. The area was especially motivated by the advent of new small devices, such as mobile phones, as well as no-keyboard touch-based technologies such as smart phones and tablets. These new interfaces required new efficient text entry methods, in order to cope with the lack of classic hardware keyboards. As a result, different hardware interfaces, algorithms and virtual keyboards have been developed to improve the situation. LetterWise [8], for example, offers a predictive text entry method, which uses probabilities of letter sequences to enhance the performance of text entry. WordWise [4], a commercial technology for text-entry on small keyboards, predicts the entered word by looking at the sequence of pressed keys, which avoids the user from pressing a key multiple times. Such methods require the user to get a continuous visual feedback through the screen, as dealing with ambiguous predictions is up to the user. Perkinput [5] on the other hand offers a nonvisual text entry on touch screens by detecting user's input fingers and assigning each finger to a Braille bit, hence enabling blind and visually impaired users to enter text on touch screens. It though requires a rather large interaction area, since both of user's hands should be on the screen. Escape-Keyboard [6] is another eyes-free text entry technologies, which recognizes user input based on the area of screen being touched as well as the flicking gestures performed by the user afterwards. Tapulator [10] also allows for nonvisual input, but it can only be used for number entry. Table 1 shows a summary of the mentioned technologies.

Technology	Eyes-Free	Mid-air	Input Type
WordWise	No	No	Word
LetterWise	No	No	Character
Perkinput	Yes	No	Character
EscapeKeyboard	Yes	No	Character
Tapulator	Yes	No	Number

Table 1. Comparing different text entry technologies

As using in-air gestures is starting to become established as a form of interaction, in-air text entry can be potentially a new form on interaction. This is mainly due to advent of advanced depth-sensing cameras such as Kinect [2] and Leap Motion [1]. These technologies enable affordable in-air hand and finger interaction, which allows the users to give inputs without touching any surface or button. This minimal interaction experience comes with its own problems. Apart from fatigue, which is one of the main problems experienced by majority of users, lack of haptic feedback is a major challenge, which makes the users dependent on some sort of visual feedback. This in turn makes eyes-free [9] interaction impossible.

While we do not expect such text entry methods to perform better in terms of speed or accuracy, the fact that such methods do not require any visual representation enables us to interact with small screens more efficiently, since there is no need anymore to display a soft keyboard on the screen.

Even more important, blind and visually impaired users will be able to type without the use of a keyboard or speech-to-text program, hence enabling them to use their voice for purposes other than text entry, e.g. speaking or chatting with other people, while entering a text on their devices.

SYSTEM DESIGN

Since our text entry method relies on detecting in-air hand pose and gesture, we need a sensor system to recognize users' hands. Over the past couple of years, we evidenced an increase in such sensors, particularly vision based ones. Hence, while developing this research, we assume the skeletal model of the hand as a given. In other words, how this skeletal model is produced is not our concern. Hence, we can use this system with any sensor, as long as it can provide a skeletal model of the hand. Currently, for practical purposes and due to its

IUI 2015, March 29–April 1, 2015, Atlanta, GA, USA.
Copyright © 2015 ACM 978-1-4503-3306-1/15/03...$15.00.
DOI string from ACM form confirmation

wide availability, we use a Leap Motion sensor which provides such a model (Figure 1).

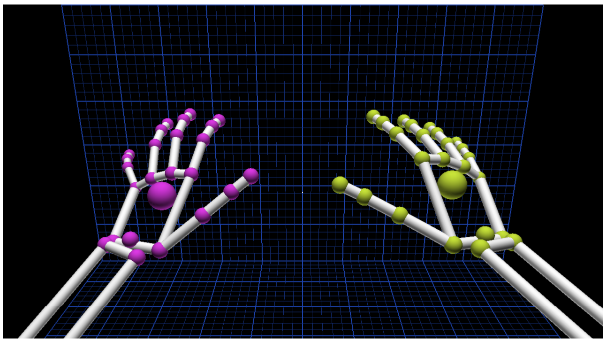


Figure 1. Skeletal model of the hand provided by Leap Motion [3]

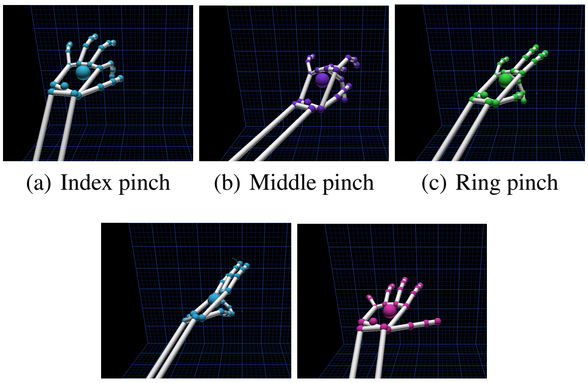
System setup

In our setup, the Leap Motion sensor is put on a desk, and the user can interact with it in the area above it. The orientation and exact location of the sensor is not important, as long as it is able to see both hands.

