Fabio Paternò   Boris de Ruyter
Panos Markopoulos   Carmen Santoro
Evert van Loenen   Kris Luyten (Eds.)

LNCS 7683

Ambient Intelligence

Third International Joint Conference, AmI 2012
Pisa, Italy, November 2012
Proceedings

Springer

Lecture Notes in Computer Science 7683

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

Fabio Paternò Boris de Ruyter
Panos Markopoulos Carmen Santoro
Evert van Loenen Kris Luyten (Eds.)

Ambient Intelligence

Third International Joint Conference, AmI 2012
Pisa, Italy, November 13-15, 2012
Proceedings

Springer

Volume Editors

Fabio Paternò
CNR-ISTI, Pisa, Italy
E-mail: fabio.paterno@isti.cnr.it

Boris de Ruyter
Philips Research, Eindhoven, The Netherlands
E-mail: boris.de.ruyter@philips.com

Panos Markopoulos
Eindhoven University of Technology, The Netherlands
E-mail: p.markopoulos@tue.nl

Carmen Santoro
CNR-ISTI, Pisa, Italy
E-mail: carmen.santoro@isti.cnr.it

Evert van Loenen
Philips Research, Eindhoven, The Netherlands
E-mail: evert.van.loenen@philips.com

Kris Luyten
Hasselt University, Diepenbeek, Belgium
E-mail: kris.luyten@uhasselt.be

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-34897-6 e-ISBN 978-3-642-34898-3
DOI 10.1007/978-3-642-34898-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012951487

CR Subject Classification (1998): I.2, H.4, H.3, C.2.4, H.5, I.2.11, K.4

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

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

Printed on acid-free paper

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

Preface

This volume contains the papers and posters selected for presentation at the International Joint Conference on Ambient Intelligence (AmI 2012) held in Pisa in November 2012.

The vision of ambient intelligence is to provide environments enhanced by intelligent interfaces supported by computing and networking technology embedded in everyday objects, and which enable users to interact with their surroundings in a seamless manner.

More specifically, such environments should result in systems that are aware of the characteristics of users, recognize their needs, learn from their behavior, and are able to intelligently and even proactively act in order to support humans in achieving their goals. Ambient intelligence should also be unobtrusive – interaction should be natural and engaging for the users.

From a scientific point of view, ambient intelligence (AmI) comprises a multidisciplinary approach covering fields such as computer science, human computer interaction, electrical engineering, industrial design, behavioral sciences, aimed at enriching physical environments with a network of distributed devices, such as sensors, actuators, and computational resources, in order to support users in their everyday activities.

From a technological perspective, AmI represents the convergence of recent achievements in ubiquitous and communication technologies, pervasive computing, intelligent user interfaces and artificial intelligence, just to name a few.

This conference started as the European Symposium on Ambient Intelligence in 2003, and has grown to an annual international event that brings together researchers and serves as a forum to discuss the latest trends and developments in this field.

These AmI 12 proceedings include the latest research into technologies and applications that enable and validate the deployment of the AmI vision.

This year the program contained 18 full papers carefully chosen from a total of 47 submissions (38% acceptance rate). There were also five short papers accepted out of 14 (acceptance rate 36%). All papers were reviewed in a double-blind review process. For some papers this included a conditional acceptance step which required further revisions finally checked by reviewers and Chairs. In addition, the program included five landscape papers (papers that brainstorm on the future evolution of AmI), ten posters, and two demos.

The competition for paper acceptance was strong and final selection was difficult. The published material originates from 27 countries, including Africa, Australia, North and Central America, Japan, Saudi Arabia, Singapore, and Europe.

Each paper had at least two independent reviews from reviewers who were matched by expertise area to the topic of each paper. The Chairs handled borderline cases, and requested additional reviews when needed.

In addition to the main conference, seven workshops were held prior to the main AmI 2012 event, and stimulated interesting discussions on specific relevant topics.

A special thanks goes to the dedicated work of the 54 Program Committee members involved in the review panel who came from Europe and North America, thus reflecting the international spirit of AmI participation. Their names are listed in the conference proceedings and on the website.

We would also like to express our gratitude to ACM SIGCHI, Interaction-design.org, SIGCHI Italy, IFIP WG 2.7/13.4 for their help in creating interest in the conference.

Finally, we would like to thank the conference Organizing Committee for their dedicated support, as well as the paper presenters and conference participants who contributed to the vibrant discussions, presentations, and workshops held at AmI 2012.

Fabio Paternò
Boris de Ruyter
Panos Markopoulos
Carmen Santoro
Evert van Loenen
Kris Luyten

Organization

The 6th European Conference on Ambient Intelligence, AmI 2012, was held in Pisa, Italy.

General Chairs

Fabio Paternò	CNR-ISTI, HIIS Laboratory, Italy
Boris de Ruyter	Philips Research Europe, The Netherlands

Full Papers

Panos Markopoulos	Eindhoven University of Technology, The Netherlands
Carmen Santoro	CNR-ISTI, HIIS Laboratory, Italy

Short Papers

Evert van Loenen	Philips Research Europe, The Netherlands
Kris Luyten	Hasselt University, Expertise Centre for Digital Media, Belgium

Landscapes

Adrian David Cheok	Mixed Reality Lab, National University of Singapore

Posters

Davy Preuveneers	University of Leuven, Belgium
Davide Spano	CNR-ISTI, HIIS Laboratory, Italy

Demos

Giuseppe Ghiani	CNR-ISTI, HIIS Laboratory, Italy
Giulio Mori	CNR-ISTI, HIIS Laboratory, Italy

Workshops

Gerrit Meixner	DFKI, Germany
Ben Schouten	Eindhoven University of Technology, The Netherlands

Doctoral Consortium

Manfred Tscheligi	University of Salzburg, Austria
Volker Wulf	University of Siegen and Fraunhofer-FIT, Germany

Local Organization

Giulio Galesi	CNR-ISTI, HIIS Laboratory, Italy

Program Committee

Emile Aarts	Philips Research, The Netherlands
Julio Abascal	University of the Basque Country, Spain
Ville Antila	VTT, Finland
Carmelo Ardito	University of Bari, Italy
Juan Carlos Augusto	University of Ulster, UK
Yolande Berbers	University of Leuven, Belgium
Regina Bernhaupt	Ruwido, Austria
Gerald Bieber	Fraunhofer IGD-R, Germany
Oliver Brdiczka	Palo Alto Research Center, CA, USA
Gaelle Calvary	LIG-IIHM, University of Grenoble, France
Javier Caminero	Telefónica I D, Spain
Luis Carriço	Universidade de Lisboa, Portugal
Stefano Chessa	University of Pisa, Italy
Karin Coninx	University of Hasselt, Belgium
Marco de Sa	Yahoo! Research, CA, USA
Monica Divitini	NTNU, Norway
Jose Carlos Dos Santos Danado	CNR-ISTI, HIIS Laboratory, Italy
Morten Fjeld	Chalmers University of Technology, Sweden
Jacqeuline Floch	SINTEF, Norway
Mathias Funk	TU Eindhoven, The Netherlands
Francesco Furfari	CNR-ISTI, Italy
Luciano Gamberini	University of Padova, Italy
Maurits Clemens Kaptein	TU Eindhoven, The Netherlands
Fahim Kawsar	Alcatel-Lucent, Belgium
Javed Vassilis Khan	Breda University, The Netherlands
Gerd Kortuem	The Open University, UK
Vassilis Kostakos	University of Oulu, Finland
Kyriakos Kritikos	FORTH-ICS, Greece
Ben Krose	University of Amsterdam, The Netherlands
Brian Lim	Carnegie Mellon University, PA, USA
Alexander Meschtscherjakov	University of Salzburg, Austria
Vittorio Miori	CNR-ISTI, Italy
Michael Nebeling	ETH Zurich, Switzerland
Laurence Nigay	LIG-IIHM, University of Grenoble, France

Zeljko Obrenovic	Eindhoven University of Technology, The Netherlands
Philippe Palanque	IRIT, France
Volkmar Pipek	University of Siegen, Germany
Davy Preuveneers	University of Leuven, Belgium
Aaron Quigley	The University of St. Andrews, UK
Joerg Rett	SAP Research Center Darmstadt, Germany
Enrico Rukzio	Ulm University, Germany
Albert Ali Salah	Bogazici University, Turkey
Thomas Schlegel	TU Dresden, Germany
Dirk Schnelle-Walka	TU Darmstadt, Germany
Johannes Schöning	DFKI, Germany
Ahmed Seffah	UTT, France
Kostas Stathis	Royal Holloway University of London, UK
Gunnar Stevens	University of Siegen, Germany
Manfred Tscheligi	University of Salzburg, Austria
Kaisa Väänänen-Vainio-Mattila	Tampere University of Technology, Finland
Kristof Van Laerhoven	TU Darmstadt, Germany
Jo Vermeulen	University of Hasselt, Belgium
Reiner Wichert	IGD Fraunhofer, Germany
Massimo Zancanaro	Fondazione Bruno Kessler, Italy

Table of Contents

Long Papers

Short Papers

Landscape Papers

Poster Papers

Demo Papers

Workshops

Context-Based Fall Detection Using Inertial and Location Sensors

Hristijan Gjoreski, Mitja Luštrek, and Matjaž Gams

Department of Intelligent Systems, Jožef Stefan Institute
Jamova cesta 39, 1000 Ljubljana, Slovenia
{hristijan.gjoreski,mitja.lustrek,matjaz.gams}@ijs.si

Abstract. Falls are some of the most common sources of injury among the elderly. A fall is particularly critical when the elderly person is injured and cannot call for help. This problem is addressed by many fall-detection systems, but they often focus on isolated falls under restricted conditions, neglecting complex, real-life situations. In this paper a combination of body-worn inertial and location sensors for fall detection is studied. A novel context-based method that exploits the information from both types of sensors is designed. The evaluation is performed on a real-life scenario, including fast falls, slow falls and fall-like situations that are difficult to distinguish from falls. All the possible combinations of six inertial and four location sensors are tested. The results show that: (i) context-based reasoning significantly improves the performance; (ii) a combination of two types of sensors in a single physical sensor enclosure seems to be the best practical solution.

Keywords: Context-based reasoning, Fall detection, Inertial sensors, Location sensors, Activity recognition.

1 Introduction

Falls are some of the most critical health-related problems for the elderly [3]. Approximately 28–35% of people over the age of 65 fall each year, and this proportion increases to 32–42% in those aged more than 70 years [20]. About 20% of all the fall accidents that involve an elderly person require medical attention [6]. Furthermore, falls and the fear of falling are important reasons for nursing-home admission [18]. Falls are particularly critical when the elderly person is injured and cannot call for help. These reasons, combined with the increasing accessibility and miniaturization of sensors and microprocessors, is driving the development of fall-detection systems.

Even though fall detection (FD) has received significant attention in recent years, it still represents a challenging task for two reasons. First, there are several everyday fall-like activities that are hard to distinguish from fast falls. Most of the current approaches define a fall as having greater accelerations than normal daily activities. However, focusing only on a fast acceleration can result in many false alarms during fall-like activities with fast acceleration, such as sitting down quickly or lying down on a bed quickly. The second reason why FD is challenging is that not all falls are

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 1–16, 2012.

characterized by a fast acceleration. Rubenstein et al. [16] showed that 22% of the falls experienced by the elderly are slow and are caused by dizziness and vertigo (13%), and drop attacks (9%). Therefore, the detection of slow falls should be an intrinsic part when creating a successful fall-detection system.

To overcome the problems of the existing fall-detection methods discussed above, we propose a new approach to FD by combining body-worn inertial and location sensors, named CoFDILS (Context-based Fall Detection using Inertial and Location Sensors). Our approach uses the context information from the both types of sensors to determine whether a fall has occurred. It exploits body accelerations, location and atomic activities to detect a fall. The evaluation was performed on a special real-life scenario that includes fast falls, slow falls and non-fall situations that are difficult to distinguish from falls. In addition, we tested 1023 possible body-placement combinations of six inertial and four location sensors in order to find the best-performing sensor placements for FD and therefore to achieve the lowest sensor burden on the user. The results showed that by combining the two types of sensors it is possible to detect complex fall situations and that the context-based reasoning significantly improves the performance.

The paper is organized as follows. Firstly, an overview of the related studies for FD is presented in Section 2. In the next two sections, the sensor equipment (Section 3) and the architecture of our system (Section 4) are described. Next, the preprocessing of the raw data is presented in Section 5. In the next two sections we describe the context components (Section 6) and the methodology (Section 7). After that, the experimental setup, the results and discussions are presented in Section 8 and 9. Finally, we conclude this work and give directions for future work in Section 10.

2 Related Work

FD approaches can be divided into those using non-wearable and wearable (i.e., body-worn) sensors. The most common non-wearable approach is camera-based [10, 15]. Although this approach is physically less intrusive to the user compared to the body-worn sensors, it suffers from issues such as low image resolution, target occlusion and time-consuming processing. However, often the biggest issue is user privacy: the user has to accept the fact that a camera will record him/her.

Most of the studies for FD are based just on inertial sensors. Usually, they are focused only on fast falls [13, 23], which are not difficult to detect using the acceleration signal. The non-fall events used to test for false positives are usually normal, everyday activities [8, 15], not events chosen specifically because they are easily mistaken for falls. In contrast, we used complex falls and safe events that appear like falls. An example where FD was evaluated on events difficult to recognize as falls or non-falls is the work by Li *et al.* [11]. By applying thresholds to two inertial sensors, they detected a fall with an accuracy of 90.1%. The recall value of their method on a fall event ending with sitting is 50% and for a non-fall event, quickly lying on a bed, is 40%. By combining one inertial and one location sensor, we were able to achieve 99% and 100%, on similar events, respectively.

A combination of inertial and location sensors was described in Zinnen *et al.* [23]. However, their goal was activity recognition for car-quality control and they did not deal with FD. Their approach was based on high-level primitives that were derived from a reconstructed human-body model by using inertial sensor data. The location data was mainly used to estimate the person's location near the car.

We are not aware of any prior publication that studies a combination of body-worn inertial and location sensors for FD, except ours [14]. There, we focused on location-based FD, and we considered only a single accelerometer to detect the impact of the fall and the orientation of the user. The main advantages of the study presented here compared to our previous work are: (i) a machine-learning model that recognizes the activity of the user; (ii) a thorough analysis of the system's complexity and invasiveness to the user by analyzing the performance of all the possible body-placement combinations of 10 sensors; and (iii) an explicit presentation of the context-based reasoning algorithm, the core of our system.

A context-based approach to FD is presented in the study by Li *et al.* [12]. However, they used a different fall-detection method and different types of sensors to extract the context information, compared to our approach. In particular, they used 5 body-worn accelerometers and 2 environmental sensors that monitor the vibration of the furniture. They combined the user's posture information, extracted from the accelerometers, and the context information, extracted from the environmental sensors, in order to detect the fall situations. Although they also analyzed slow falls and fall-like situations, their evaluation was performed on only 3 test subjects; while we tested our method on 11 subjects. The advantage of our location system, compared to the environmental sensors, is that it provides richer information about the user's situation, e.g., the user's location, the sensor's height, etc. The environmental sensors used in their research can only inform about the presence/absence of the user at a specific location where the sensor is installed. We tested all the combinations of 10 sensors and found a satisfactory performance with single sensor enclosure, while they analyzed only the fixed 5 accelerometer placements on the body.

To summarize, the improvements of our FD approach upon most related work are the following:

- combining two types of body-worn sensors: inertial and location,
- context reasoning about the user's situation,
- analysis of the FD performance for all combinations of 10 sensor body placements,
- machine learning activity-recognition model as a part of the FD,
- evaluation on a complex test scenario, including events such as slow falls and fall-like events that may be difficult to distinguish from falls.

3 Sensor Equipment

The CoFDILS sensor equipment consists of inertial and location sensors (Fig. 1). The two types of sensors were chosen because inertial sensors are relatively cheap and portable, and the location sensors provide rich information about the user, without significantly compromising the user's privacy (like with the cameras).

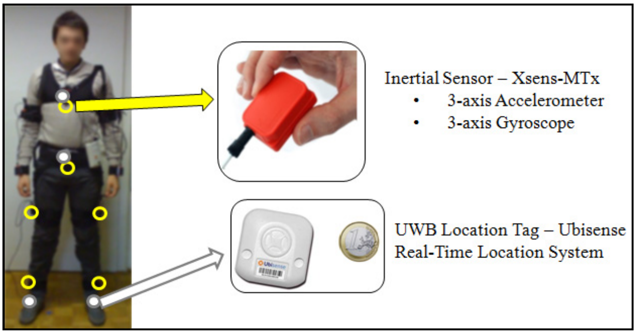


Fig. 1. Sensor equipment. The empty yellow circles represent the inertial sensors and the filled white circles represent the location tags.

Six inertial sensors were placed on the chest, waist, left thigh, right thigh, left ankle, and right ankle (non-filled circles in Fig. 1). Since only activities that are associated with the user's legs and torso were studied, the arm- and wrist-sensor placements were not considered. The inertial sensor equipment consisted of body-worn Xsens-MTx sensors [22], but the methods developed for this research are general and can be applied to any type of inertial sensor.

Four location tags were placed on the chest, waist, left ankle and right ankle (filled circles in Fig. 1). They emit UWB radio signals, which are detected by sensors fixed in the corners of a room. The tags are detected by the location system and their coordinates are computed. The location system used in CoFDILS is Ubisense [19]; it is a real-time location system used to track subjects indoors. Note that for simplicity the term sensor is also used for the body-worn location tag.

The data-sampling frequency of the sensors was set to 10 Hz because of Ubisense's hardware limitations. Although the inertial sensors do not have the same limitation, the data is sampled at the same frequency to simplify the synchronization.

4 System Architecture

The architecture of CoFDILS is shown in Fig. 2. First, the data from both types of sensors is stored and preprocessed. Next, the flow of data splits into two. On the top, firstly, a feature extraction is performed and the constructed feature vector is fed to the activity-recognition (AR) classification model, which recognizes the activity of the user. At the bottom, context-based reasoning about the user's situation is performed. The context reasoning analyzes the activity of the user and additional context information from the preprocessed data. The motivation is that the context information depends on the type of sensors. Inertial sensors provide body-movement information and the detection of a rapid-acceleration fall pattern, i.e., a threshold-based approach (TBA). Location sensors provide the location of the user in the room. The system evaluates the information from various sources in light of its contexts and concludes whether a fall alarm should be issued. Each module in Fig. 2 is presented in more detail in the sections that follow.

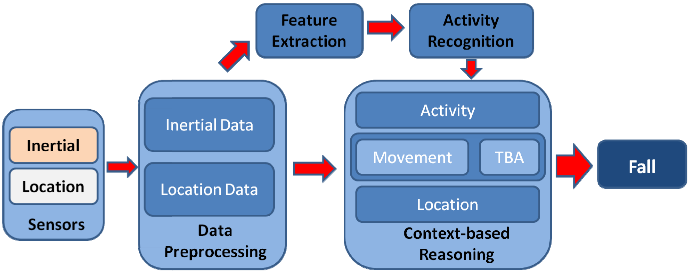


Fig. 2. CoFDILS architecture. TBA – Threshold-based approach.

5 Data Preprocessing

5.1 Inertial Data

An inertial sensor provides the raw data that consists of 3-axis accelerometer data and 3-axis gyroscope data. It measures the accelerations and the angular velocities represented in three directions.

The raw data was filtered with low-pass and high-pass filters. The low-pass filter removes the movement of the sensors, which leaves only the gravity component. This information is useful, especially for an assessment of the sensor-inclination angles. In contrast, the high-pass filter removes the gravity and only the sensor movements are left. These filters were applied separately: if the gravity component is needed, the low-pass filtered data is used; otherwise, the high-pass filtered data is used.

Finally, an overlapping sliding-window technique was applied for the AR. This means that a one-second window moves across the stream of data, advancing by half its length for each step.

5.2 Location Data

The Ubisense's output consists of the 3D coordinates of the sensors that are attached to the user's body. In a typical open-environment, the localization accuracy is about 15 cm, but in practice it may occasionally drop to 200 cm or more. Therefore, filtering was performed in order to tackle the problems with the Ubisense system [9].

First, a median filter computed each coordinate as the median of the measured values in a time window. This type of filtering removes large, short-term deviations of a measured coordinate from the true one. Second, the coordinates were corrected with a filter enforcing anatomic constraints based on the user's height and the body proportions. After that, a Kalman filter was used to smooth the data and correct some of the errors. Finally, the same overlapping sliding-window technique was applied.

6 Context Components

The most important novelty in our fall-detection method (CoFDILS) is based on the use of the context information. In general, a context is defined as any information that can be used to characterize the circumstances in which an event occurs [2]. In CoF-DILS, the context information consists of three components: (i) the user's body accelerations, (ii) the user's activities and (iii) the location of the user.

6.1 Body Accelerations

Threshold-Based Approach

The threshold-based approach (TBA) is used as one of the components in CoFDILS, as well as a baseline for comparison. The rationale for this method is that the acceleration pattern during a typical fall (i.e., fast, uncontrolled) is a decrease in the acceleration (free fall) followed by a rapid increase (impact with the ground). For our implementation of the TBA, the difference between the maximum and minimum accelerations within a one-second window was calculated. If the difference exceeded the threshold and the maximum appeared after the minimum, a fall was declared. The threshold was chosen empirically based on preliminary data [4].

Body Movement

During motion the accelerometers produce a changing acceleration signal and the fiercer the motion, the greater the change in the signal. Using these changes a feature is extracted: Acceleration Vector Changes (AVC) [4]. This feature sums up the differences between consecutive high-passed values of the lengths of the acceleration vectors, and divides the sum by the time interval (one second):

$$AVC = \frac{\sum_{i=1}^{n} | length_i - length_{i-1} |}{T_n - T_0} \tag{1}$$

T_0 is the time stamp for the first data sample in the window, and T_n is the time stamp of the last data sample. By applying a threshold to the AVC value, the movement of the appropriate sensor is detected.

6.2 Activity Recognition

Seven basic (atomic) activities that can also be interpreted as body postures were studied: *standing*, *sitting*, *lying*, *sitting on the ground*, *on all fours*, *going down* and *standing up*. We decided only for these activities because they are the most common, atomic, everyday-life activities and are also the most relevant for the detection of falls and distinguishing them from non-falls. Note that *sitting* and *sitting on the ground* are two different activities/postures and are not related to the location (e.g. chair, bed, ground) but only to the user's body orientation.

To recognize the activities of the user, machine learning (ML) was used. The idea of the ML approach was to learn a classification model that will be able to classify the target activities of the person wearing the sensors. The first step in the ML-based AR is the feature extraction procedure. The activities needed to be represented by general features, so that the ML method will also be general and work well on situations different from those in our scenario. Therefore, using the sliding-window technique (described in Subsection 5.1), the data from both types of sensors was first transformed into a number of features. Then, the feature vector was fed into the classification model, which recognized the activity of the user. The ML analysis was performed using the API of the software toolkit WEKA [21]. Among several methods tested, *Random Forest* yielded the best results in preliminary tests [4, 7]. Random Forest (RF) is an ensemble of decision trees in which the final decision is formed by a majority vote of the tree models [1].

Inertial Features

This subsection briefly describes the features extracted from the inertial sensors' data and used in the AR [5]. The total number of extracted features per sensor is 25, i.e., 8 for the gyroscope data and 17 for the accelerometer data, divided into four groups:

- Statistical features (total 20). The Mean Value and the Standard Deviation are extracted for both the acceleration and gyroscope data; additionally, the Root Mean Square (RMS) is calculated only for the accelerometer data. A feature-selection analysis showed that the RMS is a redundant feature for the gyroscope data.
- Movement feature (AVC feature, explained in the Body Movement subsection).
- Sensor inclination angles (total 3). They represent the orientation of the sensor, calculated as the angles between the actual acceleration and each of the axes.
- Difference between the maximum and minimum value of the high-passed acceleration vector in the current data window.

Location Features

The number of features extracted for the location data does not increase linearly with the number of sensors. The reason for this is that there are features that are extracted for pairs of sensors. Most of the features omit the x and y coordinates of the sensors because they refer to the specific location of the user in the room. Our goal was to build a general AR classification model that would not depend on the room's characteristics. The following features were extracted:

- the z (height) coordinate of the sensor,
- the Euclidian distances between each pair of sensors,
- the z-distances between each pair of sensors (difference in heights),
- the Euclidian distances between each pair of sensors in the xy plane,
- two velocity-based features: the first one is the absolute velocity of the sensor, and the second one is computed as the velocity of the sensor in the z direction.

6.3 Location

The location system outputs the 3D coordinates of the sensors that are attached to the user's body. In this way it captures properties such as the location of the user in the apartment, e.g., whether the user is on the floor or on the bed, the height of the sensors, etc.

7 Context-Based Reasoning

In this section the context-based reasoning in CoFDILS is presented (Fig. 3). The general idea is that each of the previously described components uses the information from the other two as context, and reasons about the user's situation. Therefore, there are three possible cases: (i) the *body acceleration* component uses the *activity* and *location* as context; (ii) the *activity* uses the *body acceleration* and *location* as context; and (iii) the *location* uses the *activity* and *body acceleration* as context. Additionally, the time component is included, giving another dimension to the reasoning. The time is important for synchronizing the values provided by each component and when analyzing the activities and locations in the time intervals.

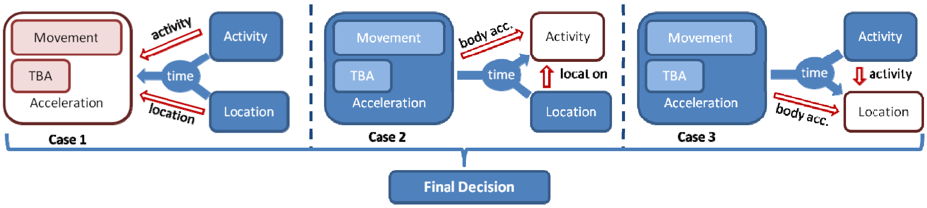


Fig. 3. Context-based reasoning schema

To explain the basic principle of the context-based reasoning, let us consider the following example in which a user is lying down quickly on a bed, i.e., a non-fall situation. In this case, the acceleration component, i.e., the TBA, recognizes a large acceleration (Case 1 in Fig. 3). If this component reasons by itself, a wrong decision would be formed: a fall would be detected. If the activity of the user is additionally evaluated, the decision would still be wrong (a large acceleration and lying activity = typical fast fall). However, when the location of the user is evaluated (which is the bed), the final decision is corrected into non-fall (quickly lying on the bed).

Besides the example described in the previous paragraph, there are 168 combinations of context dependencies, based on the values of each context component. For a final conclusion, all the potential rules are tested and confirmed. Some rules are primary in the sense that their conclusion is imperative, e.g., normal walking after a large acceleration cancels all the potential fall alarms. Other rules are evaluated using levels of importance, i.e., urgent alarms, normal alarms and warnings, and the most important rule prevails. However, in this study only falls were tested and therefore all alarms had the same level of importance, i.e., urgent. Additionally, in our tests it turned out that only a couple of context-based relations capture most of the context information in the FD domain. In particular, to describe a fall we used the activity as a basic piece of information and both location and body acceleration as additional context information. In the next subsections the rules that contain the context dependencies for each sensor type and their combination are presented.

7.1 Inertial + Location FD

When the two types of sensors are combined the FD primarily relies on the recognized activity; the additional context information consists of the location and the body movement. As an example, a fall situation is defined by each of the following rules:

- ($A[t1, t2]$ = "lying") ← ($B\_M[t1, t2]$ = "no" $\wedge$ $L[t1, t2]$ = "floor");
- ($A[t1, t2]$ = "sitting_on_the_ground") ← ($B\_M[t1, t2]$ = "no" $\wedge$ $L[t1, t2]$ = "floor");
- ($A[t1, t2]$ = " *on_all_fours* ") ← ($B\_M[t1, t2]$ = "no" $\wedge$ $L[t1, t2]$ = "floor").

where $A[t1, t2]$ represents the recognized activity in the time interval $t = [t1, t2]$, $B\_M[t1, t2]$ represents the body movement in the interval and $L[t1, t2]$ is the location in the same interval.

We used assumptions that the elderly do not usually lie or sit on the ground and are not on all fours for more than t seconds while not moving. The value for t was chosen to be 10 seconds. This is long enough for a reliable recognition, but still negligible compared to the time needed for help to arrive.

7.2 Inertial FD

In this section we present the context-based reasoning when a conclusion is inferred based on inertial sensors only and therefore a fall situation is defined by the activity and the body accelerations.

Previous experiments showed that it is possible to detect a straightforward (fast) fall by using only TBA; however, lots of false positives appeared in other fall-like events: quickly lying down on a bed, quickly sitting on a chair, etc. Therefore, a potential fall detected by TBA was confirmed by the body movement and additional context information, i.e., the user's activity. As an example, a fall situation is defined by each of the following rules:

- (TBA$[t1]$ = "yes") ← ($A[t1, t2]$ = "lying" $\wedge$ $B\_M[t1, t2]$ = "no");
- ($A[t1, t2]$ = "sitting_on_the_ground") ← ($B\_M[t1, t2]$ = "no");
- ($A[t1, t2]$ = "on_all_fours ") ← ($B\_M[t1, t2]$ = "no").

TBA$[t1]$ = "yes", represents the time when a large-acceleration fall pattern is detected.

7.3 Location FD

Since the location sensors are better at AR than detecting fall accelerations, FD is based on the activity that may result from a fall, and uses location as the context. The first advantage compared to the stand-alone, inertial FD was the location information: the system was aware of some predefined "safe" locations, like the bed. The second advantage was the z coordinate of the sensor location, enabling us to figure out the height of the body and distinguish, for example, sitting on the floor from sitting on a chair. An example of a rule structure is presented here:

- ($A[t1, t2]$ = {"lying" $\vee$ "sitting_on_the_ground" $\vee$ "on_all_fours"}) ← $L[t1, t2]$ = "floor".

8 Experimental Setup

8.1 Experimental Scenario

We designed a complex, 15-minute test scenario specifically to investigate events that might be difficult to recognize as falls or non-falls. This scenario (Table 1) was created in consultation with a medical expert. The numbers in parentheses represent the event numbers for easier referencing throughout the text. The events were recorded in a single recording, including all the events.

Table 1. Events in the scenario and their description

	#	Event	Description
Fall Events	(1)	Fast fall (tripping)	Performed in different ways: forwards, backwards or on the sides.
	(2)	Slow fall (fainting)	Losing consciousness and slowly falling to the ground (trying to hold onto furniture).
	(3)	Falling when trying to stand up	Trying to stand up from a chair, but has difficulties and slowly falls to the ground, ending up in a sitting position on the ground.
	(4)	Sliding from a chair	Person is sliding from a chair and ends up sitting on the ground.
Fall-like Events	(5)	Quickly lying down on a bed	Person is quickly lying down on a bed.
	(6)	Quickly sitting down on a chair	Person is quickly sitting down on a chair.
	(7)	Searching for something on the ground	Person first goes on all fours and after this goes to lying on the ground.
Normal Events	(8)	Sitting down	Sitting down on a chair normally.
	(9)	Lying down	Lying down on a bed normally.
	(10)	Walking	Walking sequences between events.

Because typical fast falls are easy to detect, only one such fall (1) was included. Three atypical falls (2, 3 and 4) were included to test the use of the contextual activity information, i.e., that a person is not expected to sit/lie on the ground (as opposed to the chair/bed). Furthermore, two events (5 and 6) we included that involve high acceleration and could thus be misclassified as falls by acceleration-based methods (such as TBA). However, the methods that use the activity and location as contextual information should be able to detect that these are non-fall events. An event (7) was included that involves voluntarily lying on the ground, which could mislead the methods that use information other than acceleration. The events 8, 9 and 10 are normal and were included to verify that all the methods work correctly during normal activities.

The experimental scenario was recorded with all 6 inertial and 4 location sensors. Afterwards, the FD was tested with all 1023 combinations of sensors (single type, as well as both types). The scenario was recorded by 11 young, healthy volunteers (7 males and 4 females). It was repeated 5 times by each person, resulting in 55 recordings and a total of 550 events for the FD. Testing elderly people was not feasible because of the scenario complexity and for safety reasons, but the volunteers were advised how to act by the medical expert.

Additionally, the data for three more people was recorded for tuning the basic parameters, e.g., thresholds, preliminary tests, choosing the best algorithms (details can be obtained from the authors).

8.2 Evaluation Technique

To evaluate the FD, one must decide how to weight the missed falls and the false alarms. Both are important: missing a fall may endanger a person's health, while false alarms make the system unlikely to be used in real life. Therefore, we used the F-measure (F), which weights missed falls and false alarms equally. It is defined as a harmonic mean of recall (the percentage of the events recognized as falls/non-falls from all falls/non-falls events) and precision (the percentage of the events truly being falls/non-falls of all predictions for falls/non-falls). However, for more detailed results, the true positive and true negative rates are presented for the fall and non-fall events, respectively.

9 Experimental Results and Discussion

Fig. 4 presents a matrix (5 × 7) representation of the best sensor combinations. The inertial sensors are shown on the x axis and the location on the y axis. Each rectangle in the matrix contains the sensor placements and the achieved F-measure as a percentage (marked with F in Fig. 4). For example, the (2 × 3) rectangle represents a combination of 2 location and 3 inertial sensors. It is the best of all combinations according to the F-measure = 99.7%. The dotted lines (diagonal) connect the rectangles that have the same number of sensors. Along each dotted line the best (according to the F-measure) rectangle is marked with a white circle. These rectangles represent the best combination given the number of sensors.

Tests of the statistical significance were performed. The best sensor combinations for each number of sensors (white circles in Fig. 4) and each sensor type (rectangles on the axes) were tested separately. Because of the small number of folds (11) and because the individual samples (folds) are paired (the same person's data for each combination), we used the paired Student's T-test with a significance level of 5%.

Analyzing the results achieved with the inertial sensors only (horizontal axis rectangles), one can see that the only important improvement is detected when using two sensors instead of one. After this, adding up to five sensors did not significantly improve the F-measure; including a sixth sensor even decreased the performance. In another observation, the chest sensor is the most suitable for inertial FD, because it is in all the sensor combinations.

For the location FD, an increase in the number of sensors increases the performance all the way. The statistical tests proved that there is a significant difference in the performance of a system using one, two, three and four location sensors. Like with the inertial FD, the chest is the best-performing placement.

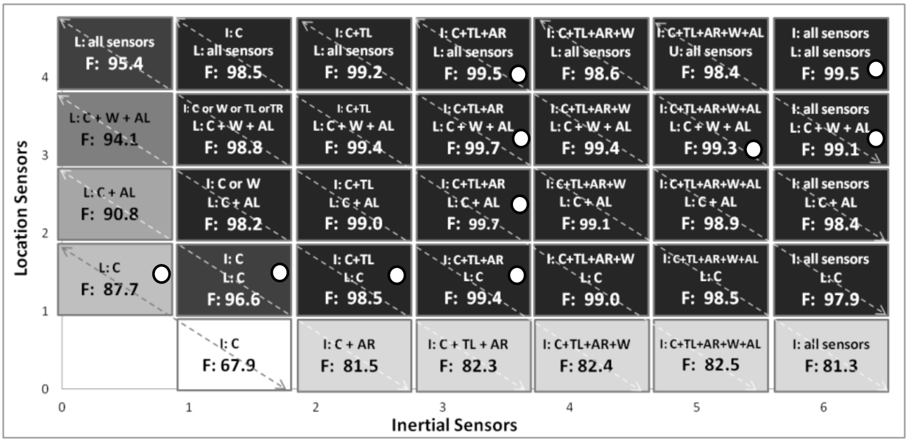


Fig. 4. Matrix representation of the best sensor combinations using the Inertial (I) and Location (L) sensors. F - overall F-measure, C - Chest, W - Waist, RA - Right Ankle, LA - Left Ankle, RT - Right Thigh, LT - Left Thigh.

The statistical tests for the combined FD showed that the difference in performance is statistically significant only when the system is using two and three sensors. Four sensors or more do not significantly increase the performance of the system.

The parts of the graph with a smaller number of sensors are of the greatest interest for practical use. The combination of sensors clearly outperforms the individual sensor types. For example, the performance values of the system using two sensors are 81.5% and 90.8%, for the inertial and location sensors, respectively. Their combination improves these results by 15 p.p. and 6 p.p., respectively. This is the case for each number of sensors (dotted lines): the combination of two sensor types is better than each of the types used separately. Furthermore, one can put the two sensor types in one sensor enclosure. The performance of the system using only one sensor type is 68% and 88% for the inertial and location sensor, respectively. Combining them in one enclosure improves these results by 29 p.p. and 9 p.p., respectively.

The rest of the discussion is a detailed analysis of the results achieved by the simplest and the best (statistically significant) combinations of the inertial-only, location-only and both types of sensors. The sensor types and placements are shown in Table 2 and the results are presented in Table 3.

Table 2. The simplest and the best (statistically significant) combinations of the inertial-only, location-only and both types of sensors

Sensor types/placements	The simplest combination	The best combination
Inertial sensors	Chest	Chest + Right ankle
Location sensors	Chest	All four sensors
Combined sensors	**Inertial**: Chest **Location**: Chest	**Inertial**: Chest + Right ankle **Location**: Chest

The events in Table 3 are divided into two groups: fall and non-fall (fall-like and normal) events. The number for each event is the percentage of all fall/non-fall events being correctly recognized as fall/non-fall (true-positive rate/true-negative rate). The last row represents the overall F-measure.

Table 3. Detailed FD results for each event and each context-based FD method

		Context-based Reasoning					
		Inertial (Activity + TBA + Movement)		Location (Activity + Location)		Combination (Activity + TBA + Movement + Location)	
		simplest	best	simplest	best	simplest	best
Fall Events	(1) Tripping – Quick falling	100	100	96	100	100	100
	(2) Fainting – Falling slowly	11	11	100	100	100	100
	(3) Falling from a chair slowly	68	98	95	95	99	99
	(4) Sliding from a chair	72	99	97	97	98	99
Non-Fall Fall-like Events	(5) Sit down quickly on a chair	55	97	75	89	91	98
	(6) Searching on the ground	85	88	25	78	80	89
	(7) Quickly lying down on a bed	34	34	100	100	100	100
Non-Fall Normal Events	(8) Sitting normally	68	98	80	93	93	98
	(9) Lying normally	100	100	100	100	100	100
	(10) Walking	97	100	92	97	100	100
	Overall F-measure in %	67.9	81.5	87.7	95.4	96.6	98.5

The first two columns show the results achieved by the inertial FD. The first event in Table 3, tripping, is a typical fall that was recognized accurately because of the TBA rule. The second event, which is falling slowly, was difficult to recognize because of the low acceleration during this event. For this event, additional contextual information was necessary (e.g., the location of the user). The effect of the activity information of the user can be seen in the fall events that end with sitting on the ground (events 3 and 4). In these cases the AR model correctly recognized sitting on the ground. On the other hand, this has a negative impact on the performance when the sitting event is analyzed (events 5 and 8). In this case, the AR model is not accurate enough and recognizes sitting on the ground, resulting in a false positive. This issue is solved by including more sensors, which improves the AR method (e.g., the column Inertial-best).

The location FD was using the activity and the location information. But because of the location, it recognized all falls with high accuracy (events 1 to 4). However, some problems still exist in non-fall events, because of the relatively low accuracy of the AR model. Namely, sitting (events 5 and 8) and searching on the ground (event 6) were misclassified as sitting on the ground or lying (on the ground), causing the system to detect a fall during non-fall events. Improvements in the performance can be seen when the number of sensors is increased (the column Location-best), due to the improvements in the AR method.

The last two columns show the results achieved with the combination of both types of sensors and the full context as presented in Subsection 7.1. The improvements are clear in all of the events. The overall performance when two sensors (one inertial and one location) are used is 96.6%. Some problems only appear in non-fall events that end with sitting (5 and 8) and the searching on the ground event (6). The reason lies in the AR method, which misrecognizes the appropriate activities (sitting and on all fours). These problems are solved by including one more inertial sensor, which significantly improves the AR model and consequently the FD (the last column in Table 3).

To summarize, the best-performing combination of sensors is two inertial sensors (chest + left thigh) and one location sensor (chest). With this combination, the context-based FD method achieves an F-measure of 98.5%, mostly because of the location sensor, which provides the context location information. However, the best practical solution with the minimum number of sensors is the combination of the chest inertial and chest location sensor, achieving an F-measure of 96.6%.

Finally, two commonly used methods in the literature, the threshold-based approach (TBA) and the machine-learning approach (MLA), are tested for comparison. The results are shown in Fig. 5, by presenting the true-positive or true-negative rate for each fall or non-fall event, respectively.

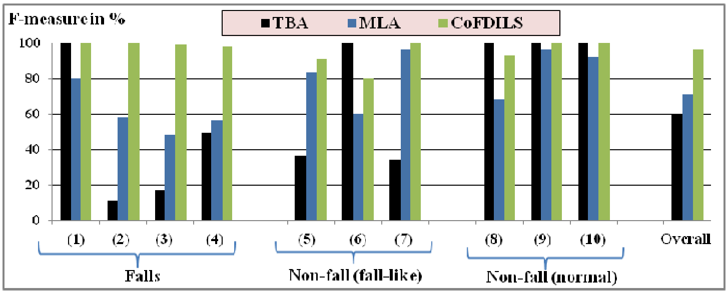


Fig. 5. Comparison of the FD results achieved by: our Context-based approach, the Machine-learning approach (MLA), and Threshold-based approach (TBA). The event numbers correspond to the events given in **Table 1**.

The TBA is described in Subsection 6.2. More details about the MLA can be found in our previous work (Luštrek et al. [14]). The basic principle of MLA is that a machine-learning model is trained to detect a fall event. In our case, features extracted from the chest-inertial and chest-location sensor data were used. Therefore, the contextual location information was implicitly (through features) introduced in the MLA.

The overall results showed that our method, in which the context is explicitly encoded with rules, outperformed the other two methods, which use: implicit context information (MLA) or only accelerations (TBA). The TBA outperformed our CoFDILS only in two events (6, 8); however, this was due to the one-sided performance of the TBA (detects only large accelerations) at the expense of the overall performance.

10 Conclusion

We presented a novel approach, i.e., CoFDILS, to fall detection by combining inertial and location sensors using a general context-based schema. The method exploits three components, i.e., the activity of the user, the body accelerations and the location, to detect a fall situation. The decision of each component based on the same type of input data is re-evaluated in the context of the other two components. There are 168 combinations of context dependencies, based on the values for each component; however, in our tests it turned out that only a couple of context-based relations capture most of the context information in the FD domain. Currently, the context-based reasoning rules were provided by an expert. The automation of learning the context relation is considered as future work.

We tested the performance with all possible combinations of the six inertial and four location sensors to find the best sensor placements, using the context-based decision schema. The evaluation was performed on a complex test scenario; it included real life, realistic events that were difficult to recognize as falls or non-falls. The results showed that by combining the two types of sensors it is possible to detect complex fall situations by using the activity and the context information from both types of sensors. It is essential that both sensor types are employed, since they provide complementary information about the user's situation. Finally, the best practical solution is the chest-sensor placement with a single sensor enclosure, combining one inertial and one location sensor.

Acknowledgements. This work was partly supported by the Slovenian Research Agency under the Research Programme P2-0209 Artificial Intelligence and Intelligent Systems and partly from the European Community's Framework Programme FP7/2007-2013 under grant agreement No.214986. The authors would like to thank Boštjan Kaluža for the helpful suggestions and discussions.

References

1. Breiman, L.: Random forests. Machine Learning 45, 5–32 (2001)
2. Dey, A., Salber, D., Abowd, G., Futakawa, M.: The conference assistant: Combining context-awareness with wearable computing. In: International Symposium on Wearable Computers, San Francisco, USA (1999)
3. Friedman, S.M., Munoz, B., West, S.K., Rubin, G.S., Fried, L.P.: Falls and Fear of Falling: Which Comes First? A Longitudinal Prediction Model Suggests Strategies for Primary and Secondary Prevention. Journal of the American Geriatrics Society, 1329–1335 (2002)
4. Gjoreski, H., Gams, M., Chorbev, I.: 3-axial accelerometers activity recognition. In: ICT Innovations, pp. 51–58 (2010)
5. Gjoreski, H., Luštrek, M., Gams, M.: Accelerometer Placement for Posture Recognition and Fall Detection. In: The 7th International Conference on Intelligent Environments (2011)
6. Gillespie, L.D., Robertson, M.C., Gillespie, W.J., Lamb, S.E., Gates, S., Cumming, R.G., Rowe, B.H.: Interventions for preventing falls in older people living in the community (Review). The Cochrane Library 4 (2009)
7. Gimon, D., Gjoreski, H., Kaluža, B., Gams, M.: Using accelerometers to improve position-based activity recognition. In: 13th International Multi-Conference Information Society, pp. 15–18 (2010)

8. Jantaraprim, P., Phukpattaranont, P., Limsakul, C., Wongkittisuksa, B.: Evaluation of fall detection for the elderly on a variety of subject groups. In: The 3rd International Convention on Rehabilitation Engineering & Assistive Technology (2009)
9. Kaluza, B., Dovgan, E.: Glajenje trajektorij gibanja cloveskega telesa zajetih z radijsko tehnologijo. In: Information Society, Ljubljana, Slovenia (2009)
10. Khawandi, S., Daya, B., Chauvet, P.: Automated monitoring system for fall detection in the elderly. International Journal of Image Processing (IJIP) 4(5) (2010)
11. Li, Q., Stankovic, J., Hanson, M., Barth, A.T., Lach, J., Zhou, G.: Accurate, Fast Fall Detection Using Gyroscopes and Accelerometer-Derived Posture Information. In: Sixth International Workshop on Wearable and Implantable Body Sensor Networks, pp. 138–143 (2009)
12. Li, Q., Stankovic, J.: Grammar-based, posture- and context-cognitive detection for falls with different activity levels. In: WH 2011: 2nd Conference on Wireless Health (2011)
13. Lin, H.-T., Hsieh, Y.-J., Chen, M.C., Chang, W.: ActionView: a movement-analysis ambulatory monitor in elderly homecare systems. In: IEEE International Symposium on Circuits and Systems, pp. 3098–3101 (2009)
14. Luštrek, M., Gjoreski, H., Kozina, S., Cvetković, B., Mirchevska, V., Gams, M.: Detecting Falls with Location Sensors and Accelerometers. In: Twenty-Third IAAI Conference (2011)
15. Machajdik, J., Zambanini, S., Kampel, M.: Fusion of Data from Multiple Cameras for Fall Detection. In: Workshop on Behaviour Monitoring and Interpretation, BMI 2010, pp. 1–7 (2010)
16. Rubenstein, L.Z., Josephson, K.R.: The epidemiology of falls and syncope. Clinics in Geriatric Medicine 18(2), 141–158 (2002)
17. Shan, S., Yuan, T.A.: Wearable Pre-impact Fall Detector using Feature Selection and Support Vector Machine. Systems, 1686–1689 (2010)
18. Tinetti, M.E., Williams, C.S.: Falls, Injuries Due to Falls, and the Risk of Admission to a Nursing Home. The New England Journal of Medicine 337, 1279–1284 (1997)
19. Ubisense RTLS website, http://www.ubisense.net (accessed June 2012)
20. World Health Organization (WHO). Global brief for World Health Day 2012. Good health adds life to years, http://whqlibdoc.who.int/hq/2012/WHO_DCO_WHD_2012.2_eng.pdf (accessed June 2012)
21. Witten, I., Frank, E.: Data Mining: Practical machine learning tools and techniques, 2nd edn. Morgan Kaufmann (2005)
22. XSens sensors website, http://www.xsens.com (accessed June 2012)
23. Zinnen, A., Wojek, C., Schiele, B.: Multi Activity Recognition Based on Bodymodel-Derived Primitives. In: Choudhury, T., Quigley, A., Strang, T., Suginuma, K. (eds.) LoCA 2009. LNCS, vol. 5561, pp. 1–18. Springer, Heidelberg (2009)
24. Zhang, T., Wang, J., Xu, L., Liu, P.: Fall Detection by Wearable Sensor and One-Class SVM. Intelligent Computing in Signal Processing and Pattern Recognition 345, 858–863 (2006)

Enhancing Accelerometer-Based Activity Recognition with Capacitive Proximity Sensing

Tobias Grosse-Puppendahl[1], Eugen Berlin[2], and Marko Borazio[2]

[1] Fraunhofer IGD, Darmstadt, Germany
tobias.grosse-puppendahl@igd.fraunhofer.de
[2] Embedded Sensing Systems, Technische Universität Darmstadt, Germany
{berlin,borazio}@ess.tu-darmstadt.de

Abstract. Activity recognition with a wearable accelerometer is a common investigated research topic and enables the detection of basic activities like sitting, walking or standing. Recent work in this area adds different sensing modalities to the inertial data to collect more information of the user's environment to boost activity recognition for more challenging activities. This work presents a sensor prototype consisting of an accelerometer and a capacitive proximity sensor that senses the user's activities based on the combined sensor values. We show that our proposed approach of combining both modalities significantly improves the recognition rate for detecting activities of daily living.

Keywords: activity recognition, capacitive proximity sensors, ambient assisted living, user context.

1 Introduction

Persons affected from physical and mental restrictions and their care givers can profit significantly from unobtrusive activity monitoring solutions. For example, formal care givers could equip a person with a wearable activity monitoring system that evaluates the course of a disease or the influence of an adapted medication [14]. This scenario is especially relevant for people suffering from dementia who have limitations in organizing their daily activities. A simple wearable activity monitoring solution could analyze activities performed in daily life, such as drinking, eating and sleeping habits.

Sensing a person's activity is an active research topic with a raising interest due to the advancement in mobile phone technology. These devices include multiple sensors and therefore enable the recognition of daily activities [7]. Current wearable activity recognition systems are able to unobtrusively capture and recognize a person's activities throughout the whole day. These systems often rely on inertial sensor data that are captured by wearable sensors embedded in a mobile device [7] or attached to the body [17]. Usually, single sensor modalities are used or duplicated to detect the activities. However, it is a great challenge to identify many activities just by using a single modality like the accelerometer.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 17–32, 2012.