Gesture selection and recognition

Our gesture selection approach has two steps. Firstly, we select a set of gestures which are easy for users to perform eyes-free. We call such gestures *basic gestures*. Since the set of such gestures is rather small, we cannot correlate them directly to English alphabet. Hence, a combination of these gestures are defined to get a larger set of gestures (Figure 4). We call these gestures *complex gestures*.

The basic gesture recognizer recognizes five different gestures per hand: one open hand gesture, and four different pinching gestures: pinching index finger, middle finger, ring finger and little finger. The corresponding skeletal model of these gestures can be seen in figure 2. We omit showing right hand gestures for brevity.



(a) Index pinch (b) Middle pinch (c) Ring pinch

(d) Little pinch (e) Open hand

Figure 2. Left hand gestures as seen by the sensor

These gestures are recognized by looking at the distance between thumb and other fingers, and if the distance between tip of thumb and another finger tip is below a certain threshold, a pinching gesture is recognized (in case of multiple fingers fitting into this criterion, we take the finger that has the minimum distance with thumb). The output from this gesture recognizer is a stream of symbols, each corresponding to a basic gesture, as shown in table 2.

Pose	Symbol
Right hand Open	o
Right index finger	i
Right middle finger	m
Right ring finger	r
Right little finger	l
Left hand Open	O
Left index finger	I
Left middle finger	M
Left ring finger	R
Left little finger	L

Table 2. List of gestures and their corresponding symbols

Left hand and right hand gestures are always emitted in pairs, as in the following example:

```
oOiIoOmM...
```

Such streaming output is used as the input of a *complex gesture* recognizer, which tests the stream against a set of criteria, each indicating a complex gesture. This check is performed using one Finite State Machine per complex gesture. For example, we can define a complex gesture for entering letter *A* with the following sequence:

```
oOiOiMoO
```

That is, in order to enter an *"A"* letter, the user should start with open hands, then pinch her left hand and then, while keeping that pinch, pinches her right middle finger, and then should release both pinches. Since we are streaming the gestures on a fixed time basis (every 100 milliseconds), our stream will generate more than one symbol per gesture, i.e. if a user keeps his hand open for one second, the system will stream out 10 *oO* symbols. This will later help us to determine gestures based on their temporal characteristics. But in order to determine corresponding gesture for letter *A*, as mentioned above, we need to disregard repetitive gestures. This can be better described by a Finite State Machine (FSM) (see Figure 3).

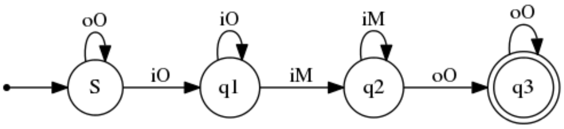


Figure 3. A finite state machine for describing a complex gesture

Looking at complex gestures in form of a finite state machine of basic gestures has four major benefits:

1. Since we have a small set of basic gestures, it would be easy for the users to learn how to interact with the system.
2. Developing a gesture recognizer for a small set of basic gestures takes little effort.

3. Defining complex gestures as a formal automata enables us to quickly define a large number of new gestures.
4. We can systematically find conflicts among different gestures, by calculating their intersection automaton.

In practice, we define such formal automata using regular expressions [11], i.e. we define the above-mentioned state machine using the following regular expression:

```
(Oo)+ (iO)+ (iM)+ (oO)+
```

Since we can define a large number of conflict-free complex gestures, we define a one-to-one relation between gestures and English letters, i.e. each complex gestures correlates to a letter of English alphabet.

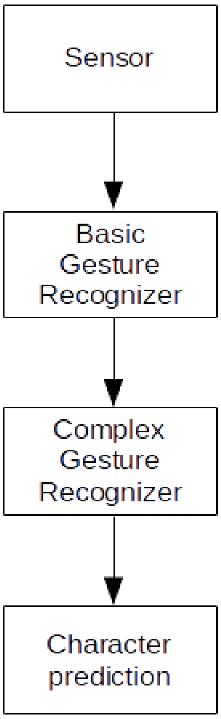


Figure 4. An overview of the system

Dealing with uncertainty

Despite significant improvement in sensory technology, the output of all of these sensors has a level of uncertainty (due to noise, occlusion, inherent system errors, ...). Since we do not want to deal with sensory technology in this research, we consider uncertainty as an input to our system. This uncertainty affects our basic gesture recognizer, as any error in skeletal model of the hand can directly affect our interpretation of gestures. The degree of such errors varies with the type of gestures (presumably due to occlusion, asymmetric relative position of hand and sensors, and so on). Figure 5 shows an example of these measurement errors, in an experiment with only one user, with 58 repetition for each gesture.