Capacitive proximity sensors on the other hand can indirectly measure the distance and nature of a grounded object within reach. This means that the measurement result depends on the object's distance, its size and the material it is made of. In this work, we show that an accelerometer and a capacitive proximity sensor can be used to improve activity recognition in activities of daily living that rely heavily upon object usage. Therefore, we obtained an open-hardware and open-source wrist-worn activity data logger [1] and integrated a capacitive proximity sensor into its wristband.

There are several intuitive examples for which a combination of accelerometer-based activity recognition with a capacitive proximity sensor reveals its strength. For example, it may be conducted which material is placed underneath the wristband. The capacitive proximity sensor would return a different measurement result for a hand placed on a couch covered with fabric than a hand placed on a wooden table. Moreover, the approximate distance from the wristband to objects can be exploited to identify activities like grasping into a locker or a refrigerator to prepare food.

The remainder of this paper is structured as follows: First, Section 2 is dedicated to related work work in activity recognition, in particular considering approaches with multiple sensing modalities. Section 3 presents the hardware based on which a wrist-worn sensor prototype was built, fusing an inertial data logger and a capacitive sensor integrated into a wristband. The experimental setup, the scenario with daily activities and the activity recognition evaluation results, showing the performance boost of the capacitive sensor unit, are given in Section 4. The paper is wrapped up with a conclusion enumerating the findings of the evaluation, as well as pointing out future research potential.

2 Related Work

Activity recognition research relying on wearable sensors mostly considers inertial data from the participants body to infer performed activities, such as in the works of [17,4,20,3]. The acceleration data is often augmented with data from sensors such as gyroscopes [12], magnetometers [2], ambient light [6] or ambient and skin temperature [13], aiming to extract a more detailed environmental user context. In [26], the authors use heart rate information as an addition to the accelerometer data to detect activities like lifting and lowering loads or even digging. In [23], workshop assembly activities are detected by augmenting acceleration sensors with microphones.

The works of Fishkin et al. [9] and Patterson et al. [15] show that detecting touched and used objects can be very helpful for activity recognition. By using RFID readers that are embedded in gloves or bracelets at the wrist and RFID tags attached to various objects of interest, one can detect the object grasped and used by the user, thus aiding the activity recognition in various application scenarios, such as activities of daily living [22] and [16], activity tracking in car manufacturing [21], or household and gardening activities [5].

Our approach to enhance the inertial data from a wearable sensor is comparable to the RFID scenarios just mentioned, as it also relies on a single wearable sensor and an unobtrusive deployment. The main difference lies in the fact that we do not consider an accurate detection of tagged objects, but the proximity to various unknown objects and the environment.

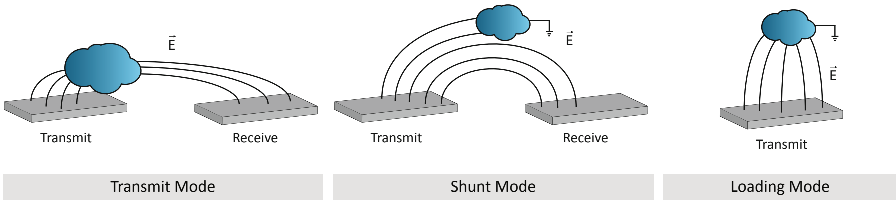


Fig. 1. Three different capacitive proximity sensing modes can be distinguished [18]. The sensing electrodes build up an electric field to objects in the environment, illustrated with a cloud.

In the field of capacitive proximity sensing, three different measurement modes (shown in Figure 1) were identified by Smith et al. [19]: transmit mode, shunt mode and loading mode. Transmit mode is based on a varying electric potential coupled to an object that can be measured by a capacitive proximity sensor next to that object. Shunt mode applies two electrodes, a transmit and a receive electrode, that can measure capacitance changes produced by objects disturbing the electric field between the two electrodes. In loading mode, a single electrode builds up an electric field to any grounded object in the environment. By measuring the capacitance, conclusions can be made upon the proximity and nature of an object. In our work we apply loading mode since it requires only a single electrode that can be integrated invisibly into the wristband.

Capacitive proximity sensing faces the great advantage of being robust against changing lighting conditions and occlusion. Moreover, sensing electrodes can be integrated invisibly into the environment. On the other hand, the exact distance to objects can only be approximated since the object's surface, its conductivity and grounding has influence on the measurement result. A single sensor will thus deliver data that has a certain degree of ambiguity. Due to the nature of capacitive proximity sensors, they can be prone to errors in environments with strong and rapidly changing electric fields. This, however, is usually not an issue when considering activities of daily living.

A great variety of capacitive sensors and measurement techniques exists [19]. The most common sensing principle, the loading mode, is based on running numerous charge and discharge cycles of the virtual capacitor that is created by the electrode and the environment. Depending on the charge and discharge times, one can infer the corresponding capacitance. This sensing principle is applied by Wimmer et al. in [24] who presented a toolkit for capacitive proximity sensing. In previous works, capacitive proximity sensors were applied in various fields of human-computer-interaction. Wimmer et al. and Grosse-Puppendahl et al. presented gesture recognition systems [25,10] as well as smart furniture that can sense human activities [25] and classify human postures [11]. Cheng et al. have investigated the possibility of using capacitive sensors for activity recognition by measuring shape changes of muscles and skin [8].

3 Hardware

This section presents the two components of our hardware prototype, the wrist-worn activity data logger tailored to capture acceleration data, and the capacitive proximity sensor used for distance measurements.

3.1 Activity Data Logger

The HedgeHog sensor [1] is a custom designed wearable data logger aiming at long-term deployments in activity recognition scenarios. Due to its small form-factor (37x32x16mm) and weight, this wrist-worn sensor is an unobtrusive way to record relevant motion data.

The sensor node itself is built around the low-power Microchip microcontroller (PIC18F46J50) featuring an accelerometer sensor (ADXL345) to capture human motion, light and ambient temperature sensors and a microSD flash card for locally storing the sensor data. The sensor is powered by a $200mAh$ lithium polymer battery, which allows for two weeks of continuous recording on a single battery charge. A USB port is used to configure the sensor (e.g. setting the sensitivity of the accelerometer), to access the stored sensor data, and to recharge the battery. A plastic case packages and protects the sensor to be worn at the wrist (Figure 2).

Fig. 2. The inertial data logger featuring a low-power microcontroller, a 3 axis accelerometer, a microSD flash card for storing the sensor data and a USB connector for accessing the data (on the right) is powered by a small lithium polymer battery and is packaged into a plastic case to be worn at the wrist (a version with an OLED display).

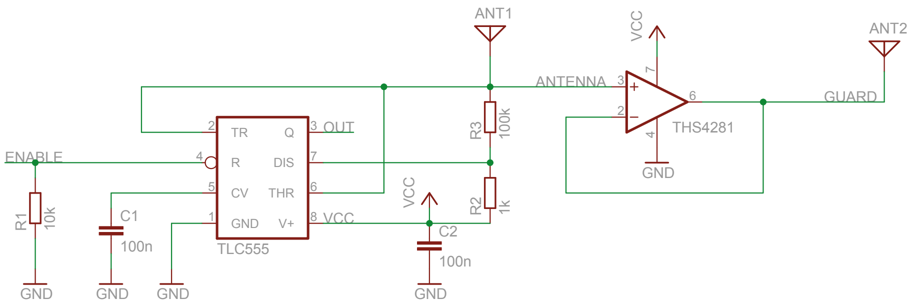


Fig. 3. The sensing circuit is based on a timer with an operational amplifier that acts as a voltage follower. The wire that is labeled with "antenna" leads to the sensing electrode, whereas the wire labeled with "guard" leads to the shield electrode.

The 3D accelerometer sensor is being sampled at 100Hz, resulting in 10ms equidistant measurements. For efficiency reasons, the sensor data is run-length encoded before being stored locally to the microSD card. The HedgeHog can be extended with further sensors tailoring different application scenarios. For our scenario, we have added a capacitive proximity sensor that is described in detail in the next section.

3.2 Capacitive Proximity Sensor

A wrist-worn capacitive proximity sensor requires a shield that eliminates the influence of the grounded arm directly underneath the sensor. Using this setup, we can detect the proximity to a grounded object in the environments for distances up to 20cm. Especially for mobile devices, it is required that the sensor draws a very small amount of power. Thus, other proximity sensing input modalities like ultrasound or optical measurements are not applicable for this type of mobile application.

The capacitive proximity sensor performs measurements in loading mode. Two electrodes are integrated into the wristband, one sensing electrode and one shielding electrode. The sensor draws a supply current of 1mA at 3.3V when active which qualifies it for wearable proximity sensing applications. In the following a virtual capacitor denotes the capacitance between the sensing electrode and the environment. The capacitance of the sensing electrode to environmental objects increases with closer distances.

The sensing circuit schematic is shown in Figure 3. It is based on a timer that controls the charging and discharging cycles of the virtual capacitor that is built by the sensing electrode and the surrounding environment. The timer toggles from charge to discharge at the time when a threshold voltage at the capacitor is reached. This results in an astable operation with succeeding charge/discharge cycles. When the capacitance of the virtual capacitor increases, the charging time will also increase and vice-versa. Therefore, the capacitance is inversely proportional to the

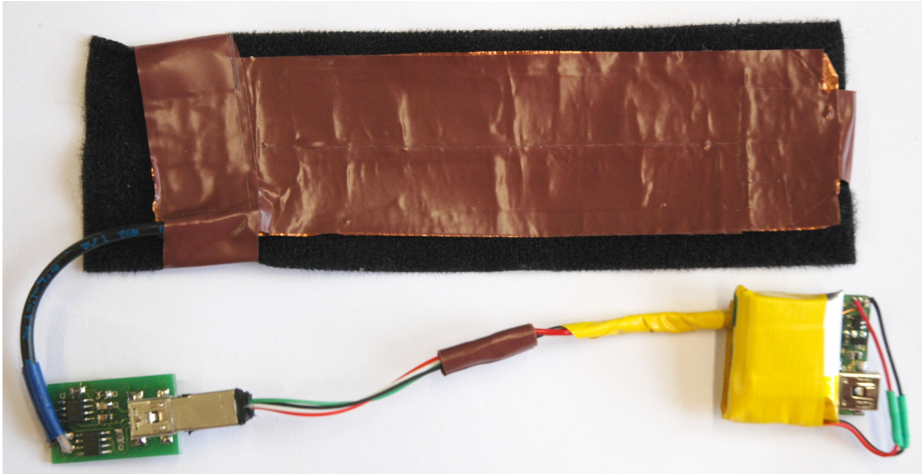


Fig. 4. The hardware prototype at a glance: HedgeHog activity logger at the lower right, the capacitive sensor unit at the lower left, and the wristband with the sensing and the shield electrodes on-top each other. The electrodes are covered with adhesive tape for isolation purposes.

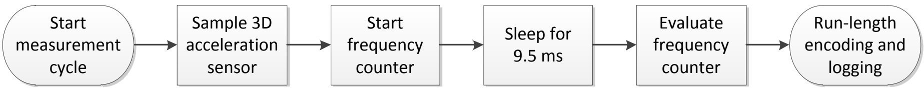


Fig. 5. Overview of the measurement procedure carried out by the HedgeHog sensor: using the microcontroller's Timer0 module in counting mode, the oscillating signal generated by the capacitive sensor circuit can be measured by counting the frequency pulses over a predefined gate time of approximately 9.5ms.

number of charging cycles in a given time span. In order to guard the sensor from measuring the capacitance to underlying objects, a shield electrode is placed directly underneath the measuring electrode. The shield is driven with the same potential as the sensing electrode, such that the capacitance between the two electrodes is negligible. Using this shielding method, the measured capacitance will only be slightly affected by the grounded underlying arm.

Figure 4 shows the the wrist-worn prototype used in the evaluation experiments, with the HedgeHog as the main data logger, the capacitive sensor circuit and the wristband holding the sensing and shielding electrodes.

The operations required for a measurement cycle are illustrated in Figure 5. The proximity sensor board generates a clock signal with varying frequency depending on the charge and discharge cycles. The HedgeHog measures the resulting capacitance by counting the signal's edges over a gate time of approximately 9.5ms. During that counting phase, the microcontroller is sent to sleep in order to reduce power consumption.

4 Experiment

This section presents the experimental setup including the activities and the participants, as well as the findings that were obtained during the evaluation.

4.1 Setup and Scenario

The experiment setup aims to depict a typical scenario of a person in daily life. Especially in the field of Ambient Assisted Living (AAL), it is desired to monitor activities like drinking, preparing lunch and sleeping. A fine-grained monitoring of such activities may help elderly or people suffering from mental diseases to maintain a healthy day/night rhythm and take action if irregularities occur. Figure 4 shows the modified HedgeHog activity logger that has been extended with a capacitive proximity sensor. The wristband has two electrodes, a sensing electrode underneath a slightly bigger shield electrode.

The recorded test set contains the following activities: opening door, sitting on a couch, lying on a couch, putting kitchen equipments from a shelf and out of a locker, making a marmalade sandwich, eating the sandwich, pouring and drinking water, walking and sleeping. The relations of these activities to environmental objects are given in Table 1. Some of those activities are very hard to recognize when the data is limited to a single modality like a 3D accelerometer. For example, sitting at the table and sitting on a couch are very similar activities. We aim to show that the data basis can be significantly improved by the additional input modality.

Table 1. Some details on the activities performed during the experiment and objects directly involved or nearby

activities	objects involved	objects nearby
open door	door knob	door
sitting	chair or couch	body, chair, couch, table
lying	couch	body, couch, cushion
get things	plate, glass, cutlery, bread, marmalade, bottle	shelf, locker, fridge, table
make sandwich	bread, knife, marmalade	table, plate
eating	marmalade sandwich	table, plate, body
drinking	bottle, glass	table, body
sleeping	bed, cushion, blanket	body
walking		body

In order to evaluate if capacitive proximity sensors in wrist-bands can enhance the performance of activity recognition, we have conducted an evaluation with 7 test persons. All test persons received a basic script with the activities they were supposed to perform. They were not given any instructions about the way they are supposed to perform the activities. After manual labelling, we used this test-set as ground truth and performed a 4-fold cross-validation on an support-vector machine (SVM) classifier on each user. The cross-validation was performed once with and once without including the data of the capacitive proximity sensor into the feature set. We chose an SVM classifier because of its high relevance in activity recognition and its fast performance.

The classifier was trained with basic features that were extracted from a sliding window of 1 second width. Our first tests have shown that greater window sizes do not provide better classification results. In order to suppress noise contained in the capacitive proximity sensing data, we applied a moving average filter with a kernel size of 10. The final feature set contained the arithmetic mean, min, max, median and standard variance for each accelerometer axis and the capacitive proximity signal. These simple feature types represent standard features applied in activity recognition. Since we aim to show an improvement using the new modality, the selection of features and classifiers does not represent the primary focus of this paper.

4.2 Evaluation Results

In the following, the performed activities will be analyzed in detail, stating the influence of the capacitive proximity sensor on the classification result. In general, the usage of data provided by the new modality showed improvements in recognition rates reaching from 2.4% up to 10.7% for single activities.

The "opening door" activity has very poor recognition rates without the data from the proximity sensor. The average F-measure could be increased from 35.5% to 46.2%. A plot of the activity is given in Figure 6. The capacitive proximity sensor shows two approaches to the door knob, one for opening the door (2s) and one for closing the door (9s). The acceleration sensor captures relevant data in the time in which the person moves into the room and the hand changes from the outer to the inner door knob (5 - 7s). The recorded data for this

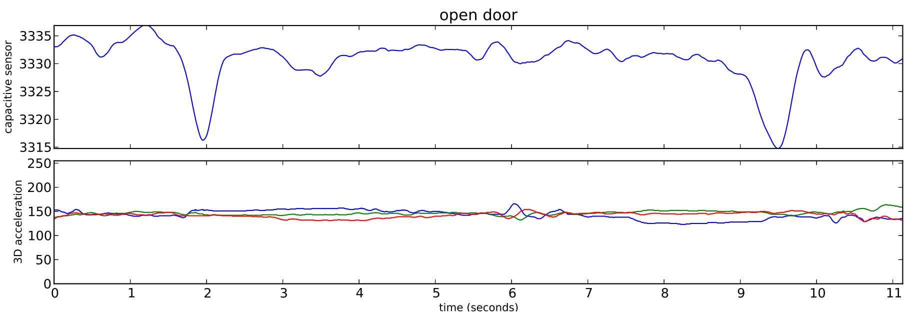


Fig. 6. When the participants entered the apartment, the wrist approached the door knob twice, at the time of opening and closing the door. This fact can be observed in the capacitive proximity data (upper plot) at the beginning and at the end of the activity, whereas the acceleration has little characteristic information (bottom plot).

activity also shows strong correlations between all experiment participants. The confusion matrices show that the "open door" activity was often confused with the "sitting" activity, probably because of the amount of motion on the one hand and the proximity to nearby objects (door, couch or cushions) on the other hand. By using the capacitive sensor data, the recall for that class and confusion with the sitting activity could be improved.

After closing the door, the participants were supposed to sit down on the couch. It turned out that there are great variations of the sitting posture and the corresponding hand positions. Many users tapped with their fingers or hands while sitting, changed their sitting positions very frequently, or were even talking and gesticulating, as shown in Figure 7. In this case, it is obvious that the data from the acceleration sensor is very difficult to interpret as there are numerous changes in the axial orientation of the sensor. However, the capacitive proximity sensor is able to indicate when a hand is placed on the surface of the couch. Especially for this particular participant, the F-measure increased from 50.3% to 60.4%, while the average F-measure improved from 68.0% to 74.4%.

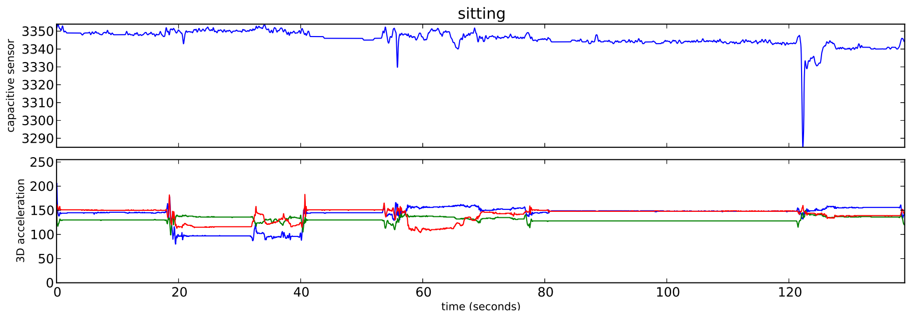


Fig. 7. Example of the "sitting" activity in which the user moved his hands quite frequently (bottom plot). Most of the time the values of the proximity sensor stay more or less constant, probably due to the hands position on the couch's fabric. The sharp peak in the capacitive sensor data (upper plot) occurred when the participant scratched the back of his head.

In the following, the participants were instructed to lie down on the couch. Again, there were great variations in how this activity was performed by the participants. For example, some of them crossed their hands under their head, or placed them on their body. For this class, the average F-measure could only be increased by 2.4%, from 81.2 to 83.6%. Considering some participants, the activity was often confused with the "make sandwich" class. By using the proximity modality, the confusion between the two classes could be reduced.

After that, the participants were asked to walk over to the kitchen and to put food and dishes from a shelf and a locker on the table. This activity involved direct interactions with various objects as well as proximity to furniture in the room (see Table 1). The capacitive proximity sensor was able to capture the proximity to the shelf and to the table (see Figure 8). The average F-measure

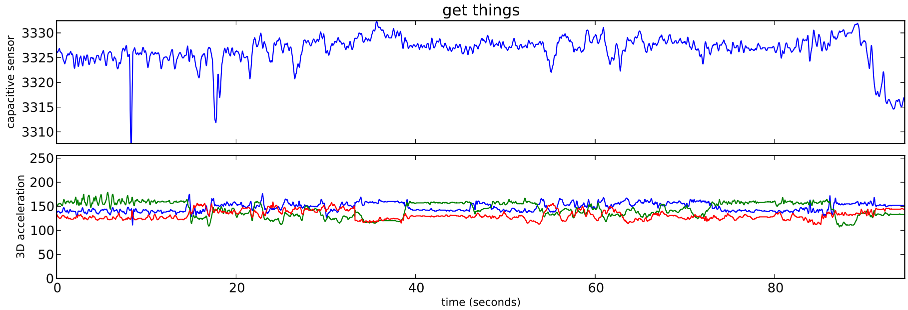


Fig. 8. An example of the "get things" activity, where the participants had to get food and dishes from shelves and lockers. The proximity sensor peaks in the beginning (9s and 19s) indicate immediate proximity to shelf, and to the locker (55-63s) in the kitchen (upper plot). The signal drop at the end results from the participant placing his hand on the table when she was finished.

for this activity is rather low, but improved by 6.8% from 53.8% to 60.6%. The worst performing participants for this activity reached an F-measure of 46.8% without and 53.5% with the capacitive sensor, while the best performing one reached 62.0% and 64.5% respectively. The low performance results from confusions with other activities, with a higher tendency to the "make sandwich" activity across all participants. This is most likely due to various objects involved in both activities, and the fact that 1-second features are obtained. The capacitive sensor modality has a more positive impact reducing the confusion with other activities.

Figure 9 shows an example instance of preparing a bread with marmalade. It is notable that the acceleration data does not seem to provide any characteristic patterns, while the proximity sensor indicates a table, plate, or other objects

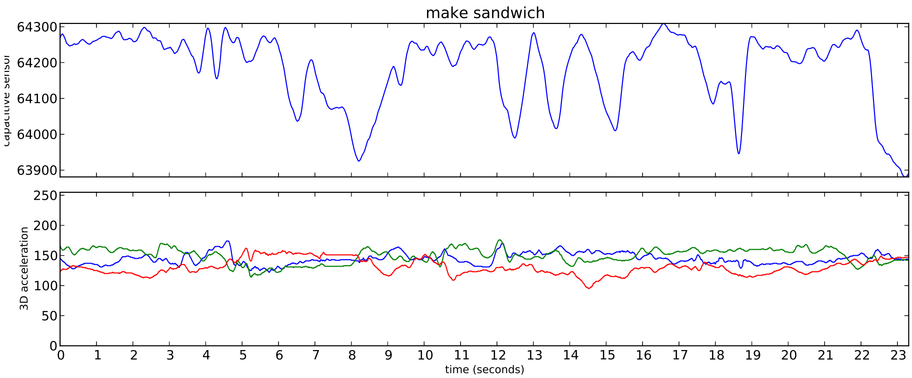


Fig. 9. An example of the "make sandwich" activity, where the participants had to put marmalade on a slice of bread. The proximity sensor indicates the closeness to the table, while the acceleration sensor shows recurring hand motions.

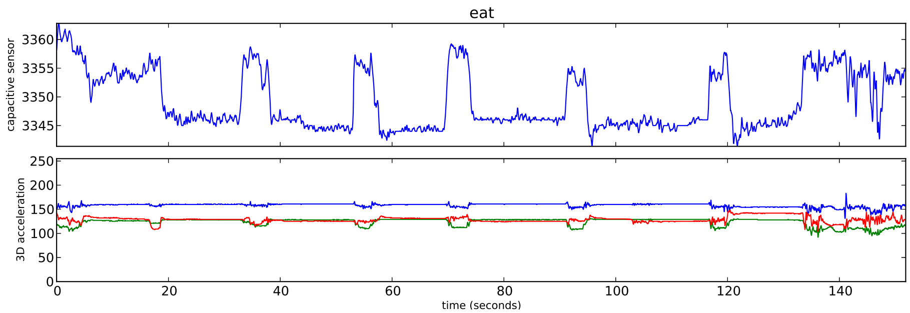


Fig. 10. An example of a participant eating a marmalade sandwich, taking 5 bites from it. After each bite, the hand is placed on the table, which can be recognized both in the acceleration as well as the proximity data plots.

in immediate distance. This activity showed a high improvement in the average F-measure by 10%, from 49.0% to 59.8%, where the data delivered by the capacitive proximity sensor is taken into account. The "make sandwich" class was often confused with the "sitting" class for some users, probably due to lots of motion during the sitting, as mentioned previously. For other users, "make sandwich" was confused with "eating" or "drinking". Using the new input modality, confusion across users could be reduced.

When considering the eating activity, the impact of the new capacitive proximity sensor on the classification performance is quite low, as the chosen features are able to distinguish it from other activities. Some of the participants ate their sandwich leaving their hand close to the mouth, while others moved their hand up and down putting their sandwich aside on the plate, which also results in the performance range from 71.2 to 88.1% without and 77.5 to 90.6% with the proximity data. An example of an eating activity is shown in Figure 10, where the participant took a few bites from the sandwich while putting it down every time. The average F-measure increased slightly by 2.7% (from 79.5% to 82.2%).

The activity "drinking water" is depicted in Figure 11. The participant took a few sips from the glass, with leaving the hands positioned. These motions can be easily detected in the acceleration as well as proximity data. The accelerometer data shows that there periodic up- and down-movements while the capacitive proximity sensor delivers data that is associated to the proximity of the table. The F-measure lies at 48.4% without and 54.4% with the proximity data taken into account, resulting in a gain of 6%.

Regarding the walking activity one can identify periodic changes in the acceleration as well as in the measured capacitance, illustrated in Figure 12. While walking, the capacitance between the wristband and the leg increases when the wristband is located close to the body and decreases when the wristband moves away. There were problems distinguishing this activity from "get things" that could be improved by using the new input modality. The classification improvement for this activity accounts to 12.2% boosting the average F-measure from

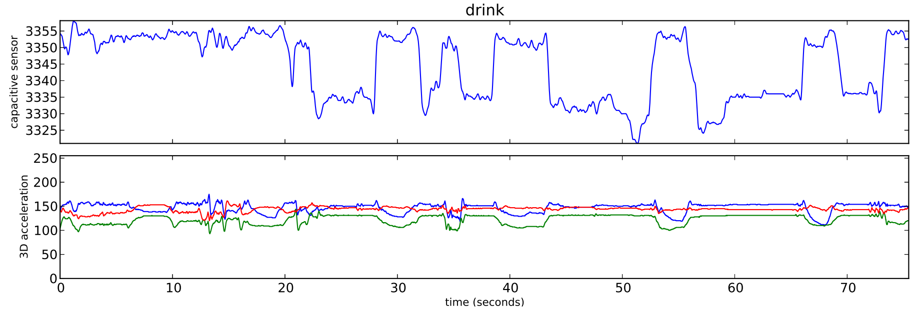


Fig. 11. An example of the "drinking" activity. The participant first pours some water into the glass and then takes three drinks of water. After each sip, he returns his arm to the table which can be observed in the characteristic patterns of the proximity sensor.

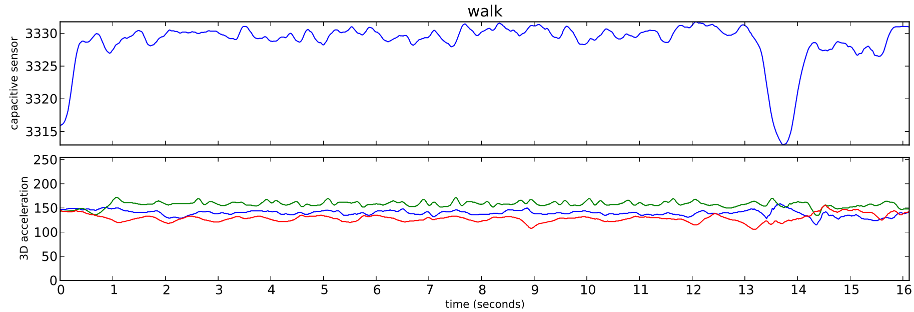


Fig. 12. An exemplary instance of the class "walking". The acceleration sensor and the proximity sensor show periodic recurring patterns that are related to the pendulum-like arm movement and the proximity to the person's body during those movements.

41.7% without to 53.9% with the new sensor. Due to the low performance results and the characteristic periodic signal shape, it would help to consider frequency domain features, as it is often applied in related work.

The sleeping activity (an example shown in Figure 13) was classified with an average F-measure of 83.2%, which increased to 86.9% when using the proximity data. In this case, the capacitive proximity sensor is able to capture the surrounding cushions and blankets, as well as the body or head of the participant. The accelerometer data and the capacitive proximity data have larger periods of low variance, which is a cause for confusion with the "lying on the couch" activity.

Figure 14 depicts two confusion matrices for an exemplary participant from our evaluation, once without and once with including the proximity data. In most activities, an enhancement in the number of correctly classified instances is observable. The "lying" activity's recognition performance could benefit a lot from the proximity data, improving both precision and recall.

A better classification performance can also be observed for the "get things" class that includes interactions with a multitude of objects in the environment.

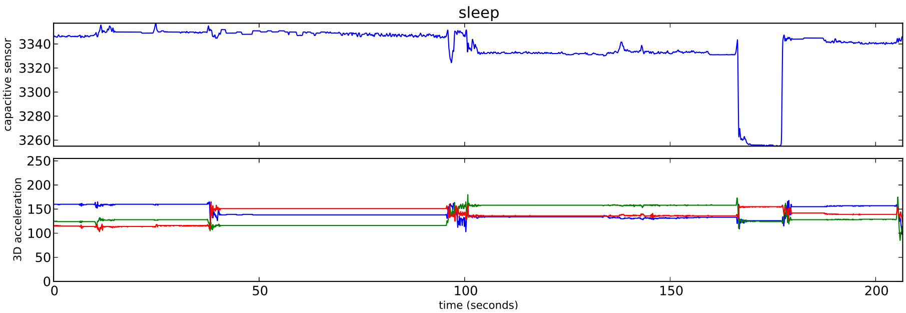


Fig. 13. During the sleeping activity the data from both sensors remains constant for large time spans. The capacitive sensor plot shows the coverage of the arm with either cushions, blankets or the proximity to the mattress, other parts of the bed, the head or body of the participant.

without proximity data									with proximity data									
a	b	c	d	e	f	g	h	i	a	b	c	d	e	f	g	h	i	← classified as
8	0	0	0	0	0	0	2	1	**11**	0	0	2	1	0	0	0	1	a = open door
0	**140**	0	5	2	0	0	0	0	0	**140**	0	5	2	0	0	0	0	b = sitting
0	0	**294**	1	9	10	1	1	0	0	0	**309**	1	0	4	0	0	1	c = lying
1	7	1	**50**	17	9	1	4	1	2	6	0	**67**	14	9	1	2	1	d = get things
0	2	7	7	**102**	42	5	0	2	1	1	0	13	**120**	24	7	1	1	e = make sandwich
0	0	3	5	21	**288**	4	0	0	0	0	2	1	19	**300**	3	0	0	f = eating
0	5	11	1	29	40	**18**	0	0	0	4	4	0	26	57	**16**	0	0	g = drinking
2	0	0	2	0	0	0	**11**	0	2	0	0	0	0	0	0	**14**	0	h = walking
0	2	2	0	2	0	0	0	**271**	0	0	0	0	1	0	0	0	**276**	i = sleeping

Fig. 14. Activity recognition evaluation revealing the positive impact of the capacitive proximity sensor. Here, we are comparing SVM classification presented as confusion matrices for an exemplary user, without the proximity data on the left, and with the proximity data on the right. Note that the reject class (background data not annotated as an activity) is not included in the confusion matrix.

Regarding the activity "make sandwich", the capacitive proximity could reduce the number of confusions with the "eating" class significantly.

Due to the high similarity of eating and drinking (cf. Figure 10 and 11), the number of confusions between those two classes increases when considering the proximity data. However, for the "drinking" class the new modality limits the confusions to related activities such as "eating" and "make sandwich" only, while lowering the number of false recognitions for the other activities.

5 Conclusion and Outlook

Our paper presented a wrist-worn activity data logger prototype, which consists of an accelerometer in combination with a capacitive proximity sensor integrated into the wristband. Our experiments with seven participants and nine daily activities show that this additional input modality can significantly boost the activity recognition performance. Regarding the classification performance, we obtained an improvement in the average F-measure of 6.3%, from 67.2 to 73.5%. Specifically, the activity classes "walking", "make sandwich" and "open door" could benefit a lot from proximity-related sensor data. For such classes, the classification performance could be boosted by 12.2%, 9.0% and 10.8% respectively.

With this proof of concept we show that the proximity information can provide an information gain regarding the evaluated activities. In future work we aim at evaluating other relevant feature types, such as frequency domain features, as well as feature sets to extract the most discriminative ones. Additionally, using other classifiers (such as HMMs) might also improve activity recognition performance.

The classification results could also be improved by using more than one sensing electrode in the wristband. For example, the wristband could integrate up to four electrodes that are placed on each side of the arm. However, this will lead to smaller electrode surfaces thus resulting in a decreased sensing distance. The sensor's power consumption can be decreased by shorter measurement windows and the choice of more energy efficient hardware components as well as software implementation. Measuring pulse width lengths instead of counting the number of pulses of the sensor's signal may reduce the required time needed for a measurement, thus increasing the time the microcontroller is able to sleep.

Capacitive proximity sensors represent a suitable new input modality for future activity recognition systems. The low power consumption as well as the unobtrusive integration of a sensor into the wristband meets an essential requirement of wearable applications. Especially in AAL environments, these systems can help monitoring the course of chronic diseases by recognizing activities of daily life. This may improve the quality of life of persons affected and their caregivers.

Acknowledgements. We would like to thank our evaluation participants from the Technische Universität Darmstadt and Fraunhofer IGD.

References

1. HedgeHog activity logger, http://www.ess.tu-darmstadt.de/hedgehog (accessed June 17, 2012)
2. Altun, K., Barshan, B.: Human Activity Recognition Using Inertial/Magnetic Sensor Units. In: Salah, A., Gevers, T., Sebe, N., Vinciarelli, A. (eds.) HBU 2010. LNCS, vol. 6219, pp. 38–51. Springer, Heidelberg (2010)
3. Amft, O., Junker, H., Tröster, G.: Detection of eating and drinking arm gestures using inertial body-worn sensors. In: Proceedings of the 9th IEEE International Symposium on Wearable Computers (ISWC 2005), pp. 160–163. IEEE (2005)
4. Bao, L., Intille, S.: Activity Recognition from User-Annotated Acceleration Data. In: Ferscha, A., Mattern, F. (eds.) PERVASIVE 2004. LNCS, vol. 3001, pp. 1–17. Springer, Heidelberg (2004)
5. Berlin, E., Liu, J., van Laerhoven, K., Schiele, B.: Coming to grips with the objects we grasp: Detecting interactions with efficient wrist-worn sensors. In: Proceedings of the Fourth International Conference on Tangible, Embedded, and Embodied Interaction, TEI 2010, pp. 57–64. ACM, New York (2010)

6. Borazio, M., Van Laerhoven, K.: Combining wearable and environmental sensing into an unobtrusive tool for long-term sleep studies. In: Proceedings of the 2nd ACM SIGHIT Symposium on International Health Informatics, IHI 2012, pp. 71–80. ACM Press (2012)
7. Brezmes, T., Gorricho, J.L., Cotrina, J.: Activity Recognition from Accelerometer Data on a Mobile Phone. In: Omatu, S., Rocha, M.P., Bravo, J., Fernández, F., Corchado, E., Bustillo, A., Corchado, J.M. (eds.) IWANN 2009, Part I. LNCS, vol. 5518, pp. 796–799. Springer, Heidelberg (2009)
8. Cheng, J., Amft, O., Lukowicz, P.: Active Capacitive Sensing: Exploring a New Wearable Sensing Modality for Activity Recognition, pp. 319–336 (2010)
9. Fishkin, K., Philipose, M., Rea, A.: Hands-on rfid: wireless wearables for detecting use of objects. In: Proceedings of the 9th IEEE International Symposium on Wearable Computers (ISWC 2005), pp. 38–41 (October 2005)
10. Grosse-Puppendahl, T., Braun, A.: Honeyfish - a high resolution gesture recognition system based on capacitive proximity sensing. In: Embedded World Conference 2012, pp. 1–10 (2012)
11. Große-Puppendahl, T.A., Marinc, A., Braun, A.: Classification of User Postures with Capacitive Proximity Sensors in AAL-Environments. In: Keyson, D.V., Maher, M.L., Streitz, N., Cheok, A., Augusto, J.C., Wichert, R., Englebienne, G., Aghajan, H., Kröse, B.J.A. (eds.) AmI 2011. LNCS, vol. 7040, pp. 314–323. Springer, Heidelberg (2011)
12. Holleczek, T., Schoch, J., Arnrich, B., Troandster, G.: Recognizing turns and other snowboarding activities with a gyroscope. In: Proceedings of the 14th IEEE International Symposium on Wearable Computers (ISWC 2010), pp. 1–8 (October 2010)
13. Krause, A., Siewiorek, D., Smailagic, A., Farringdon, J.: Unsupervised, dynamic identification of physiological and activity context in wearable computing. In: Proceedings of the 7th IEEE International Symposium on Wearable Computers (ISWC 2003), pp. 88–97. IEEE Computer Society, Washington, DC (2003)
14. Mühlsteff, J., Such, O., Schmidt, R., Perkuhn, M., Reiter, H., Lauter, J., Thijs, J., Müsch, G., Harris, M.: Wearable approach for continuous ECG and Activity Patient-Monitoring. In: Complexity, pp. 2184–2187 (2004)
15. Patterson, D., Fox, D., Kautz, H., Philipose, M.: Fine-grained activity recognition by aggregating abstract object usage. In: Proceedings of the 9th IEEE International Symposium on Wearable Computers (ISWC 2005), pp. 44–51 (2005)
16. Philipose, M., Fishkin, K.P., Perkowitz, M., Patterson, D.J., Fox, D., Kautz, H., Hahnel, D.: Inferring activities from interactions with objects. IEEE Pervasive Computing 3(4), 50–57 (2004)
17. Ravi, N., Dandekar, N., Mysore, P., Littman, M.: Activity recognition from accelerometer data. In: Proceedings of the National Conference on Artificial Intelligence, vol. 20, p. 1541. AAAI Press, MIT Press, Menlo Park, CA, Cambridge, MA (2005)
18. Smith, J.R.: Field mice: Extracting hand geometry from electric field measurements. IBM Syst. J. 35(3-4), 587–608 (1996)
19. Smith, J.R., Gershenfeld, N., Benton, S.A.: Electric Field Imaging. Technology (1999)
20. Srinivasan, R., Chen, C., Cook, D.: Activity recognition using actigraph sensor. In: Proceedings of the Fourth Int. Workshop on Knowledge Discovery form Sensor Data (ACM SensorKDD 2010), Washington, DC, pp. 25–28 (July 2010)
21. Stiefmeier, T., Roggen, D., Ogris, G., Lukowicz, P., Tröster, G.: Wearable activity tracking in car manu- facturing. IEEE Pervasive Computing 7, 42–50 (2008)

22. Stikic, M., Huynh, T., Van Laerhoven, K., Schiele, B.: ADL Recognition Based on the Combination of RFID and Accelerometer Sensing. In: Proceedings of the 2nd International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health 2008), pp. 258–263. IEEE Xplore, Tampere (2008)
23. Ward, J., Lukowicz, P., Troster, G., Starner, T.: Activity recognition of assembly tasks using body-worn microphones and accelerometers. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(10), 1553–1567 (2006)
24. Wimmer, R., Kranz, M., Boring, S., Schmidt, A.: A Capacitive Sensing Toolkit for Pervasive Activity Detection and Recognition. In: Fifth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom 2007), pp. 171–180 (2007)
25. Wimmer, R., Kranz, M., Boring, S., Schmidt, A.: CapTable and CapShelf - Unobtrusive Activity Recognition Using Networked Capacitive Sensors. Group Networked (2007)
26. Wyss, T., Mader, U.: Recognition of Military-Specific Physical Activities With Body-Fixed Sensors. Military Medicine 175(11), 858–864 (2010)

Adaptive User Interfaces for Smart Environments with the Support of Model-Based Languages

Sara Bongartz[1], Yucheng Jin[1], Fabio Paternò[2], Joerg Rett[1], Carmen Santoro[2], and Lucio Davide Spano[2]

[1] SAP AG, Darmstadt, Germany
{sara.bongartz,yucheng.jin,joerg.rett}@sap.com
[2] CNR-ISTI, Pisa, Italy
{fabio.paterno,carmen.santoro,lucio.davide.spano}@isti.cnr.it

Abstract. This article presents a solution for supporting adaptive user interfaces in work environments. Its architecture is built upon the concept of model-based UI design extended by context aware and adaptive features. Model-based languages provide the software development process with useful support for, building design prototypes and actual implementations for devices with various interaction resources. The proposed architecture is able to adapt to selected aspects of the context during run-time by communicating with a context server and applying the specified adaptation rules. In order to show the possibilities of the proposed solution, we report on its application in the development of an adaptive user interface prototype to be used in a warehouse picking system.

Keywords: Adaptive service front-ends, Context-aware user interfaces, model-based user interface languages, Warehouse picking system.

1 Introduction

Work environments are continuously becoming richer in sensors and devices and it is important that users are efficient and perform well in such contexts, without having to spend too long in understanding how to interact with the system. Adaptive interfaces can be useful for this, because they can provide the required information in the most suitable modality by taking into account the current context of use. For this purpose, it is useful to consider various contextual aspects, including the user-related aspects (tasks to accomplish, personal preferences and knowledge, etc.), the aspects related to the technology (available interaction resources, connectivity support, etc.) and the environmental aspects (level of noise, light, etc.).

In order to show the possible application of our methods, tools and languages; we consider a specific application domain: warehouse picking, which is a part of a logistics process often found in retail and manufacturing industries. Warehouses store the goods and products incoming from suppliers until they are collected and shipped to the stores or customers. The process of picking items from a shelf, collecting them in some sort of container and bringing them to certain locations is usually conducted by

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 33–48, 2012.

one or more persons. This leaves a maximum degree of freedom for the enterprise to change or rearrange the environment when needed. The costs of maintenance and the down-times are lower compared to a fully automated system based on e.g. conveyor belts. In addition, only people can react and adapt to unforeseen situations.

Since the beginning of warehouse picking, people have thought on how the pickers can be supported in their task. Technological solutions have been provided, like Voice-directed warehousing (VDW). VDW refers to the use of the voice direction and speech recognition software in warehouses and distribution centres. In a voice directed warehouse, workers wear a headset connected to a small wearable computer, which tells the worker where to go and what to do using verbal prompts. Workers confirm their tasks by speaking pre-defined commands and reading confirmation codes printed on locations or products throughout the warehouse.

A significant drawback of VDW is that the information is volatile. Once information, such as the amount of items to be picked has been given, the system might not be able to repeat it. If the picker forgets such information, its retrieval might become laborious. Thus, visual UIs have gained importance in the task of supporting the picker. In previous work [1] the authors report on a 12 participant within-subjects experiment, demonstrating the advantages of a head-mounted display based picking chart over traditional text-based pick lists, paper-based graphical pick charts, and mobile pick-by-voice systems. No multimodal solution was investigated in that study. A different study [3] aimed at comparing various design solutions for head-mounted displays with different levels of information. The order was not only presented on a graphical UI (GUI) but also supported by some kind of sensors installed in the environment. This system exhibited a kind of Ambient Intelligence by detecting the picker's action of reaching into a shelf and comparing the respective box and its containing item with the order from the backend system. Such systems based on HMDs can be extended to support Augmented Reality, as shown in [2]. One limitation of these contributions is that they provide solutions that are implemented with ad hoc techniques and thus cannot be easily generalised to other similar applications. The use of model-based languages provides designers and developers with a general vocabulary for describing their solutions that can be refined in concrete terms for various interaction platforms, and then used to obtain implementations in a variety of languages, even for various combinations of interaction modalities. Thus, this model-based approach can be interesting in this application domain in which the combination of vocal and visual interaction has not been investigated so far. The level of multimodality, i.e. emphasising one of the two modalities, should depend on the context of use. For instance, in a noisy environment the interaction should rely mainly on the GUI. In addition, also the user's preferences and the capabilities of the platform should be considered. If the environment is fully intelligent, the picker can receive support for navigating through the location and picking the right items. Ideally, the system should be able to receive this context information and adapt the UI accordingly.

In this paper we present a solution consisting of an adaptive, context-sensitive UI which is based on an architecture for context-sensitive service front-ends. The solution is based on the use of model-based languages for interactive application descriptions in order to facilitate the possibility of deriving versions adapted to various contexts of use, particularly in terms of interaction device resources. Such languages are currently under consideration for standardisation in W3C, because of their useful support in creating versions of interactive applications that adapt to different interactive devices. However, they have been mainly used in academic environments, with very few cases of use in real world applications.

In the paper, after discussing related work, we provide some background information related to the approach we have developed and how it has been extended in order to better support adaptive multimodal user interfaces for smart environments. Next, we describe the example application considered and how it varies depending on the context of use; followed by a description of the architecture supporting the adaptation of the multimodal service front ends. Then, we introduce some adaptation rules for the application considered, which are formalised in terms of *event*, *condition*, *action* (ECA) rules. We then show the corresponding prototype and report on an early user test focusing on the adaptation rules considered. Lastly, we draw some conclusions and provide indications for future work.

2 Related Work

Mobile applications often require highly-focused visual attention, which poses problems when it is inconvenient or distracting to continuously look at a screen (e.g., while walking). Aural interfaces support more eyes-free experiences, as users can primarily listen to the content and occasionally look at the device. However, designing aural information architectures remains a challenge. For example, recent studies [15] have highlighted that backward navigation is inefficient in the aural setting, as it forces users to listen to each previous page to retrieve the desired content. Thus, they have introduced topic and list-based back navigation strategies in order to enhance aural browsing and improve the navigation experience, reducing the perceived cognitive load. This shows the potential of multimodal interfaces in ubiquitous scenarios, but also the need for some specific design solutions, which depend on the interaction modalities exploited.

The problem of designing user interfaces that are able to be rendered on multiple types of platforms, including multimodal and vocal ones, has been addressed in some previous work, but still needs more general, better engineered solutions. Damask [10] includes the concept of layers to support the development of cross-device (desktop, smartphone, voice) UIs. Thus, the designers can specify UI elements that should belong to all the user interface versions and elements that should be used only with one device type. However, this approach does not consider the support of multimodal user interfaces. XFormsMM [8] is an attempt to extend XForms in order to derive both graphical and vocal interfaces. The idea is to specify the abstract controls with XForms elements and then use aural and visual CSS respectively for vocal and graphical rendering. However, aural CSS have limited possibilities in terms of vocal interaction and the solution proposed requires a specific ad-hoc environment. For this purpose we propose a more general solution which is able to derive implementations in various languages.

Obrenovic et al. [11] have investigated the use of conceptual models expressed in UML, in order to derive graphical, form-based interfaces for desktop, mobile or vocal devices. However, since UML is a software engineering standard, aimed at supporting the specification of the internal software application functionalities, it seems unsuitable to capture the specific characteristics of user interfaces. A different approach to multimodal user interface development has been proposed in [9], which aims to provide a workbench for prototyping UIs using off-the-shelf heterogeneous components. In that approach, model-based descriptions are not used and it is necessary to have an available set of previously defined components, which are able to communicate through low-level interfaces; thus making it possible for a graphical editor to easily compose them.

Sottet and others [13] have presented a set of general principles relevant for supporting model-based adaptation while in our case we present a software architecture supported by engineered tools that can be applied in real world applications.

Octavia et al. [12] have considered the use of a model-based approach to facilitate adaptation in virtual environments, also using the event-condition-action paradigm, we provide a more general architecture for this purpose able to support adaptation involving various interaction modalities.

To summarise, we can say that the few research proposals that have also considered multimodal interaction have not been able to obtain a suitable engineered solution in terms of logical descriptions and corresponding software architectures and provided limited support in terms of the generation of the corresponding user interface implementations. For example, in [14] the transformations were specified using attributed graph grammars, whose semantics is formally defined but have considerable performance limitations.

3 MARIA and Its Support for Multimodal Interaction

We exploit the MARIA model-based framework [5] for obtaining adaptation able to better support various interaction modalities. The framework provides a language for the abstract description (the so-called “Abstract User Interface” level, in which the UI is defined in a platform –independent manner) as well as multiple platform-dependent languages (which are at the level of the so-called “Concrete User Interface”), which refine the abstract language depending on the interaction resources at hand. Examples of platforms are the graphical desktop, the graphical mobile, the vocal platform, etc.

At the abstract level, a user interface is composed of a number of presentations, it has an associated data model, and can access a number of external functions. Each presentation is composed of a number of interactors (basic interaction elements) and a set of interactor compositions. There are four types of interactor composition operators: grouping, relation, composite description and repeater. These composition operators support the structuring of the elements inside a presentation. A *grouping* is a type of interactor composition used when a logic composition of interactors is needed. Therefore, grouping basically represents a generic group of interactor elements. A *relation* is an interactor composition which expresses a relation between an interactor (or an interactor composition) and other interactors (or interactor compositions). A *composite_description* represents a group aimed to present contents through a mixture of only_output elements (namely: description/object/feedback/alarm) with navigator elements, while a repeater is used to repeat the content according to data retrieved from a generic data source. Each presentation is also associated with a dialogue model, which describes how the events generated by the interactors can be handled. With respect to previous languages in this area a number of substantial features have been added, such as a data model, a dialogue model, the possibility to specify typical Web 2.0 interactions, and the support to access Web services.

In its current version, MARIA consists of a set of languages: one for abstract user interface descriptions and a set of concrete refinements of such language for various target platforms (Vocal, Desktop, Smartphone with touch, Mobile, Multimodal desktop, Multimodal mobile). Moreover, user interface generators for various implementation languages are available starting with such concrete languages. The multimodal concrete language provides the possibility to indicate how to distribute the user interface elements across modalities through a simple and intuitive vocabulary that can be applied at various granularity levels. While previously this multimodal concrete language was associated with a generator of X+V implementation [6], in this work we consider a new generator able to create HTML 5 applications with multimodal features obtained using the Google support. Such support allows sending the user's utterance to a remote vocal recogniser in order to determine the corresponding input value, together with a Google Chrome extension that provides Text-to-Speech (TTS) access directly from JavaScript code. An HTML5 page developed with this extension consists of two parts: the graphical and the vocal one. These parts are linked by a rule applied to the id attributes of the HTML elements: the id of the vocal element is obtained adding "_vocal" at the end of the id of the graphical object. This allows us to obtain different implementations for the same interface element in the two modalities. Once a graphical element gets the UI focus, the extension retrieves its corresponding vocal element and invokes the TTS engine. The synthesis properties (e.g. speech, break, emphasis) are represented by a set of pre-defined CSS classes. If the graphical element is an input, the extension also starts the possibility of recording the voice and, when it detects a long silence after the vocal input, it invokes the ASR passing it the user's utterance. The result of the ASR is then used for filling the corresponding graphical element, or simulating a link or button click.