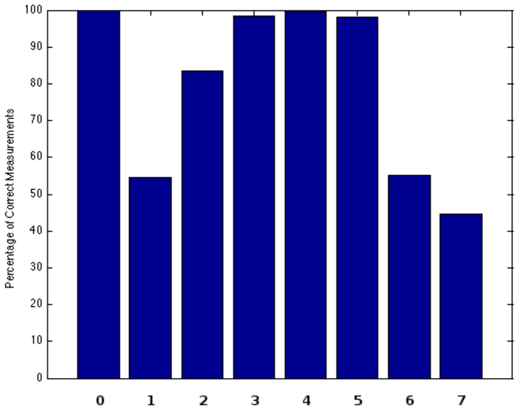


Figure 5. Correct measurement of pinching gestures. 0: Left Index, 1: Left Middle, 2: Left Ring, 3: Left little, 4: Right Index, 5: Right Middle, 6: Right Ring, 7: Right little

We try to address this issue by converting the problem to a bigram model. We assign each state to a basic gesture. Since each letter is represented by a sequence of symbols (complex gestures), the state transition probability of the model can be inferred from two known parameters: transition probability of basic gestures in the set of complex gestures, and frequency of English letters. We can then marginalize the English letter's frequency probability out to get bigram probability of basic gestures, which is the transition matrix of our bigram Hidden Markov Model (Figure 7).

$$P(s_i|s_j) = \sum_l P((s_i|s_j), l) = \sum_l P((s_i|s_j)|l)P(l)$$

Finally, the observation probability matrix of the model, $P(s_i|o_i)$, is measured by an experiment. In this experiment, we measure the observation rate of different gestures when a specific gesture is expected. In figure 6, we see the observation probability of different gestures when the user performs an *oR* gesture.

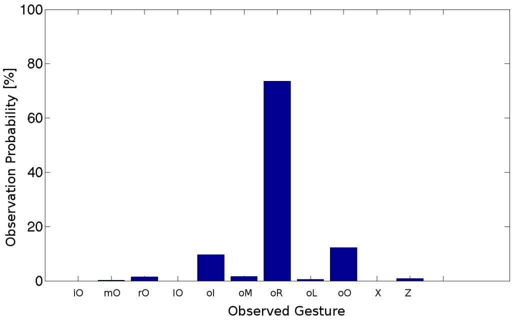


Figure 6. Observation probability of different gestures when basic gesture oR is performed. Unobserved gestures not shown.

CONCLUSION AND FUTURE WORK

In this paper, we described a new technique for in-air eyes-free text entry. We managed to define a set of gestures which

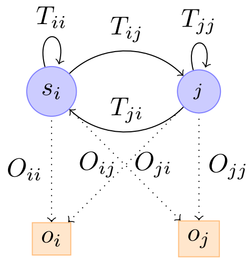


Figure 7. A partial representation of the Hidden Markov Model. s_i and s_j represent states (hence basic gestures), and o_i and o_j represent observations. $T_ij = P(s_i|s_j)$ and $O_ij = P(s_i|o_j)$

could be performed eyes-free, and systematically defined a larger set of gestures based on those basic gestures. Our future work will focus on testing the effectiveness of the system, particularly for accessibility use cases. Moreover, our basic gesture recognizer is a very naive one, and we are working on extending it by using a programming by demonstration [7] approach, where users can define their own set of basic gestures in a more efficient way. Finally, comparing this method of text entry with other methods will conclude this research.

REFERENCES

1. Leap motion, https://www.leapmotion.com/, [accessed 10 january 2015].
2. Micsosoft kinect, http://www.microsoft.com/en-us/kinectforwindows/, [accessed 10 january 2015].
3. Skeletal model of hand by leap motion, https://developer.leapmotion.com/documentation/javascript/devguide/intro_skeleton_api.html, accessed 10 january 2015.
4. Wordwise, http://www.eatoni.com/wiki/index.php/wordwise, accessed 10 january 2015.
5. Azenkot, S., Wobbrock, J. O., Prasain, S., and Ladner, R. E. Input finger detection for nonvisual touch screen text entry in perkinput. In *Proceedings of Graphics Interface 2012*, GI '12, Canadian Information Processing Society (Toronto, Ont., Canada, Canada, 2012), 121–129.
6. Banovic, N., Yatani, K., and Truong, K. N. Escape-keyboard: A sight-free one-handed text entry method for mobile touch-screen devices. *International Journal of Mobile Human Computer Interaction (IJMHCI) 5*, 3 (2013), 42–61.
7. Kosbie, D. S., and Myers, B. A. Extending programming by demonstration with hierarchical event histories. In *Human-Computer Interaction*, B. Blumenthal, J. Gornostaev, and C. Unger, Eds., vol. 876 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 1994, 128–139.
8. MacKenzie, I. S., Kober, H., Smith, D., Jones, T., and Skepner, E. Letterwise: Prefix-based disambiguation for mobile text input. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, ACM (New York, NY, USA, 2001), 111–120.
9. Oakley, I., and Park, J.-S. Designing eyes-free interaction. In *Proceedings of the 2Nd International Conference on Haptic and Audio Interaction Design*, HAID'07, Springer-Verlag (Berlin, Heidelberg, 2007), 121–132.
10. Ruamviboonsuk, V., Azenkot, S., and Ladner, R. E. Tapulator: A non-visual calculator using natural prefix-free codes. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '12, ACM (New York, NY, USA, 2012), 221–222.
11. Thompson, K. Programming techniques: Regular expression search algorithm. *Commun. ACM 11*, 6 (June 1968), 419–422.