4 Example Application

In this work we want to exploit the model-based approach in supporting adaptation in real world applications relevant in the ambient intelligence domain. Thus, we provide some further detail on the application considered and how interaction can vary in it depending on the ambient intelligence available. The example application is situated at a distribution centre of a supermarket chain in the domain of retail industries. The task of the so-called pickers is to collect items from the shelves in the warehouse and place them into containers. One collection belongs to an order issued by a specific store of the supermarket chain.

The process starts when the picker signs-up for an order. Fig. 1 represents the scenario in Business Process Modelling Notation [4]. The simplified model for the warehouse picking process consists of Events of types: message (circle with envelope), Gateways of the type data-based exclusive (diamond with X), Tasks (rounded rectangles) processed by whether the system (scroll) or the picker (people), Data objects (document) and Text annotations (square bracket).

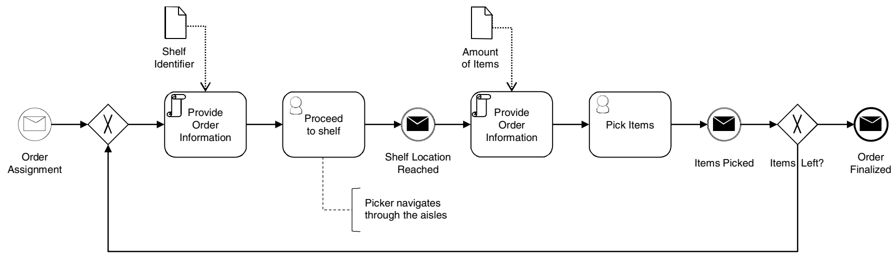


Fig. 1. Scenario for a warehouse picking process expressed in Business Process Modelling Notation [4]

After the Order Assignment, the system will provide the shelf identifier to the picker so they know where to find the items. The picker will then proceed to the shelf by navigating through the aisles. Depending on the initial location and the destination, s/he may pass through several halls. The system needs to be informed that the picker has reached the shelf location (i.e. a message needs to be sent). The system will then provide the amount of items to be picked and the picker can begin picking them. After this, the system needs to be informed about the completion of the picking. As long as there are still some items left, the process will loop back to the first Task. The process ends when the order is completed, i.e. when all items have been picked.

Now, we are going to consider the example application in two different contexts of use, with and without ambient intelligence. In the first situation (A) we assume that the warehouse is equipped with a kind of sensor which could support Ambient Intelligence. Concerning the example scenario described in this article, we assume that the system supports:

- the input of vocal prompts by the picker,
- tracking the position of the picker, e.g. through indoor navigation,
- identification of the actions of the picker, e.g. reaching into a shelf or,
- tracking the location of the items, e.g. by means of RFID tags.

In situation (A) the environment will trigger the event. In our example scenario the event “Shelf Location Reached” can be triggered by the module that tracks the position of the picker. The event “Item picked” can be triggered by a module that tracks the location of an item. In the second situation (B) we assume that the system can only support the input of vocal prompts by the picker. In this case the picker needs to trigger the event, i.e. issue messages to the system. In our example scenario, the event “Shelf Location Reached” can be triggered by the picker through a vocal interface. It is common that the picker reads a number from a sign which is attached to the shelf. The event “Item picked” can be triggered in a similar fashion. The picker then repeats the amount of items that have been picked.

It is clear, that the situation at a specific warehouse might be somewhere in between situation A and B. For example, one warehouse might have a tracking system for the picker but not for the items and vice versa. Thus, it would be desirable to have an interactive application that could adapt to the specific situation of the warehouse environment.

5 Architecture

Our architecture shows how we can provide support for adaptation through the use of model-based descriptions of interactive applications. At design time the various initial versions of the applications are developed in terms of the three concrete descriptions (vocal, graphical, and multimodal) specified according to the MARIA language. In addition, the relevant adaptation rules are specified in terms of events, conditions, and actions according to a language for adaptation rules that will be described later on. Such adaptation rules are triggered by contextual events, which can depend on various aspects (user preferences, environmental changes, application-related events, etc.). The impact of the adaptation rules can have various granularities: complete change of user interface (e.g. from vocal to graphical if the environment becomes noisy), change of some user interface parts (e.g. change from map view to order view), and even change of attributes of specific user interface elements (e.g. change of font size).

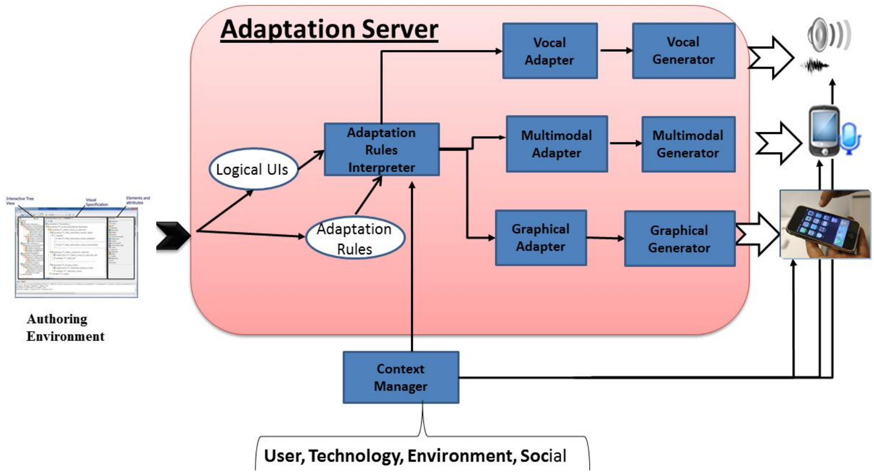


Fig. 2. The software architecture supporting the adaptive application

Thus, at design-time it is possible to specify the relevant logical descriptions of the interactive application versions and the associated adaptation rules.

At run-time we have an adaptation server that is able to communicate with the context manager server in order to receive information on subscribed events and the interactive devices available in order to update the application according to the adaptation rules. The context server is also able to communicate with the applications, which can manage directly some contextual events according to their specifications.

More specifically, the adaptation rules interpreter has access to the list of the relevant adaptation rules. Changes in the context communicated by the context manager can trigger some of them. The adaptation rule interpreter considers the action part of the rules and depending on its content it can either trigger the activation of the application in a different modality (which means to trigger a new application generator for the most relevant modality in the new context of use) or indicate some change to perform to the adapter associated with the current modality (in this case the adapter will then request the corresponding generator to update the interactive application accordingly). The languages to specify the adaptation rules and the interactive applications are distinct, so that it is possible to modify one without having to change the other one, but with clear relations defined among them so that the actions of the adaptation rules can be specified in terms of required modifications to the model-based descriptions of the interactive applications. Thus, the logical description of the interactive application can dynamically change from the version that was initially provided by using the authoring tool.

6 Adaptation Rules

We developed a XML-based high-level description language intended to declaratively express advanced adaptation logic defining the transformations affecting the interactive application when some specific situations occur both in the context (e.g. an entity of the context changes its state), and in the interactive application (e.g. an UI event is triggered). In particular, the three parts of the language are: *event*, *condition*, *action* (ECA). The *event* part of the rule should describe the event whose occurrence triggers the evaluation of the rule. This part could specify elementary events occurring in the interactive application, or a composition of events. The *condition* part is represented by a Boolean condition that has to be satisfied in order to execute the associated rule action(s). The condition part is optional. In the *action* part there might be 1 to N simple actions occurring in the interactive application or even 1 to N other adaptation rules. In practise, the action part often contains indications on how the concrete description of the interactive application should change in order to perform the requested adaptation. Event Condition Action is an approach that was originally introduced for the structure of active rules in event driven architecture and active database systems, and has already been used for supporting adaptive user interfaces (see for example [12]). In our case, we have structured it in such a way to easily connect it to the events generated by the context manager and the interactive application specification.

Below there is a list of example adaptation rules supported by the prototype. For each rule we provide a title with a brief explanation/rationale and the three key parts of its specification (event, condition, action).

- *Fragile object* - The rationale of this rule is that when the worker is about to pick a fragile object, the multimodal UI should switch to only-vocal modality in order not to distract the user while picking the item.
    - *Event*: the right shelf has been reached

    - *Condition*: the worker has to pick a fragile item and the current modality is not only-vocal
    - *Action:* Switch from multimodal to only-vocal modality,
- *Picking timeout* - The user has just reached the destination shelf of the item but there is no confirmation of the actual item picking. The application then assumes that the worker is distracted/confused and/or not able to recognize the item to pick, then it provides again info on the item, both graphically and vocally.
    - *Event*: the user has reached the destination shelf
    - *Condition*: there has not been confirmation of the item picking and the user interface is multimodal
    - *Action*: the application visualizes an image representing the item to pick, and simultaneously repeats the item name vocally.
- *Order visualization for experienced workers* - If the user has good knowledge of the warehouse shelf organisation, there is no need to show associated path information: the application adapts accordingly.
    - *Event*: beginning of a session with the HMD
    - *Condition*: the user is a warehouse expert
    - *Action*: the application hides the information about how to reach the different shelves.
- *Traffic Jam* - There are multiple workers who are expected to approach the same path at the same time: the application adapts in order to minimise the risk of workers to wait for other people before picking the items.
    - *Event*: order completed
    - *Condition*: multiple pickers are expected to approach the same path at the same time and the path optimization preference is selected.
    - *Action*: the application shows the blocked path suggesting a different route.
- *Noisy environment* – The environment gets noisy, then the multimodal application switch to ‘only-graphical’ modality.
    - *Event*: the environment gets noisy
    - *Condition*: the application is using both the graphical and vocal modality for interacting with the user.
    - *Action*: the application switches to the only-graphical modality

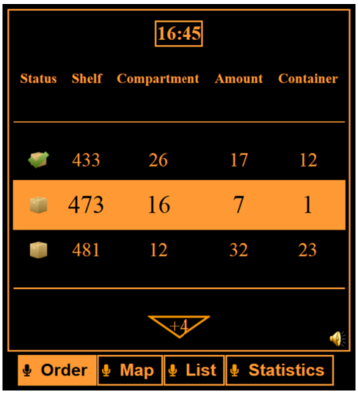
We show how it is possible to express such adaptation rules through our high-level description language with one example. We consider (Fig. 3) the rule for the order visualization for experienced workers. When the interaction starts (the presentation raises the *onRender* event), if the current user is an expert one (represented as an attribute in the ‘user’ part of the context model), the application hides the path to reach the shelf, which is represented by an interactor with id *path_to_shelf*, setting its *hidden* attribute to *true*. A symmetrical rule manages the case of inexperienced workers.

```xml
<rule>
    <event>
        <simple_event event_name="onRender"
            xPath="/interface/"
            externalModelId="uiModel"/>
    </event>
    <condition operator="eq">
        <entityReference xPath="/context/users/user/@experience"
            externalModelId="ctxModel"/>
        <constant value="high" type="string"/>
    </condition>
    <action>
        <update>
            <entityReference
                xPath="/interface[@current_presentation]/interactor[@id = 'path_to_shelf']/@hidden"/>
            <value>
                <constant value="true" type="boolean"/>
            </value>
        </update>
    </action>
</rule>
```

Fig. 3. Formalization of the Order visualization for experienced workers rules

7 Adaptive Application

We start with a description of the graphical version of the interactive application considered. The GUI consists of four views (Order, Map, Task and Statistics), for the sake of brevity only the Order view and the Map view are discussed. The Order view, shown in Fig. 4, mainly contains information on the previous (i.e. shelf 433), the current (i.e. shelf 473) and the next (i.e. shelf 481) items to be picked. This sequence of picks is represented in three rows starting with the previous pick and having the current pick highlighted (i.e. inverted) and magnified.

a) 

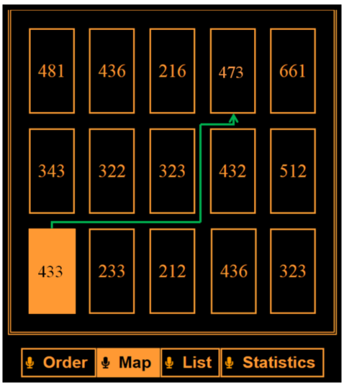
b) 

Fig. 4. Design of the graphical user interface (GUI). a) Order view b) Map view

The columns reflect the types of information available for the pick (status, shelf, compartment, amount and container) while only the status of the pick (e.g. open), the shelf identifier (e.g. 473) and the amount of items to be picked (e.g. 7) are relevant here. The active view is reflected as a highlighted tab in the bottom area. The main information in the Map view is a simplified representation of the location of the shelves (in Bird eyes view) showing the current location of the picker (i.e. the previous shelf), the destination shelf (i.e. 473) and a suggested route (line with arrow). In general it is possible to navigate between the screens by using voice commands, but this functionality is not used in the actual setting.

Based on a list of requirements for the prototype a Head-Mounted Display (HMD) and a wearable computer are used to access the application. The UI generated from the MARIA specification is implemented in HTML5, JavaScript and AJAX. The navigation route in the Map view is drawn using the canvas label of HTML5. Speech recognition is realized using the speech input label of HTML5 and calling a respective API. The architecture of the application implementation is shown in Fig. 5.

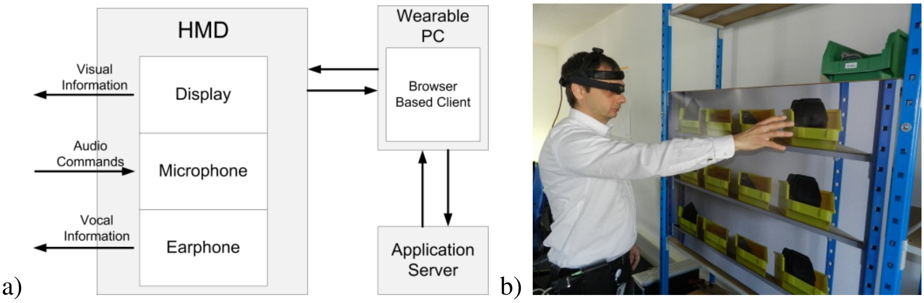


Fig. 5. a) Architecture of the prototype. b) Picking from a shelf using a Head-Mounted Display.

The adaptation server sends the updated data to the wearable computer after a change in the context has triggered the execution of an adaptation rule. The display is used for the visual output, the earphone for the vocal output and the microphone for the vocal input of the user. Some changes might be triggered by the smart environment (e.g. tracking of the picker's position or the item's location). Table 1 lists the five variations of the context and its consequences for the interaction modalities with respect to the basic interaction flow. The variations are based on the *Condition* and the consequences derive from the *Action* stated in the respective adaptation rule.

Table 1. Variations of the context and its consequences for the interaction modalities

Context variation	Interaction consequence
The items to be picked are fragile	After vocally confirming the arrival at the destination by the picker, the visual output will be switched off, only vocal remains.
The route is blocked by other pickers.	The Map view marks the blocked path and suggests an alternative route.
The picker is experienced	The Map view is omitted.
The environment is noisy	The vocal input and output is switched off, only visual output remains
The picking is not performed due to some confusion or distraction	An image of the item to be picked is shown, the vocal output is repeated.

Finally we present the basic interaction sequence (i.e. the basic interaction flow) with an example for an adaption in Fig. 6: the picker is presented with three screens and two vocal outputs (upper balloons) and needs to perform two vocal inputs (lower balloons). Assuming that a picker, who is experienced, i.e. has been working for a long time in the warehouse environment and thus should know by heart the location of the shelves, and the Map view can be omitted. We assume that an indicator of the experience level is stored within the profile of the picker and is added as context information at run-time during the log-in procedure.

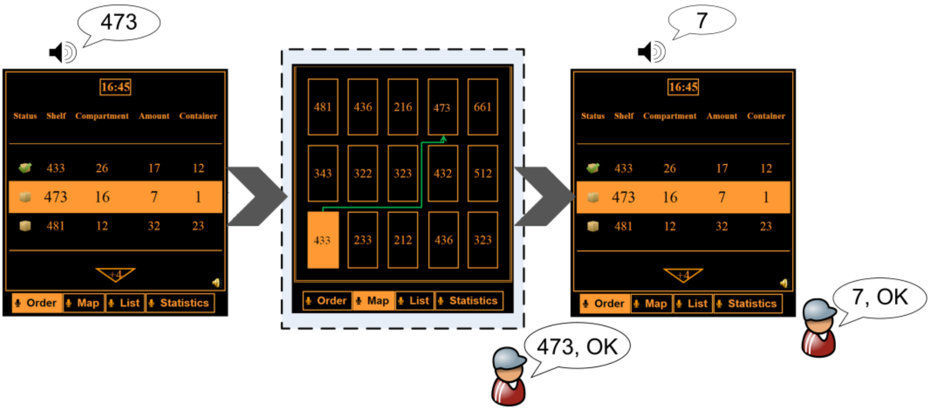


Fig. 6. Basic interaction flow with adaptation: the execution of the rule for an experienced picker omits the appearance of the Map view (dotted line)

8 User Feedback

We have conducted a first user study in order to evaluate the five adaptation rules from the end-users point-of view. The study aimed at evaluating the applicability and usefulness of the adaptation rules, specified as described in Section 6 through an XML-based ECA style, by assessing the quality of the adaptation rules as subjectively perceived by the participants. The general concept "quality" was operationalized by several more specific constructs, e.g. usefulness, comprehensibility or simplicity, which were assessed by a questionnaire.

To address such issues, the five adaptation rules were the independent variables. We had a within-subject design, meaning that every participant was confronted with every adaptation rule. The dependent variables were the subjectively perceived quality of the adaptation rule as assessed in a 9-item questionnaire. The questions originated from a list of non-functional requirements for the prototype identified in user studies in the beginning of the project and aimed at assessing the following aspects: the user's awareness for the adaptation rule, its appropriateness and comprehensibility, its effectiveness with respect to performance and usability, its error-prevention, continuity, intuitiveness, and general likeability.

Participants were company staff or students of the local university. A total of 10 participants took part in the study, 9 were male and 1 was female. The average age of participants was 24 years (SD = 1.82). The technical set-up consisted of an HMD with earphone worn by the participants. The device presented the GUI and the vocal output as shown in section 7. The sequence of the interaction was controlled by the moderator simulating the change of context and the execution of the adaptation rule.

Participants were first introduced into the scenario and the interface, i.e. getting familiar with the hypothetical situation in the warehouse and learning how to interact with the interface. Participants were asked to play through a "basic interaction flow" which started with the systems request to pick items from a certain shelf, required the user to hypothetically walk to that shelf and ended with the user's confirmation that he picked a certain amount of items. Participants were asked to comment their hypothetical actions, e.g. by saying "I walk to the shelf 473 now" or "I pick 7 items from the shelf". After ensuring that the participants understood the basic interaction flow of the interface, the study started by introducing the first alternative flow. All alternative flows (flows containing adaptation rules) were applied to the same scenario as practiced in the basic flow. Prior to playing through the alternative flows, participants were informed about the condition of the adaptation rule (e.g. "imagine you are now in a noisy environment"), but not about the actual rule (i.e. the action of the rule). All five rules were played through and the sequence of the adaptation rules was permutated to avoid order effects. After each rule, the 9-item questionnaire was filled out.

Since most of the scales of the questionnaire were not normal-distributed, we applied non-parametric tests for the data analysis. We calculated the Friedman test for every single questionnaire scale and the aggregated overall rating from all 9 scales (Bonferroni-corrected) to assess differences between the five adaptation rules. In case of significance, we calculated a post-hoc Wilcoxon signed-rank test for each pair of adaptation rule (Bonferroni-corrected as well).

The Friedman test revealed significant differences for the aggregated overall rating over all 9 scales ($\chi^2(4) = 18.74$, $p = .001$) and for 4 of the subscales: Appropriateness ($\chi^2(4) = 19.26$, $p = .001$), Performance ($Z = -2.69$, $p=.007$), Error-Prevention ($\chi^2(4) = 22.73$, $p = .000$), Intuitiveness ($\chi^2(4) = 22.31$, $p = .000$) and General Likeability ($\chi^2(4) = 18.92$, $p = .001$). Only these significantly different scales are regarded in detail here. Post-hoc tests revealed a significant difference in the rating between the rules Fragile Objects and Traffic Jam ($Z = -2.60$, $p = .009$) and Experienced Worker and Traffic Jam ($Z = -2.70$, $p=.007$). The significant differences in the subscale Appropriateness are between the rules Fragile Objects and Traffic Jam ($Z = -2.62$, $p = .009$) and Fragile Objects and Pick Timeout ($Z = -2.69$, $p = .007$). For the subscale Error prevention, the significant differences can be found between the rules Fragile Object and Pick Timeout ($Z = -2.71$, $p = .007$), Traffic Jam and Experienced Worker ($Z = -2.81$, $p = .005$) and Pick Timeout and Experienced Worker ($Z = -2.68$, $p = .007$). Intuitiveness shows significantly different values for the rules Fragile Objects and Traffic Jam ($Z = -2.69$, $p = .007$). Finally, although the Friedman test revealed significant differences between the rules for the scales: general Likeability and Performance; direct pairwise comparison failed reaching significance due to Bonferroni correction.

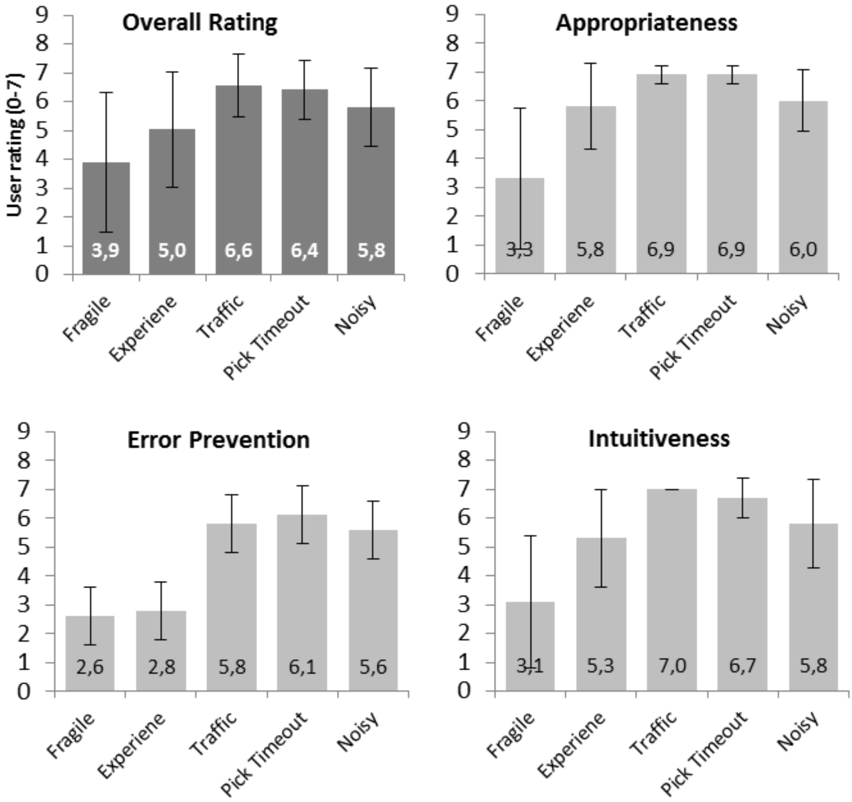


Fig. 7. Overall rating and the subscales Appropriateness, Error-Prevention and Intuitiveness

The big picture of the results (see Fig. 7) shows a clear trend: all quality aspects of the Fragile Object rule are consistently rated the worst, and the Traffic Jam and Pick Timeout rule are consistently rated best. This pattern can be observed for all quality scales, indicating a clear and coherent preference pattern. Traffic Jam and Pick Timeout are consistently and undoubtedly preferred by the users (with very good overall ratings of 6.6 and 6.4 on a scale from 0-7). Alongside the good rating of these two rules, the standard deviation is very small, indicating a very high agreement between the participants. However, the Fragile Object rule, as the worst rated one, shows the highest variance in the ratings between the subjects. This indicates that there is no strong agreement between the subjects, yet still most of the subjects gave comparably low ratings for that rule. A possible explanation for this finding can be drawn from the subject's comments. While all subjects gave a positive opinion about the idea to support the process of picking a fragile object, most of the subjects noted that the actual realisation of that rule was poor. Turning off the display was irritating and non-intuitive to the subjects. The abrupt darkness in the HMD was perceived as a break-down of the system and therefore caused confusion. Rather, subjects had wished to receive a short warning message before turning off the display.

We found similarities between those rules that were ranked well and those that were ranked poor. The group of poorly ranked rules was omitting information like the visual output and the Map view with regard to the Basic Interaction Flow. The Fragile rule takes a prominent position as a very strong modality, the visual channel, is shut off. Those rules that were ranked well however delivered additional information like the blocked path or the image of the item. This noticeable difference between the adaptation rules is presumably the reason for the striking difference in the preference ratings. It is worth investigating the role of adding vs. removing information as well as amount of information in the course of interface adaptation.

9 Conclusions

Work applications often need intelligent environments able to provide adaptive user interfaces that change the interaction modalities taking into account contextual aspects. In this paper we have reported a solution exploiting the use of model-based descriptions of interactive applications that facilitate the development and the dynamic update of versions that depend on the interaction resources available. The models allow the generation of different versions of the interactive application that exploit the different modalities according to policies defined in adaptation rules, which are separated from the UI definition and can be modified as a separate aspect. The solution proposed is able to support adaptation at various granularity levels ranging from changing the interactive application since the interaction modality has changed to small modifications of the current version. We have discussed its application to a warehouse picking case study, indicating how a set of integrated tools (authoring environment, adaptation server, dynamic interactive application generators) have been exploited, and we have also reported on an early user test related to the adaptive rules considered.

The result of the user test showed in general that adaptive rules are received well if they trigger the delivery of additional information. Omitting some information without notification, however leads to some usability problem.

The MARIA authoring environment including the interactive application generators used in this work are publicly available at http://giove.isti.cnr.it/tools/MARIAE/home

Future work will be dedicated to further engineering the solution proposed and apply it to other case studies. In addition, we plan to do some studies targeting designers and developers of adaptive interactive applications in order to better assess how their work is facilitated through a model-based approach.

This work has been supported by the SERENOA EU ICT Project, http://www.serenoa-fp7.eu/

References

1. Weaver, K.A., Baumann, H., Starner, T., Iben, H., Lawo, M.: An empirical task analysis of warehouse order picking using head-mounted displays. In: 28th International Conference on Human Factors in Computing Systems (CHI 2010). ACM, New York (2010)
2. Schwerdtfeger, B., Klinker, G.: Supporting Order Picking with Augmented Reality. In: 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 91–94 (2008)
3. Ali, S., Lewandowski, A., Rett, J.: A SOA based context-aware order picking system for warehouses using Laser Range Finder and wearable computer. In: 2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 1–8 (2011)
4. Object Management Group (OMG): Documents Associated with Business Process Model and Notation (BPMN) Version 2.0, `http://www.omg.org/spec/BPMN/2.0/`
5. Paternò, F., Santoro, C., Spano, L.D.: MARIA: A universal, declarative, multiple abstraction-level language for service-oriented applications in ubiquitous environments. ACM Trans. Computer-Human Interaction 16(4), 1–30 (2009)
6. Manca, M., Paternò, F.: Supporting Multimodality in Service-Oriented Model-Based Development Environments. In: Forbrig, P. (ed.) HCSE 2010. LNCS, vol. 6409, pp. 135–148. Springer, Heidelberg (2010)
7. Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., Young, R.: Four Easy Pieces for Assessing the Usability of Multimodal Interaction: the CARE Properties. In: Proceedings INTERACT 1995, pp. 115–120 (1995)
8. Honkala, M., Pohja, M.: Multimodal interaction with XForms. In: Proceedings ICWE 2006, pp. 201–208 (2006)
9. Lawson, J., Al-Akkad, A., Vanderdonckt, J., Macq, B.: An open source workbench for prototyping multimodal interactions based on off-the-shelf heterogeneous components. In: Proceedings ACM EICS 2009, pp. 245–254 (2009)
10. Lin, J., Landay, J.A.: Employing Patterns and Layers for Early-Stage Design and Prototyping of Cross-Device User Interfaces. In: Proc. CHI, pp. 1313–1322 (2008)
11. Obrenovic, Z., Starcevic, D., Selic, B.: A Model-Driven Approach to Content Repurposing. IEEE Multimedia, 62–71 (January, March 2004)
12. Octavia, J., Vanacken, L., Raymaekers, C., Coninx, K., Flerackers, E.: Facilitating Adaptation in Virtual Environments Using a Context-Aware Model-Based Design Process. In: England, D., Palanque, P., Vanderdonckt, J., Wild, P.J. (eds.) TAMODIA 2009. LNCS, vol. 5963, pp. 58–71. Springer, Heidelberg (2010)
13. Sottet, J.-S., Ganneau, V., Calvary, G., Demeure, A., Favre, J.-M., Demumieux, R.: Model-Driven Adaptation for Plastic User Interfaces. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4662, pp. 397–410. Springer, Heidelberg (2007)
14. Stanciulescu, A., Limbourg, Q., Vanderdonckt, J., Michotte, B., Montero, F.: A Transformational Approach for Multimodal Web User Interfaces based on UsiXML. In: Proc. ICMI, pp. 259–266 (2005)
15. Yang, T., Ferati, M., Liu, Y., Ghahari, R.R., Bolchini, D.: Aural Browsing On-The-Go: Listening-based Back Navigation in Large Web Architectures. In: Proceedings ACM CHI 2012, pp. 277–286. ACM Press (2012)

Back of the Steering Wheel Interaction: The Car Braille Keyer

Sebastian Osswald, Alexander Meschtscherjakov, Nicole Mirnig,
Karl-Armin Kraessig, David Wilfinger, Martin Murer, and Manfred Tscheligi

CD Laboratory, ICT&S Center, University of Salzburg
Sigmund Haffner Gasse 18, 5020 Salzburg, Austria
firstname.lastname@sbg.ac.at

Abstract. In this paper, we present a novel text input approach for car drivers: The Car Braille Keyer combines a keyer concept (defined as keyboard without an actual board) and the braille code (i.e. blind writing method) at the back of the steering wheel. This concept allows eyeless text input while driving and simultaneously leaving the hands on the steering wheel. We present a prototype of the Car Braille Keyer along with an expert evaluation and a user study. The prototype consists of two sets of three buttons each, both of which are fixed on the back side of the steering wheel (one on the left, the other on the right side). The six buttons are designed to match a braille character like they can be found in the braille language. This approach allows for entering a character or command with only a single input combination without the need to look at the keys. In our prototype we added visual output in the head up display (HUD) as well as auditive feedback to enhance the interaction. To evaluate the system, we performed a heuristic evaluation with five HCI experts. Based on their feedback, we iterated the design of the prototype and added a learning tool for interaction using the Car Braille Keyer. An initial user study with the iterated prototype and twelve participants showed a good overall usability (SUS score=73.75) as well as a good acceptance rate based on the Technology Acceptance Model (TAM).

Keywords: automotive user interface, steering wheel, braille, keyer, chorded keyboard, acceptance, user studies.

1 Introduction

Entering text in the car is a challenge for the driver - especially while driving. It is sometimes necessary or desired to enter text in the context of a moving car, e.g. when a driver wants to enter a new destination into a navigation system or to enter a song title which should be played next. Although it is our opinion that people should not text while driving, some people want to be able to twitter while sitting in the driver's seat (e.g. in a traffic jam or at a red traffic light), and some people even want to enter text and send a message while driving the car. To do so they either use technology in the car itself, such as a touch screen or a voice input system, or they use their smartphone while driving [17]. Both

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 49–64, 2012.

approaches are potentially dangerous since interacting with such systems often requires the driver to take his/her hands off the steering wheel and loose sight of the road. Speech input does not suffer from these shortcomings but has other problems such as a high rate of speech recognition errors and lower input speed [10]. Thus, there is still a need for convincing ways to input text in cars.

Our general approach is driven by an old paradigm for car drivers: Anyone who is driving a car should keep his/her eyes on the road and his/her hands on the steering wheel. In accordance with this paradigm we envision a concept where the input modality is attached to the steering wheel and output is displayed on the windshield. Since the driver has to look through the windshield to see the road, a head-up display (HUD) is a promising approach to visualize information. In modern cars, the front side of the steering wheel is already being used for various purposes (e.g. knobs and switches to operate an in-car entertainment system, handle automated cruise control, interact with a mobile phone, or access car information).

Within the HCI community, the steering wheel has been used for novel input and output concepts. Enriquez et al. [9] used pneumatic elements on the steering wheel to alert drivers by means of haptic feedback. Gonzalez et al. [11] used a number of small touchpads positioned on the steering wheel that the driver can manipulate to interact using his/her thumbs. Pfeiffer et al. [19] suggested a multitouch-enabled steering wheel with the whole inner circle being a touch interface and Döring et al. [7] developed a steering wheel gesture set for the very same prototype. In [18] the authors evaluated the effect of technology acceptance towards different input modalities on a steering wheel and proposed the steering wheel as an area for an ambient agent, which could be displayed on a touch screen surface. All these approaches have in common that they use the front side of the steering wheel for input and/or output and most of them not explicitly for text input. A place, which has been rather neglected, is the back side of the steering wheel. Our approach uses this place in the car to integrate a text input modality in an ambient and unobtrusive way.

This paper is structured as follows: First, we outline current approaches for text input in the automotive domain. Second, we present the Car Braille Keyer concept including its foundation in other keyer approaches as well as the braille language. Third, we describe the prototype and its implementation. Thereafter, we present the two-step evaluation of the Car Braille Keyer (expert and user). Finally, we discuss our findings and give an outlook on future work.

2 Related Literature

Most of the state-of-the-art solutions for text input in the car use either multiple button configurations or touch screens located in the central console. A different approach has been taken mainly by German car manufacturers (e.g. Mercedes, BMW, Audi), who use a rotary knob as a multifunctional input device and a display as the corresponding output device, both of which are also integrated in the central console. Letters are entered by rotating the knob and pushing it to confirm the selection. A recent approach is to use gestures for character input. Audi, for instance, integrated a touchpad in the area of the gearshift or on top of the rotary device that can be used for character input via gestures.

Text input via the steering wheel was investigated by the Kern et al. [12], who compared different positions for an interface which enables handwritten text input while driving. The results of their study indicate that handwritten text input on the steering wheel is well-received by the users and that the visual feedback should be presented in the dashboard area or directly on the steering wheel. Nevertheless, handwritten input on the steering wheel requires a frequent hand-eye coordination to control the manual input process, which leads to an increasing driver distraction.

Besides manual text input, speech is often used as an input modality in the automotive context. Gartner et al. [10] showed that speech is superior to manual input in many ways (e.g. improvement of the driving quality with speech instead of manual input, less glances of the road). Disadvantages of speech input, however, were a lower input speed as compared to manual text input and the occurrence of speech recognition errors that were more frustrating for the users. Kun et al. [13] evaluated the effect of speech interface accuracy on the driving performance. They showed how the accuracy of the speech engine, the use of the push-to-talk button, and the type of dialog repair supported by the interface influences the driving performance. Barón and Green [2] summarized 15 human factor studies examining the use of speech interfaces. They found that most of the time, people drove "better" in terms of lane keeping, when they used a speech interfaces compared to a manual interfaces. Certainly, using a speech interface was often worse than just driving. Otherwise, the authors gave the example of one participant who could dial a phone number with a cell phone faster while driving by using his thumb than saying the phone number. Concluding, it can be stated that speech interfaces are often likely to perform better than other interfaces in the car but are strongly dependent on the accuracy of the deployed system.

Here, we want to point out that it is still valuable to take the effort in developing alternative text input systems and do not aim for a comparison. On the one hand, this is reasoned by acknowledging that speech text input systems are superior when they are flawless recognizing natural speech and on the other hand due to the circumstance that we are aiming for a laboratory setup and thus environmental noise that affect the speech recognition cannot be reproduced in an appropriate way.

3 Background: Car Braille Keyer Concept

In this section, we describe the Car Braille Keyer concept. Following the *eyes on the road and hands on the wheel paradigm*, we use the steering wheel as input and an HUD as output modality. More precisely we envision the back of the steering wheel as the input area of interest. We want to use this area for the more complex task of text input than what it is used for currently like changing gears, operating windshield wipers and turn signals. Since a traditional keyboard seems not useful for text input via the back of the steering wheel, we take the chorded keyboard and related keyer concept as a base for our approach.

Keyer Systems. Generally, a chorded key set is an input device that allows to enter commands by simultaneously pressing a set of keys as introduced by Douglas Engelbart in the 1960s [8]. A keyed system, which is a device with keys arranged on it that can often be held in one hand, takes this approach and thus provides multiple input potential. The Twiddler[1] [14] (see Fig. 1a) and the Septambic Keyer[2] (see Fig. 1b) are examples of one-handed chording keyboards. Both systems are designed to be held in one hand while text is being entered. The Twiddler is equipped with six keys to be used with the thumb, and twelve with the fingers, whereas the Septambic Keyer has three thumb and four finger switches. Based on the number of keys, the Septambic Keyer allows for 47 different combinations of key presses, while the Twiddler allows over 80,000.

(a) The Twiddler 2 (b) The Septambic Writer

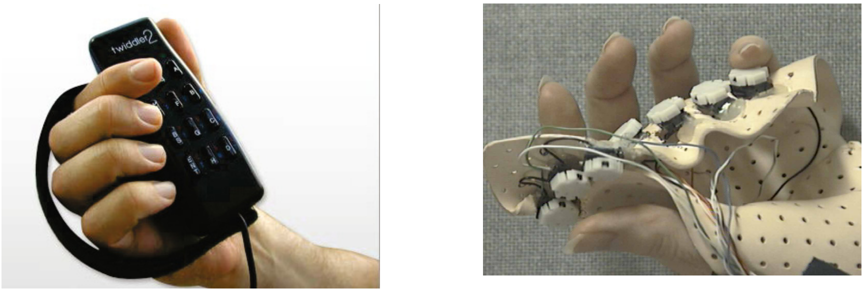


Fig. 1. Keyer Systems that allow for Text Input

The idea of our approach is to allow for combinational input through pressing buttons in varying constellations on the back of the steering wheel. According to Baudisch et al. [3], back-of-device interaction has the advantage that the user's fingers do not occlude content that prevents precision. They addressed mostly touch interaction and proposed four different form factor concepts that further showed the applicability of the back-of-device interaction. We, however, do not cluster the front side of the steering wheel with additional buttons but only use the backside area.

The next question which arises is the number of buttons to be attached to the back of the steering wheel and how they should be arranged. Above that it has to be decided how the interaction with the keyed system exactly should look like (i.e. which keys have to be pressed to enter a character or command). Since

[1] `http://www.handykey.com`
[2] `http://www.wearcam.org/septambic/`

the keyer is attached to the back of the steering wheel, writing "blind" has to accomplished. For that the braille language might be usefull.

Braille Writer. The braille language was developed by Louis Braille [4] to allow blind people to read and write through haptic sensation. Braille is designed to let blind people read by means of physically raised dots in a three by two binary matrix. Characters can be identified within this matrix through the tactile perception of different dot combinations. The combinations are used to picture letters, numbers, special characters or letter combinations (see Fig. 5).

(a) Braille Writer

(b) Braille Touch Writer

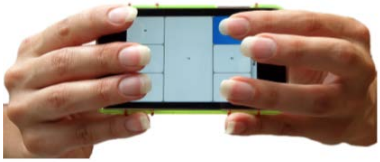


Fig. 2. Braille Writing Systems

Writing in the braille language is often carried out by so-called braille writers that mostly provide seven horizontally orientated buttons (see Fig. 2a - image from http://www.perkins.org/). These writers look like a typewriter, except for less input buttons. Typing in braille is easier as reading, as it only requires to learn the braille character code. For reading, in contrast, it is necessary to train the senses to distinguish between physically raised dot combinations. Taking the braille writer a step further, Romero et al. [20] described a system called BrailleTouch, that allows for eyeless text input on mobile devices. Their prototype provided the functionality of entering braille characters with a tablet or a smartphone with the help of an adaptive interface (see Fig. 2b[20]).

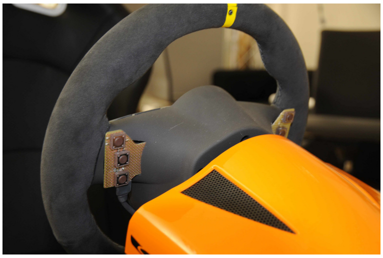
Regarding performance, they reported that braille writers are a valuable text input alternative. In a user study conducted with a touch-based braille writing system, some braille typists were able to write faster than they could on a QWERTY keyboard. Furthermore, the researchers reported that one visually impaired person, who was already familiar with the braille language, typed at a rate of 32 words per minute with an accuracy of 92%. An untrained person with no visual impairment who had never used a braille writer before, was only able to type 25 words per minute but with 100% accuracy after one week of training.

Car Braille Keyer. For the Car Braille Keyer concept we merged the approach of a keyed system with the interaction design of the braille code. The concept consists of six buttons on the backside of the steering wheel that allows for text input. On both sides of the steering wheel (left and right) three buttons are aligned vertically. Taken together, these six buttons match a braille cell, like they can be found in the braille language. The main idea behind the concept is to enter a character or command with only a single input combination without the need to look at the buttons for input (e.g. for A the driver press the top left button). When entering a character, the driver's hands remain on the steering wheel. Based on this concept we built a functional prototype that consists of the hardware and a software to translate input. In the next section a prototype description is given.

4 Car Braille Keyer Prototype

For our prototype we used a Porsche GT3 RS steering wheel[3] as the basis and aligned three buttons on each side of its horizontal axis. For the implementation we chose to use mini push buttons as they provide a well-recognizable haptic feedback when pressed and released (see Fig. 3a and 3b for details). To receive and use the button input, we connected the buttons to our study computer. We utilized an Arduino UNO board that served as the information bridge between the computer and the steering wheel hardware. For more detail see [15].

(a) The back-of-device buttons



(b) The prototype in use

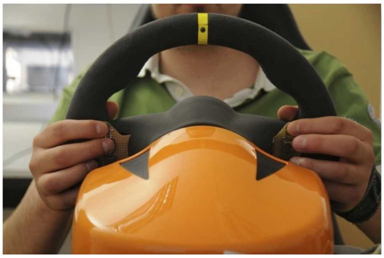


Fig. 3. The Car Braille Keyer Prototype

To make use of the signals we implemented an administration tool, programmed in JAVA to interpret the received inputs of the Arduino UNO board. The tool uses text processing software as described below, to translate the braille-based input into characters and commands. It is also responsible for the system output, which is visual feedback implemented in a head up display (see Fig. 4). The administration tool allows to configure the visualization frames as for example the font settings, colors or output position in runtime.

Sound output was implemented to generate auditive feedback. The translated input from the buttons is matched with a predefined set of characters and audibly confirmed if correct. For single characters we used a database of equivalent

[3] `http://www.fanatec.de/`

MP3 files. For the reading passages (i.e. the whole message is confirmed verbally at the end of entry) the Mary TTS Engine[4] was used. We further wanted to have an overview of the whole input performed with the buttons why we integrated a logging tool to track every action performed (e.g. character input, errors, timestamps).

The above mentioned text processing software *braille mode* allows for transferring the text into a second program. The output of braille mode is visualized in a simulated HUD display (see Fig. 4). The driver can enter characters with the push buttons (in braille) and receives visual feedback on the HUD as well as verbal feedback for each letter. After finishing the text, the driver needs to confirm the message, which was implemented as an additional command. The whole text is then read out loud and transferred to the application handler, where a service can be chosen to send the message (in this case Twitter). For the Twitter application a warning message is additionally given upon reaching 120 digits and the limit of 140 digits (limitation of Twitter posts). In principle the message can be sent to any service (e.g. email, short-message-service, Facebook).We integrated a Twitter client with Internet connection and a registered Twitter account. Thus the prototype enables the driver to enter a message through pressing button combinations subsequently (stepwise like the button on the top left side for A), confirm the message (press all six buttons) and choose Twitter as a service to deliver his/her message.

Fig. 4. The HUD Mode of the Braille Writer

For input we utilized the standard braille code and extended it with additional commands to meet the requirements of the braille mode. These included typical text input commands such as *move cursor to the left*, *move cursor to the right* and *enter*. In Fig. 5 the whole code used is shown. To enter one of this characters the driver has to press the buttons according to the dot pattern. For example, in order to enter the letter C the top left and the top right button need to be pressed.

[4] `http://mary.dfki.de/`

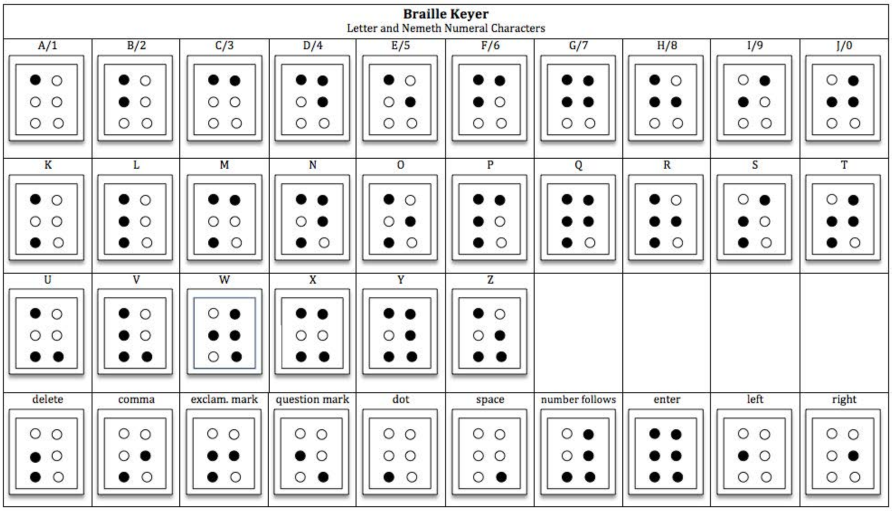


Fig. 5. The Letter, Nemeth Numeral and Functional Characters of the Car Braille Keyer

The entire setup of the prototype includes the steering wheel with the Car Braille Keyer mounted on a rod system, a driver's seat, two projectors for a car simulation software and the simulated HUD, two loudspeakers, and two computers that run the prototype software and the driving simulator environment. To simulate the HUD we used two overlaying beamer pictures. One beamer (BenQ SP840) was placed behind the car simulator seat to project the car simulation. The second beamer (Sony XGA VPL-EX1) was mounted in front of the steering wheel to project the braille interface into the image of the car simulation. To equalize any light differences we adjusted the brightness level. We used two PCs, one was running the car simulator and thus connected the steering wheel and the pedals to get the driving-relevant data. The other ran the administration tool which enabled the HUD interface projection and the buttons.

5 Evaluation

To evaluate and iterate the the Car Braille Keyer concept we conducted an expert and a user evaluation. In a first step, we explored the general feasibility of the system and gathered ideas to refine the setup with the help of an expert evaluation (heuristic evaluation). In a second step, we evaluated the usability (System Usability Scale - SUS) and acceptance (Technology Acceptance Model - TAM) of the prototype in a user study. In table 1 we summarized our research goals and questions as well as the method we applied in our evaluation to answer the research questions. In the following, we describe the expert evaluation, the follow-up prototype iteration, and the user study.

Table 1. Research Goals and Questions

	RG/RQ	Method
RG1	Explore feasibility and generate iteration ideas	Expert Evaluation
RQ1a	Which usability problems can be identified?	Heuristics
RQ1b	What recommendations can be given for refinement?	Heuristics
RG2	Evaluate usability and acceptance in a driving situation	User Study
RQ2a	How is the system usability rated?	SUS, Questionnaire
RQ2b	How is the system perceived in terms of enjoyment?	Questionnaire
RQ2c	Do the users accept the system?	TAM

5.1 Expert Evaluation

For the expert evaluation we relied on usability heuristics based on Nielsen's heuristic evaluation method [16]. It is an informal, low-cost evaluation technique, which was rated as one of the top techniques in use in a survey of usability practitioners. The major difference between evaluating automotive user interfaces and evaluating traditional displays comes from the way users interact with the system. Automotive interface users have a primary task they have to spend their attention on: the driving task. Obtaining information from a display or modality thus competes with the driving task. Drivers do not use such displays as they would use a desktop-based computer but perceive information in short glances and have only little time for hand-eye coordination to interact with a system. Consequently, some of Nielsen's original heuristics are less useful while other aspects that effect automotive user interfaces were not addressed.

For example, the heuristics do not consider the perceived distraction and the perceived mental workload of a driver, which is a central element in a human-car interaction scenario. These factors highly affect the main task of driving but cannot be addressed with Nielsen's heuristics. Our conclusion was that the methodology of heuristic evaluation could still be applied by means of adding three new heuristics while leave out three less important heuristics.

Setup. Our goal in this study was to obtain feedback on a set of car-focused heuristics that were based on Nielsen's heuristics, but were modified to be more applicable to automotive user interfaces. The set of heuristics, each single heuristic consisting of a title and definition, were handed out to the experts prior to the evaluation. We used the following heuristics:

1. Visibility of system status - The system should always keep users informed about what is going on.
2. Accessibility and organization of input elements (new) - The system should be within reach of the driver or passenger. The modalities should support the user's action and clearly communicate its intended use.

3. User control and freedom - Users need a clearly marked “emergency exit" to leave the unwanted state. Support undo and redo.
4. Consistency and standards - Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.
5. Error prevention/recovery - Careful design is even better than good error messages, which prevents a problem from occurring in the first place.
6. Safety awareness (new) - The system should raise awareness about both, the safety issues that arise from its use and safety threats through the driving situation.
7. Recognition rather than recall - Minimize the user’s memory load by making objects, actions, and options visible.
8. Flexibility and efficiency of use - Accelerators - unseen by the novice user - may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users.
9. Aesthetic & minimalistic design - Dialogues should not contain information which is irrelevant or rarely needed.The same is valid for input elements and the overall design of the modalities.
10. Distraction and mental workload (new) - Distraction of the driver from the primary task of driving should be avoided in the car.

We invited five HCI experts to take part in the expert evaluation. Each expert received a brief introduction into the Car Braille Keyer system. Then, each expert was asked to evaluate the system with the provided heuristics and to provide a short description of the violation as well as the corresponding heuristic. They further were asked to write down what they liked about the system.

Results. The evaluation lasted for 1.5 hours for each of the five experts. We collected the notes and categorized them to create an overview about usability violations. Results show that the heuristics 7, 8 and 10 were violated most often (see table 2). One of the main issues the experts addressed was that the barrier in learning the system was too high. They proposed a learning tool which would help to lessen learning issues. Positively mentioned was the placement of the buttons, which was perceived as intuitive. Some experts stated that the auditive feedback was very helpful and well-designed and that they liked that no touch screen or speech control was necessary. They further pointed out that the system allowed them to look straight ahead while typing and keep their hands on the steering wheel. The haptic feedback of the buttons was also praised and allowed for a clear distinction between different input gestures (i.e. braille characters). Although some experts complained about the high workload, they admitted that being familiar with the braille code would reduce this workload vitally. Nevertheless, the experts stated that they do not see the system being accepted from the users in this early stage.

Prototype Iteration. According to the findings from the expert evaluation, we decided to implement a learning tool that aimed for easing the Car Braille Keyer

Table 2. Expert Evaluation: Violation of Heuristics

Heuristic	Violation count	Example
1	4	It is not obvious that in the application menu an entry immediately leads to sending a message
2	3	The buttons in the application menu should not stick together - they should be separated
3	3	There is no off-switch or exit button
4	2	It is better to have six items rather than four in the secondary menu
5	2	The system does not provide support to prevent errors
6	3	The menu and the text block the view on the road
7	10	The user has to learn the braille code
8	10	Sending a message does not clear the text entry field
9	4	It would be nice if the position of the interface could be adjusted to meet the driver's requirements
10	9	Four buttons, two on each side, would be easier to coordinate finger movements

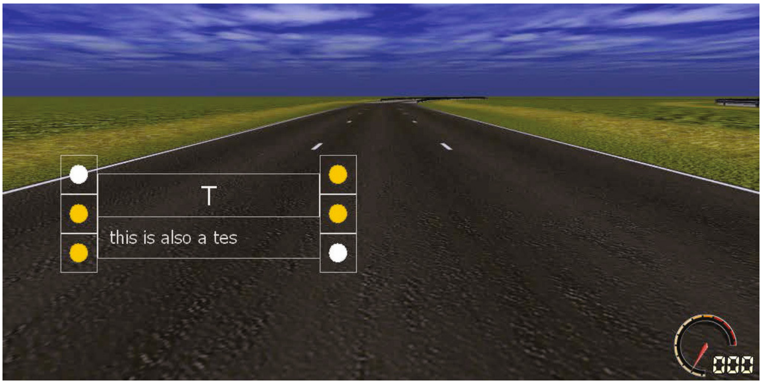


Fig. 6. The HUD Learing Mode of the Braille Writer

interaction in the beginning. The tool was designed to work in the same way like the braille mode. For the learning tool input was the same way through the six buttons on the back of the steering wheel, and the visual output was given in the HUD. As visualized in figure 6 the learning mode showed the letter that needed to be written in the middle of the HUD segment and the text already written was presented below this letter. The buttons that needed to be pressed to generate the letter were displayed with yellow marks on either side of the HUD segment. The colors of the buttons changed to green when the input was correct or red if not. The learning mode had to be initialized via the admin tool, which allowed to input text for different learning sessions and enabled the researcher to design sessions with varying duration. In addition to the learning tool, we implemented some minor changes within the system (e.g. we included a cursor and redesigned the visualization for a better match with the keyer setup).

5.2 User Study

With the iterated prototype and the newly implemented learning tool we conducted a user study. The main goal of this study was to get feedback on usability and acceptance of the system.

Setup. The user study was conducted with twelve participants. We asked for demographical data and for how long they possess a driving license and if they have any experience with the braille language. Additionally, we asked them if they ever texted while driving with their cellphone or with a text input system in the car.

Then we explained the idea of the system to the users and allowed them to make themselves comfortable first with the driving simulator and then with the Car Braille Keyer. In the next step, participants used the learning tool with the aim to learn seven characters: E, B, T, M, A, R, and L. They were chosen based on a pretest in which the criteria were the numbers of buttons required to enter the letter, the side of the steering wheel on which the required button need to be pressed vary, and if the buttons form a easy to remember combination. Accordingly, we balanced the chosen characters and created ten different words that consisted of theses characters. Participants learned the characters with the learning tool developed in the iteration phase. The letters and combinations were randomly displayed in the HUD and the participants performed the corresponding button combination until they felt comfortable in remembering all seven letters. Afterwards, participants were allowed to make themselves comfortable with the driving simulator.

At the beginning of task one that was performed without driving, we described a usage scenario: “You want to twitter a message to a friend while driving. You use the Car Braille Keyer to send five different words via the Twitter application. Every word is a single message that you need to write, confirm and send via Twitter to complete the task." Participants needed to remember and write every letter with different button combinations that they learned in advance. Failures could be fixed with a delete combination (see figure 6) and the message was send by pressing all six buttons. The researcher said each word to the participant and additionally showed the written word on a sheet of paper to overcome misunderstandings. The second task was performed in the same way under a driving condition. Before the driving task started they were asked to drive straight ahead for a short while at a constant speed of 60 km/h.

To evaluate the usability we handed out the SUS questionnaire [5]. For assessing the acceptance of the system the TAM questionnaire [6] was used. At the end of the session four additional questions were posed in a questionnaire:

1. What is your overall impression of the system?
2. Did you have fun while using the system?
3. Would you be willing learn the braille language to text while driving?
4. Do you have any ideas on how to improve the system?

Results. In total, 12 people (7 male, 5 female) participated in our study with an average age of 27.7 years (SD=2.89). All participants owned a driving license for an average of 9.3 years (SD=4.15). Ten out 12 participants stated that they texted (SMS) at least once while driving and six reported that they had used an information system for navigational input while driving at least once. None of the participants had previous experience with the braille language. All participants were confident that they have learned the combinations after a 3 minute training with the learning tool.

Regarding question 1 (overall impression), 6 participants mentioned that they were really surprised how fast the system reacted and how easy it was to perform the gestures. Two participants said that their "hands" were able to memorize the movements after a short time, which made it easier to remember the letters. They overall liked that the hands remained on the steering wheel and that the interaction steps were easy. On the other hand, 7 participants mentioned that the interaction requires a lot of learning and that the interaction is thus not intuitive. Question 2 (Fun while using) was answered positively by 9 participants. They stated that it was challenging to learn the input combinations and that they would like to see a score on how well they performed.They remarked that it was fun that the input system differs so much from common devices. Six participants answered Question 3 (willingness to learn braille) with yes, 6 with no. Valuable feedback was collect regarding question 4 (ideas for system improvement). One idea was to integrate a T9-feature (like used in cellphones) to allow for faster text input. Another idea was to physically raise the middle button on each side, to facilitate finding the correct button.

The SUS scored a value of 73.75. The SUS score ranges between 0 and 100, the higher higher a score the higher the system usability is rated. Scores above 60 are considered as valuable, in the sense that people accept a system. Referring to the rating scales described in [1], the SUS score of our system can be described in three ways: the acceptability is "acceptable", a school grade of a C was reached and the user rating of the system can be described as good. Regarding the technology acceptance questionnaire, which was answered on a 5-point Likert scale an overall good acceptance could be assessed. The TAM reached a mean score of 3.5 (ranging from 1(very good) to 5 (very bad)). In contrary to the expected non-acceptance as predicted from the experts, the TAM results showed a fairly good acceptance.

The error rate during the tasks was rather low. Participants made in total 21 failures (M=2,1/trial) No difference could be noticed between the *not driving* and the *driving condition*. The researchers had to support two participants in cases were they forgot a combination.

Summarized it can be said that the prototype was accepted by the participants and that the usability was considered as valuable. Further, the error rates as well as the answers to the questions showed the potential of the Braille Keyer.

6 Discussion

We developed a prototype for back-of-steering wheel text input by combing a keyer approach with the braille language. The advantages of the approach are threefold. First, with the Car Braille Keyer it is possible for the driver to enter a text while driving without taking his/her eyes from the road and hands off the steering wheel. Second, all characters and commands can be entered in a single interaction step by pressing a chord of keys. This is beneficial compared to systems that need a selection process (e.g. the multi-functional controllers in the BMW iDrive or the Audi MMI concept). Third, the interaction works eyeless with haptic feedback that needs no hand-eye coordination, which is more advantageous compared to touch screens or touchpads.

We are aware that the Car Braille Keyer has not only advantages. Even though the road is constantly in sight of the driver it takes some mental effort to read text in a HUD. The switch of focus between the road and the windshield is problematic. On the one hand, it might be tiring for the eyes to switch focus often while driving. Therefore, we suggest to visualize the text as far away from the driver as possible. On the other hand, a mental focus on the text could be disturbing and the reaction rate to unpredicted events could be increased. Participants in the user study mentioned that they would not need a visualization of the text after a learning phase as the audio feedback would then be sufficient. Further the audio feedback can not just used to read the entire message but also during the editing process. When space is pressed to finish a word, the audio feedback could read the whole word as this is widely used by blind persons to be more sure about the words written.

As the expert evaluation revealed, it is a challenge for users to learn the braille language in order to operate it efficiently. However, the learning tool we developed has proven to be a step in the right direction as the user study showed. Nevertheless, it remains as a certain threshold for people to start using it.

In terms of methodological issues, we can conclude that the heuristic expert evaluation in the early stage of prototyping was valuable. It led to a refinement of the prototype and resulted in the development of the learning mode. The user study showed potentials of the Car Braille Keyer concept. In general it can be argued that entering text while driving is always problematic since it increases distraction from the primary driving task. The Car Braille Keyer follows the paradigm hands on the wheel and eyes on the road, but it does not address the issue of cognitive distraction. Nonetheless, the positive feedback of both, experts and users, approves our approach and highlights its potential.

7 Conclusion and Future Work

Information technology in the car becomes increasingly complex. One of the major challenges is to design feasible solutions that support the driver but do not affect the driving performance. Addressing this challenge, this paper proposed an ubiquitous device that allows for text input while driving. The Car Braille Keyer enables the drivers to keep their hands on the steering wheel and the eyes on the road. We extensively described the background of this devices in areas of keyer systems and automotive input systems and addressed its potential beyond common text input systems in the car that lacks of usable input concepts.

We especially want to point out that in the car context it is mandatory to address safety concerns. As researchers we see the difficulty in designing interaction that allows for texting, but we want to point out that drivers are already texting while driving. The high accident rate caused by texting, documents that the reach of regulations and prohibitions are limited which is why our goal was to design an in-car system that allows for texting but in a safer way. We are aware of how the cognitive workload can rise through using the system while driving. Thus, it is important to evaluate the workload and distraction in a next step. We further see potential in a comparative study between state of the art systems for text input and the Car Braille Keyer to evaluate the performance and especially pay attention on how the system causes distraction.

Acknowledgements. The financial support by the Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged (Christian Doppler Laboratory for Contextual Interfaces).

References

1. Bangor, A., Kortum, P., Miller, J.: Determining what individual sus scores mean: Adding an adjective rating scale. Journal of Usability Studies 4(3), 114–123 (2009)
2. Barón, A., Green, P., Michigan, U.: Transportation Research Institute of Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review. Tech. rep., University of Michigan, Transportation Research Institute (2006)
3. Baudisch, P., Chu, G.: Back-of-device interaction allows creating very small touch devices. In: CHI 2009, pp. 1923–1932. ACM, New York (2009)
4. Braille, L.: Procedure for writing words, music and plain song using dots for the use of the blind and made available to them. Royal Institution of Blind Youth, Paris (1829)
5. Brooke, J.: Sus-a quick and dirty usability scale. Usability Evaluation in Industry 189, 194 (1996)
6. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 13(3), 318 (1989)
7. Döring, T., Kern, D., Marshall, P., Pfeiffer, M., Schöning, J., Gruhn, V., Schmidt, A.: Gestural interaction on the steering wheel: reducing the visual demand. In: CHI 2011, pp. 483–492. ACM, New York (2011)

8. Engelbart, D., English, B.: A research center for augmenting human intellect. In: Proceedings of the 1968 Fall Joint Computer Conference, San Francisco, CA, vol. 33, pp. 395–410 (December 1968)
9. Enriquez, M., Afonin, O., Yager, B., Maclean, K.: A pneumatic tactile alerting system for the driving environment. In: PUI 2001, pp. 1–7. ACM Press, New York (2001)
10. Gärtner, U., König, W., Wittig, T.: Evaluation of manual vs. speech input when using a driver information system in real traffic. In: Human Factors in Driver Assessment, Training and Vehicle Design. University of Iowa Public Policy Center, Iowa (2001)
11. González, I.E., Wobbrock, J.O., Chau, D.H., Faulring, A., Myers, B.A.: Eyes on the road, hands on the wheel: thumb-based interaction techniques for input on steering wheels. In: GI 2007: Proceedings of Graphics Interface 2007, pp. 95–102. ACM, New York (2007)
12. Kern, D., Schmidt, A., Arnsmann, J., Appelmann, T., Pararasasegaran, N., Piepiera, B.: Writing to your car: handwritten text input while driving. In: CHI EA 2009, pp. 4705–4710. ACM, New York (2009)
13. Kun, A.L., Shyrokov, A., Heeman, P.A.: Spoken tasks for human-human experiments: towards in-car speech user interfaces for multi-threaded dialogue. In: AUI 2010, pp. 57–63. ACM, New York (2010)
14. Lyons, K., Starner, T., Plaisted, D., Fusia, J., Lyons, A., Drew, A., Looney, E.W.: Twiddler typing: one-handed chording text entry for mobile phones. In: CHI 2004, pp. 671–678. ACM, New York (2004)
15. Murer, M., Wilfinger, D., Meschtscherjakov, A., Osswald, S., Tscheligi, M.: Exploring the back of the steering wheel: Text input with hands on the wheel and eyes on the road. In: AUI 2012. ACM (2012)
16. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People, pp. 249–256. ACM (1990)
17. Noder, S.: Talking and texting while driving: A look at regulating cell phone use behind the wheel. Val. UL Rev. 44, 237–1243 (2009)
18. Osswald, S., Meschtscherjakov, A., Wilfinger, D., Tscheligi, M.: Steering wheel-based interaction: Potential reductions in driver distraction. In: International Joint Conference on Ambient Intelligence, AmI 2011 (2011)
19. Pfeiffer, M., Kern, D., Schöning, J., Döring, T., Krüger, A., Schmidt, A.: A multi-touch enabled steering wheel: exploring the design space. In: CHI EA 2010, pp. 3355–3360. ACM, New York (2010)
20. Romero, M., Frey, B., Southern, C., Abowd, G.D.: Brailletouch: designing a mobile eyes-free soft keyboard. In: MobileHCI 2011, pp. 707–709. ACM, New York (2011)

PermissionWatcher: Creating User Awareness of Application Permissions in Mobile Systems

Eric Struse[1], Julian Seifert[2], Sebastian Üllenbeck[1], Enrico Rukzio[2], and Christopher Wolf[1]

[1] Horst Görtz Institute, Ruhr University Bochum, Germany
{firstname.lastname}@rub.de, chris@Christopher-Wolf.de
[2] Media Informatics Group, Ulm University, Germany
{firstname.lastname}@uni-ulm.de

Abstract. Permission systems control access of mobile applications to other applications, data, and resources on a smartphone. Both from a technical and a social point of view, they are based on the assumption that users actually *understand* these permissions and hence they can make an informed decision about which permission to grant to which piece of software. Results of a survey conducted for this article seriously challenges this assumption. For instance, over a third of participating Android users were not able to correctly identify the meaning of the permission *Full Internet Access*. We developed *PermissionWatcher*, an Android application which provides users with awareness information about other applications and allows to check on the permission set granted to individual applications. In a field study with 1000+ Android users, we collected data that provides evidence that users are willing to follow security principles if security awareness is created and information is presented in a clear and comprehensive way. Therefore, we argue that it is essential for security policies to take the abilities of the target audience into account.

Keywords: Usable Security, Mobile Phones, Android, Access Rights.

1 Introduction

Modern mobile phones or *smartphones* have become truly ubiquitous computers. For instance, they enable users to edit texts, browse the internet, access all kinds of online services at any place. Also, increasing storage capacities allow users to keep a multitude of data on their devices. Some of these data and files are regarded as highly sensitive and private by the users.

To fully exploit the capabilities of smartphones, modern mobile operating systems allow users to install applications of their choice (also referred to as *app*). Obviously, this creates new threat scenarios: in particular, users unwillingly installing malicious software which steals data or uses the smartphone's resources (*e.g.* making calls, sending text messages). One approach to prevent this is to totally close the system. Users are only allowed to install reviewed and signed

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 65–80, 2012.

applications. For instance, Apple's *App Store* follows this approach to an extent. A true security review would lead to high costs.

A different approach is leaving the system completely open and allowing users to install whichever application they like. However, this requires rules that regulate what data and which functionalities a certain piece of software may access for security reasons. For instance, in the case of Android, a comprehensive set of *permissions* for applications exist which can be reviewed by a user before installing an application and consequently granting access to the smartphone. However, such a system is entirely based on the assumption that users are familiar with these access rights and are further able to understand them in order to make a qualified decision whether to install an application or not. However, it is not enough that the user understands *one* isolated access right, but many, and in particular their dependencies among each other.

Therefore, this work aims to investigate whether users of smartphones with a permission-based application security model actually do understand these access rights and the implications they have. Further, this work addresses the question if increasing the users' awareness of what a piece of software potentially could do with a specific set of permissions has an impact on their decision which application to install.

Android is a good example for the rule-based application security model and was hence chosen for our investigations. In an initial online study we explored the understanding of Android users of the permission concept. The results show that a large amount of users does not understand basic access rights and what consequences they may result in (cf. sect. 3). Also, a majority of participants indicated that they are willing to uninstall applications which have permission to accessing too many resources of their phones if they were aware of them. Thus, we designed and implemented *PermissionWatcher*, an application for Android phones that analyses permissions of other applications installed on the phone. Based on a custom set of rules, which we developed, PermissionWatcher classifies applications as *suspicious* if any of the rules apply. Through a home screen widget PermissionWatcher increases the user's awareness of potentially harmful applications. In a field trial we collected usage data from 1.000+ different users. About 9% of them used PermissionWatcher to delete suspicious applications directly with PermissionWatcher. For comparison: On 98.7 % of all phones applications with suspicious permissions were found.

The remainder of this article is organized as follows: We discuss related work and illustrate selected basics about Android. In the following, we report on an initial survey in which we assessed the users' understanding of Android permissions. Further, we detail the design and development of PermissionWatcher and further report on a field trial in which PermissionWatcher was tested. Finally, this article draws conclusions and outlines possible next steps.

2 Related Work and Android Basics

Research related to this work can be classified into the following categories: *user interaction*, *smartphone security*, *application market places*, and work concerning the Android system.

User Interaction. Egelman *et al.* [5] shows the necessity to simplify complex security decisions into easy to understand yes/no warnings. In particular, they have investigated phishing sites and used an automated process to rate if a site is dubious or not. If a certain threshold was reached, the site was marked a potentially dangerous, and the user was given the chance to abort loading. Also, there was a potential risk that this engine made mistakes, this was less likely by at least one order of magnitude than the user surfing on a phishing site and becoming a target.

Amer *et al.* [1] consider different ways of displaying warning messages to users. Their main goal it to maximize the impact of a security warning. At the same time, they want to avoid that warnings are perceived as "rude" and also being ignored by users. They show a dramatic difference in user response depending on the way the warning is displayed.

Smartphone Security. General security considerations, but also specific attack vectors for smartphones informed the design of our survey. According to the Microsoft Security Intelligence Report [16] 44.8% of all malware detected, required user interaction for propagation. Note that [16] deals with computers, not smartphones. Still, its information on malware propagation is clearly relevant to this work.

An extensive overview on attack vectors and differences between normal and mobile security can be found in Becher *et al.* [4]. A technology review with a special focus on threats and attacks concerning smartphones is given by Li and Im [15]. The importance to *"scrutinize permission requests"* is highlighted by Hogben and Dekker [14] who analyze risks particularly for smartphone users and give practical recommendations to avoid them.

Market Places and Overall Comparison. Anderson *et al.* [2] examine the application markets for ten platforms, including desktop operating systems, mobile phones, web browsers and social networks. They define *incentives*, *goals* and *stakeholders* concerning application markets, and analyze case studies on each platform. Concerning smartphones, they "*find that these OSes* [Symbian, iOS and Android] *provide significantly more protection to both users and the system than their desktop counterparts, but that there are very important differences among them in terms of a user's control over applications*" (p. 19). Android receives the best protection rating but it is noted that it offers less assurance of system protection compared to the other smartphone operating systems.

Android. To derive specific rule sets (*cf.* Table 1), we considered the following work concerning possible attack vectors in Android.

Shin *et al.* [18] analyze the Android permission-based security system using state machines, leading to a formal security model [19].

There is extensive work on Android malware and data leakage detection by Enck *et al.* . We quote but a few. Enck *et al.* [7] introduce *Kirin*, a framework to enforce security policies in Android. They give a formal representation of the Android security model and develop a policy model featuring a subject-object-rights access matrix.They also present a novel procedure [8] for identifying requirements and creation of rules relating to the analysis of permission sets, and develop a set of rules. Enck *et al.* revise *Kirin* based on these rules and use it to analyze 311 applications. They find five applications with dangerous functionality, and five applications with dangerous but reasonable functionality. Note that our work takes a similar approach on malware detection, however our focus is on awareness and usability; our tool is clearly designed for users who are not knowledgeable in security.

A study on 1,100 applications by Enck *et al.* [6] also detects common misuse of personal information. Enck *et al.* statically analyzed recovered source code of applications, using data flow analysis, structural analysis and semantic analysis. They observe that the International Mobile Equipment Identity is misused as a cookie by many applications, and 51% of (free) applications connect to advertising or analytic networks.

Permissions. The following works focus on the permission system of Android. This is particularly relevant for our work as we focused on permissions both for our survey as for PermissionWatcher.

Barrera *et al.* [3] propose a methodology for empirical analysis of permission-based security. They analyze 1,100 Android applications using self-organizing maps. Felt *et al.* [10] introduce *stowaway*, a tool for detection of *overprivileges* in compiled Android applications. They test 940 applications and argue that one in three applications is overprivileged. Felt *et al.* reason that *"developers attempt to obtain least privilege for their applications but fall short due to API documentation errors and lack of developer understanding"* (Page 11). This is somehow complementary to our approach where we deal with the demand rather than the supply side of Android applications. Another study of Felt *et al.* [11] deals with the effectiveness of permissions. 1,000 chrome extensions (for the Google Chrome browser) and 956 Android applications are analyzed. Felt *et al.* conclude that permissions can be effective, but improvement is possible and necessary. The fact that 93% of the analyzed Android applications have at least one *dangerous* permission is particularly relevant to our work because it implies that users receive warnings about permissions during the installation of almost every application and are hence likely to ignore them.

Furthermore, Felt *et al.* [12] conducted two usability studies (an Internet survey of 308 Android users and a laboratory study of 25 users). They note that only 17% of the participants (in both surveys) paid attention to permissions during application installation, and only 3% of the participants (of the Internet survey) were able to correctly answer all questions on permissions. Again, this supports our claim that end-users are likely to ignore security warnings if they are presented in an unintelligible way.

2.1 Android Basics

Android is a mobile operating system for mobile devices such as smartphones and tablet computers. It is based on the Linux kernel version 2.6 and was developed by the Open Handset Alliance. Android comprises of an operating system, a middleware and key applications. Source code, along with any data and resource files, are compiled by the Android SDK tools into a single *Android package*. In this context, all code in a single package is considered to be one application.

2.2 Security

Android's security architecture relies on two basic *security mechanisms* [13], namely *sandboxing* and *permissions*. We focus on the latter in this article, but describe both for completeness.

Sandboxing. Each application is given a distinct, constant identity (Linux user ID and group ID). Each application runs in a separate process and can only access files that belong to the same user ID (with an exception concerning data stored on SD-cards). The kernel isolates applications from each other and from the system. This process is assisted by the *Dalvik* virtual machine, which is specifically designed for Android. Still, as any application can run native (C or C++) code, the Dalvik virtual machine is no strict security boundary.

Permissions. In Android, a permission mechanism enforces restrictions on the operations an application can perform. Applications have to statically declare their required permissions at install time to gain access to certain hardware features and user data, and to be able to share resources and data. There is no mechanism to grant or withdraw permissions dynamically. Therefore, it is an *all-or-nothing decision*: either, a user grants all required privileges, or cannot install the application. There is a system of grouping labels into four different categories (normal, dangerous, signature, signatureOrSystem). For more detailed discussion on Android permissions, cf. Felt *et al.* [10] or Enck *et al.* [9].

3 User Understanding of Application Permissions

Departing from the assumption that existing means for regulating application permissions in Android do not meet user requirements in terms of clarity, we designed and conducted a survey to investigate the following hypothesis: (H1) User awareness about application access rights in Android is deficient. In particular, users do not know and understand the Android security concept and corresponding access rights. Further, (H2) awareness concerning potential threats can be supported by providing users with clear and comprehensible information. To substantiate this hypothesis, we state two additional sub- hypotheses: (H2.1) users are willing to restrict access rights of applications and (H2.2) they would delete potentially harmful applications.

3.1 Setting

In order to investigate these questions we designed a questionnaire targeted to Android smartphone users. The survey included 15 questions structured in the sections 1) smartphone usage and experience, 2) Android application access rights, 3) user attitudes towards privacy, and 4) general understanding of IT security aspects. In addition, demographic data was collected.

We promoted the online survey via four email lists about IT security (with ≈3,000 receivers) and an email list of the students of the computer science department at the (blind for review). As an incentive, each participant who completed the questionnaire automatically took part in a lottery where they could win one of three gift vouchers (value: 15, 10, and 5 (blind for review)). The survey is biased in two ways: First, the participants are far more likely college students than the average population. Second, the number of security professionals is also clearly too large. Interestingly, we could not find this bias to be reflected in the results: People with a security background gave similar answers as participants with other backgrounds. Moreover, even if the survey were biased in terms of more security awareness and security friendly behavior, this would only make our findings *worse* for the general population.

3.2 Survey Results

In total, 113 complete answer-sets were collected from participants (89 male). Overall, they were aged between 17 to 50 years (Mdn=25), while 87% were aged between 20 and 30 years. 73 participants (65%) were college students, while the others had highly diverse backgrounds such as psychologists, social workers, or engineers. Concerning the usage duration of their smartphones, 73% of the participants indicated to own and use their devices for at least six months (45% longer than one year).

Understanding of Android's Security Concept. Participants rated their knowledge of the Android application security concept on a five point Likert scale (1=none; 5=very good). On average, participants rated their knowledge to be mediocre (Mdn=3.0). To assess a general understanding of the Android security concept, we asked the participants whether applications released in the Android Market are subject to a security vetting process, which is not the case. 14 participants (12%) assumed that all applications would go through such as process, 29 (26%) were not sure, and 70 (62%) choose the correct answer. Accordingly, more than a third of users are not sure or assume a security mechanism which is not existing. Further, we asked if users know at which point application access rights are granted. Three possible answers were provided. Here, a large majority of 91 participants (81%) chose the correct answer ("At application installation"), while 12 (11%) picked the neutral answer, and 10 (9%) choose the wrong answer. 63 (56%) participants answered both questions correctly (3 answered both incorrect) while the remaining picked only one correct. This result does not allow for conclusions concerning the overall understanding of the security concept. Yet, it indicates that fundamental security mechanisms are not fully understood. Concerning the understanding of Android application permissions, we found that participants understand partially what access rights mean. For instance, 99 participants (88%) understood correctly what an application with *read contacts* is allowed to do. However, only 71 participants (63%) knew the correct answer to the question what an application with *full internet access* is allowed to do. Even lower was the percentage of correct answers to the question which actions could be performed by an application with the *make phone call* permission: here only 19 participants (17%) gave the correct answer.

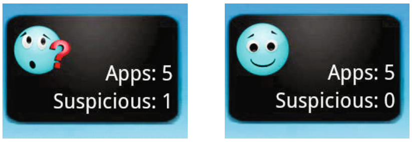
Benefit of Security Mechanisms. Participants were asked to rate their level of agreement to the statement "The available information on Android permissions is sufficient." on a five point Likert scale. A separate neutral answer was available, picked by 10 participants. On average participants rated this statement with 2.0 (Mdn). Further, we asked participants to rate their agreement to the statement "It would be helpful to be able to prohibit access to contacts or sending files to the Internet" where only 3 participants selected the neutral answer. On average, they rated their agreement with 5.0 (Mdn).

In response to the question what they would do if they were warned that an application requires permissions that are potentially harmful, 44% answered that they would delete the application. 28% stated to delete the application in case it would be none of their favorites. 77% of the users agreed to that they would not install the application and search for an alternative. Only 5% stated that they would take no action at all.

In summary, we can conclude that the existing security model is only partially understood by Android smartphone users. This is surprising, as the survey was advertised via email lists received by users with an IT security background. Also, the sample of participants included mainly well educated persons which indicates that even those have difficulties understanding basic Android security mechanism. However, users indicated that warnings concerning security and privacy threats would result in actions such as uninstalling applications which implies that providing users with information would result in a higher level of security.

4 PermissionWatcher Application

In this section, we introduce PermissionWatcher, a mobile application for Android smartphones which increases user awareness about potentially harmful applications installed on her phone. The concept is based on the assumption that users are willing to take security increasing actions (such as uninstalling a potentially harmful application) once they gain knowledge about a potential threat. In order to increase the user's awareness about potential threats, PermissionWatcher provides a widget that can be installed and displayed on the home



(a) (b)

Fig. 1. The PermissionWatcher widget: (a) a worried smiley face indicates suspicious applications. (a) after uninstalling suspicious applications, the smiley appears happy.

screen of the mobile phone. Widgets are a common way to provide users with small pieces of information such as weather information, news tickers, or upcoming assignments taken from the calendar application. Amongst these widgets, the PermissionWatcher widget provides the information of how many applications are installed on the system and how many of these are *suspicious*. When PermissionWatcher detects suspicious applications, a smiley face on the widget emphasizes that the user should take action (see Figure 1). That is, the user can touch the widget to launch the PermissionWatcher application for reviewing details and uninstalling other applications. In case, no applications are detected as suspicious, the smiley face appears as happy.

4.1 Rule Set

PermissionWatcher evaluates the risk of a given application being a potential threat to the user based on a set of rules. As we focus on raising user awareness of Android permissions, this leads to the following two limitations:

1. We inspect single applications. Therefore our rules do not cover permission re-delegation. An application with a certain permission can act as a proxy and perform a task for another application that does not have the permission. Moreover, it is possible to divide a dangerous combination of permissions to separate applications.
2. We do not analyze the code nor the behavior of applications (this would require root privilege or the modification of the Android system). Consequently, we can not detect conventional malware that bypasses the Android security system. For example, we can not detect applications that exercise a root exploit to gain root privileges.

We followed a structured approach to deduce the rule set: First, identifying relevant targets of possible attacks. Then, deriving attack scenarios. This is followed by determining which permission set is required for each scenario which leads to rules based on the permission sets. Finally, similar rules were combined. We have identified three groups of targets:

1. Stored data (*e.g.* contact data, text messages).
2. Hardware features (*e.g.* camera, microphone).
3. Functions (*e.g.* answering calls, send/receive text messages).

Table 1. The set of rules applied in PermissionWatcher

Number	Title	Permission Set
1	Relay Contact Data	READ_CONTACTS and INTERNET
2	Relay SMS Messages	READ_SMS and INTERNET
3	Send SMS Messages	SEND_SMS
4	Make Phone Calls	CALL_PHONE
5	Make Phone Calls	CALL_PRIVILEGED
6	Covert Listening Device	RECEIVE_BOOT_COMPLETED, RECORD_AUDIO, and INTERNET
7	Covert Camera Device	RECEIVE_BOOT_COMPLETED, CAMERA, and INTERNET
8	Movement Profile	RECEIVE_BOOT_COMPLETED, INTERNET, and ACCESS_FINE_LOCATION or ACCESS_COARSE_LOCATION
9	Eavesdrop on Phone Calls	RECORD_AUDIO, INTERNET, and READ_PHONE_STATE or PROCESS_OUTGOING_CALLS
10	Relay and Falsify SMS Messages	RECEIVE_SMS, WRITE_SMS, and INTERNET or SEND_SMS
11	Falsify SMS Messages	RECEIVE_SMS and WRITE_SMS
12	Activate Debugging	SET_DEBUG_APP
13	Permanently Disable Device	BRICK

We have derived eight different attack scenarios. Examples are *direct monetization*, *attack mTAN based online banking*, or *manipulation of text messages*. For a complete list please refer to sect. 4.2. In the last step, we combined three pairs of rules:

1. Rules that allow creating *movement profiles*.
2. Rules concerning *manipulating text messages*.
3. Rules concerning *eavesdropping calls*.

The resulting rule set includes 13 rules that are used to determine if an application is to be considered as suspicious because it is potentially harmful for the user. The list of rules is given in Table 1. Please note that rules 8, 9, 11, and 12 have been previously defined in [7].

4.2 Scenarios

The following set of attack scenarios were derived:

Extracting Information. Reading data stored on the device and relay it to an attacker. We focus on attacks that use network connection to relay data. Using text messages to relay data is possible. However, we neglect text messages because they are noticeable by the user (since they may cause costs) and related attacks are complex (data has to be split into separate messages and infrastructure to receive messages is required). Furthermore, we consider the unwanted sending of text messages as a separate attack. We disregard relaying data via *Bluetooth* connections as it depends on physical proximity to the target.

Direct Monetization. Send *premium rate text messages* or making *premium rate phone calls*. Furthermore, it is possible to use text messages to distribute *junk* messages.

Compromise Emergency Call System. Makeing *emergency calls*. Unwanted emergency calls may cause serious punishment (depending an national laws). In addition, attackers could use phones of numerous targets to attack the emergency call system (a *critical infrastructure*) by constantly making short emergency calls.

Covert Surveillance. Employing the microphone or the camera to spy on the user. Moreover, it is possible to utilize the GPS sensor to create a *movement profile*.

Eavesdropping on Phone Calls. Use the microphone and system functions to eavesdrop on phone calls. In addition to eavesdropping on conversations, it is possible to extract information from a phone call. For example, Schlegel *et al.* demonstrate how to extract a PIN or credit card number using a *"sound trojan"* [17].

Attacking mTAN Based Online Banking. Relay or falsify text messages that contain a mTAN (mobile transaction authentication number). An attacker intercepts a mTAN and uses it to authorize an online banking financial transaction from the victim's account.

Manipulating Text Messages. Falsify or forge text messages. In addition to annoying the victim (for example, by rendering messages useless or displaying *Spam*), the ability to falsify or forge text messages can be a serious security threat. Although falsifying or forging a text message can not be used to relay mTAN, it may enable an attacker to intercept important messages. For example, an online banking system may send alerts or confirmation messages that can be intercepted. Furthermore, forged messages can be used for *phishing* attacks to trick the user into relaying the mTAN.

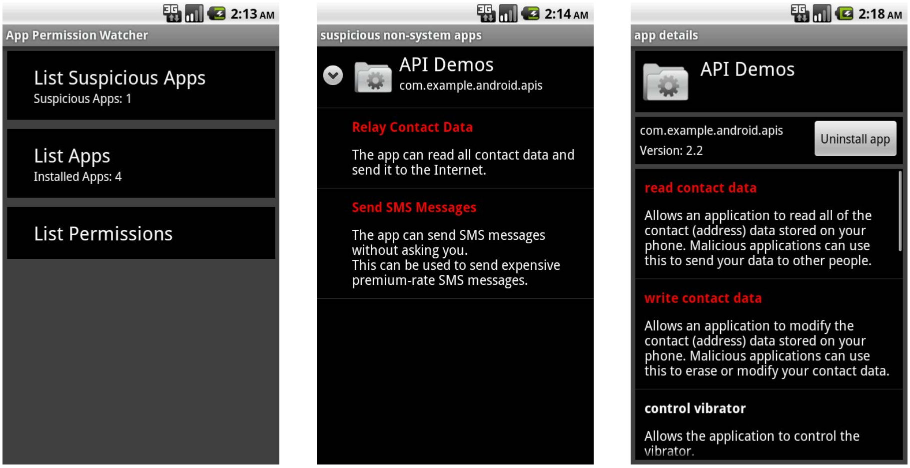
Dangerous System Functions. Android offers a function called *brick* to permanently disable the device. According to a related work by Enck *et al.* , it is possible to gain the permission *set debug app* by manipulating the Android API [8].

4.3 Application User Interface

The design and implementation of the PermissionWatcher user interface follows the basic principle of *details on demand*, to provide the user with only as much pieces of information as necessary. Further, we designed the application to integrate seamlessly in the Android system.

After launching PermissionWatcher (either using the home screen widget or the default launcher), the main screen is presented (see Figure 2(a)). Here the user can select three different options: 1) reviewing which applications are detected as suspicious, 2) reviewing all non- system applications, and 3) reviewing a list of all permissions and which applications are using them.

Figure 2(b)) shows the list with the suspicious applications (in this case only one application is contained) while details regarding rules for responsible marking the application as suspicious are expanded. When a user performs a *long touch*, the application details view is started (see Figure 2(c)). Here the user can uninstall the application by pressing the corresponding button. Further, all details about the permissions are provided.



(a) The main view of PermissionWatcher providing three options: suspicious apps, all apps, and all permissions.

(b) The suspicious applications list shows apps and corresponding permissions.

(c) Application details are summarized in application detail view. Users can uninstall an suspicious apps from here.

Fig. 2. Screen shots of the PermissionWatcher graphical user interface

5 Empirical Usage Evaluation

To gain in depth insights how users make use of the presented PermissionWatcher application, we conducted a field trial in which we collected rich data about usage patterns. To do so, we published the PermissionWatcher application with a built in user data aggregator and advertised the application via different (international) email lists. These list have $\approx 3,000$ subscribers and a partial focus on Germany.

5.1 Data Collection

Data collected by the aggregator included phone status information, *e.g.* how many and which suspicious applications were installed. Moreover, data about the PermissionWatcher usage were recorded. For instance, how often PermissionWatcher was launched, which view (UI/screen) was used, which applications were marked as trusted and which were uninstalled. In addition, the aggregator recorded how often context menus were opened.

Users were informed about the data logging process before installation. In order to ensure that data could not be used for privacy violations all data was fully anonymized and encrypted before sending them to the collection server.

5.2 Results

Within 6 weeks after release, PermissionWatcher was downloaded and installed by over 1,000 users. Most users were from Germany (72%), followed by Austria (8%), USA (7%), Switzerland (3%), and the UK (2%). Other countries (*e.g.* Brasil, China, Japan, US) had a share of less than 1%. In the Android Market users rated the application 40 times with an average rating of 4.7 (1=poor, 5=best).

After removing data logs which were not readable, data logs of 1.036 distinct users remained containing usage data of 3,669 usage sessions. That is, the application was used in 3.54 sessions (Mdn=2) on average per installation. About 400 users used the PermissionWatcher at least three times within the period of data collection. The duration of the sessions was in 60% of the cases less than 60 s (Mdn=33.5 s).

Users of PermissionWatcher had on average 53.6 applications installed on their Android phones (Mdn=43; SD=45.6). Among these, 10.31 (Mdn=8; SD=9.6) were identified as suspicious by PermissionWatcher which corresponds to 20% of all applications installed on the users' phones. The top five list of suspicious apps installed (number of installations) is shown in following table 2.

Table 2. The top five of most frequently installed applications that were identified as suspicious

Application Name	# Installations
Whatsapp	476
Google translate	328
Barcode scanner	309
Facebook	292
Skype	239

The 25 most frequently installed apps that were identified as suspicious were each installed on at least 10% of the users. The most frequent reason for tagging an application as suspicious was permission to read and relay the address book information (READ_CONTACTS and INTERNET).

We found that 94 (9.1%) of the users uninstalled at least one application using PermissionWatcher. On average 3.1 applications were uninstalled (Mdn=2).

Three users uninstalled more than 10 applications. In total 247 different applications were deleted, whereas only 26 were deleted by multiple users. The top five list is shown in table 3. Note that each application was uninstalled in four cases.

Note: In our initial survey, 44% of all users indicated that they would deinstall applications if they were warned. However, it is well known that people tend to give "socially expected" answers. In addition, our initial survey was even more biased towards security professionals then the field trial; this could also explain a part of the gap.

Table 3. The top five list of most frequently uninstalled applications using PermissionWatcher

Application Name	# Uninstallation
Compass	4
Ebay	4
Barcode scanner	4
HRS hotel search	4
Skype	4

In more than half of the cases, users were reviewing the list of *suspicious apps* before uninstalling applications (55%). In 29% of the cases the list *all apps* preceded the uninstallation and in 16% the list of *all permissions* was reviewed.

Users had the opportunity to mark suspicious applications as trusted, which removed them from the corresponding list. Only 47 users marked applications as trustworthy (137 in total) which were previously detected through the system. Each of these users marked 5.7 (Mdn=3; SD=6.3) applications as trusted. The top five list applications that were marked as trusted (# cases) using PermissionWatcher is listed in table 4.

Table 4. The top five list of applications that were marked as trusted

Application Name	# Trusted
Whatsapp	11
Skype	10
DB Navigator	9
Google+	8
AVG Antivir	7

Turning towards which features of the application were chosen, the overview list *suspicious apps* was used by 92% of the users. That is, each user opened this list on average 3.2 times (Mdn=2). On average, the usage duration was 60 s (Mdn=19 s; SD=34.4). The overview list *all apps* was used by 45% of the user and was opened 1,077 times. That is, on average each user used this list for 1.9 times (Mdn=1, SD=2.4). In this case, users spend on average 54 s (Mdn=13.1 s; SD=436.9). The list giving an overview of *all permissions* was used by 70% of the users. In total 1,171 times, which is on average 1.6 times per user (Mdn=1; SD=1.5). Users spend on average 83 s using this list (Mdn=32 s; SD=274.7). Surprisingly, only 16% of the users viewed the *application details* overview. In total 513 times were application details reviewed which corresponds to 3.1 times on average (Mdn=2) per user. On average, when reviewing the application details, users spend 16.8 s (Mdn=9.3). Figure 3 gives an overview of how many percent of the users applied which data view of PermissionWatcher. This correlates to our finding that users prefer easy to understand presentation rather than the full information. On the other hand, we think that this feature gave credibility to PermissionWatcher.

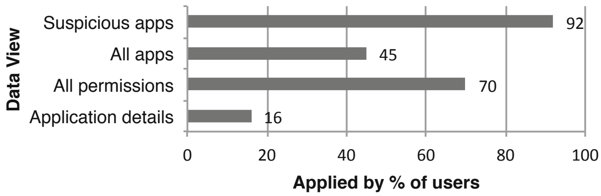


Fig. 3. Usage of different features of PermissionWatcher

In summary, we observed over 2,200 installations within a time period of six weeks. We collected data from over 1,000 users. It appeared that only free applications were considered as suspicious. Many of them were listed in the Android Market top application lists. The main reasons for identifying applications as suspicious are access to the address book or access to GPS sensor data. Interestingly, lists of deleted and trusted applications are overlapping. It is noteworthy that 18 applications that were uninstalled by users were removed later on from the Android Market due to unknown reasons.

6 Conclusions and Future Work

Our user study has indicated clearly that users are actually *willing* to use secure applications as long as the information is presented in an easy understandably way. The most prominent example is the right *full internet access*, which only 63% of all participants in our initial survey could give (*cf.* sect. 3). This points to a serious security problem: if users cannot understand the permissions they grant they are likely to allow Trojan software directly through their front door by clicking at dangerous permission sets. On the other hand, our initial survey also indicated that users *wanted* security on their phones—as long as it was presented in a clear and understandable way.

Therefore, we have developed PermissionWatcher especially for users without a technical and security background. Our aim was to put permission based systems on a stable footing by informing users about dubious permission sets. Results of a field study with 1000+ users indicate that a considerable amount of users carefully examined *suspicious* apps and even uninstalled them. Due to the field study design, it was not possible to contact users and interview them which were reasons for uninstalling applications. Even more interesting, what were reasons for keeping applications that are clearly requiring too many permissions. This needs to be investigated in followup users studies.

There are several ways to extend the concept of creating awareness after applications are installed. First, it would be beneficial if the user would be informed *before* installation about suspicious applications. Preferably, the user should be given alternatives, such as "This torch light app needs 124 permissions, including 15 dangerous ones. Alternatively, we have found a torch light application, that only needs 2 permissions, none of them being dangerous. Which one would you like?"

PermissionWatcher does not use the "wisdom of the crowd" yet. In particular, when choosing alternative applications or identifying dangerous permission sets, users could rely on (indirect) input from others. Judging from our previous experience, this needs to be done in a transparent, and in particular none-interrupting way.

Finally, it would be highly beneficial for the security of the Android Marketplace if alternative suggestions and highlighting problematic permissions were *directly* integrated into the Marketplace itself: instead of *pretending* to offer the user security by showing a mostly unintelligible heap of permissions, it needs to be far more user friendly. Otherwise, large portions of Android users will easily fall prey to attacks by seemingly innocent permissions. On the up-side: The presentation of all permissions needed for one app made it possible for us to identify dangerous permission sets quickly. So the Android market place is certainly on the right track—but would need to take this little extra step to be beneficial to *all* users.

Acknowledgements. This work has been conducted within the Emmy Noether research groups Long Term Security and Mobile Interaction with Pervasive User Interfaces, both funded by the German Research Foundation.

References

1. Amer, T.S., Maris, J.B.: Signal words and signal icons in application control and information technology exception messages – hazard matching and habituation effects. Working Paper Series 06-05, Norther Arizona University (2006)
2. Anderson, J., Bonneau, J., Stajano, F.: Inglorious installers: Security in the application marketplace (2010)
3. Barrera, D., Kayacik, H.G., van Oorschot, P.C., Somayaji, A.: A methodology for empirical analysis of permission-based security models and its application to android. In: CCS 2010, pp. 73–84. ACM (2010)
4. Becher, M., Freiling, F.C., Hoffmann, J., Holz, T., Uellenbeck, S., Wolf, C.: Mobile security catching up? revealing the nuts and bolts of the security of mobile devices. In: SP 2011, pp. 96–111. IEEE, Washington, DC (2011)

5. Egelman, S., Cranor, L.F., Hong, J.I.: You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In: CHI 2008, pp. 1065–1074. ACM (2008)
6. Enck, W., Octeau, D., McDaniel, P., Chaudhuri, S.: A study of android application security. In: 20th USENIX Security Symposium (2011)
7. Enck, W., Ongtang, M., McDaniel, P.: Mitigating android software misuse before it happens. Technical report, Pennsylvania State University (2008)
8. Enck, W., Ongtang, M., McDaniel, P.: On lightweight mobile phone application certification. In: CCS 2009, pp. 235–245. ACM, New York (2009)
9. Enck, W., Ongtang, M., McDaniel, P.: Understanding android security. IEEE Security and Privacy 7, 50–57 (2009)
10. Felt, A.P., Chin, E., Hanna, S., Song, D., Wagner, D.: Android permissions demystified. In: CCS 2011, pp. 627–638. ACM (2011)
11. Felt, A.P., Greenwood, K., Wagner, D.: The effectiveness of application permissions. In: WebApps 2011, p. 7. USENIX Association, Berkeley (2011)
12. Felt, A.P., Ha, E., Egelman, S., Haney, A., Chin, E., Wagner, D.: Android permissions: User attention, comprehension, and behavior. Technical Report UCB/EECS-2012-26, University of California at Berkeley (2012)
13. Google Inc. Security and permissions (2011), http://developer.android.com/guide/topics/security/security.html (last access June 20, 2012)
14. Hogben, G., Dekker, M.: Smartphones: Information security risks, opportunities and recommendations for users. Technical report, ENISA (2010)
15. Li, B., Im, E.G.: Smartphone, promising battlefield for hackers. Journal of Security Engineering 8, 89–110 (2011) ISSN : 1738-7531
16. Microsoft Corporation. Microsoft security intelligence report, vol. 11 (2011), http://www.microsoft.com/security/sir/default.aspx (last access June 20, 2012)
17. Schlegel, R., Zhang, K., Zhou, X., Intwala, M., Kapadia, A., Wang, X.: Soundcomber: A stealthy and context-aware sound trojan for smartphones. In: NDSS 2011, San Diego, CA, pp. 17–33 (February 2011)
18. Shin, W., Kiyomoto, S., Fukushima, K., Tanaka, T.: Towards formal analysis of the permission-based security model for android. In: ICWMC 2009, pp. 87–92. IEEE, Washington, DC (2009)
19. Shin, W., Kiyomoto, S., Fukushima, K., Tanaka, T.: A formal model to analyze the permission authorization and enforcement in the android framework. In: SOCIALCOM 2010, pp. 944–951. IEEE, Washington, DC (2010)

Exploring Non-verbal Communication of Presence between Young Children and Their Parents through the Embodied Teddy Bear

Kaisa Väänänen-Vainio-Mattila[1], Tomi Haustola[1], Jonna Häkkilä[2], Minna Karukka[3], and Katja Kytökorpi[3]

[1] Tampere University of Technology, Korkeakoulunkatu 6, 33720 Tampere, Finland
[2] Center for Internet Excellence, P.O. Box 1001, 90014 University of Oulu, Finland
[3] Nokia Research Center, Oulu, Finland
{kaisa.vaananen-vainio-mattila,tomi.haustola}@tut.fi,
{jonna.hakkila,minna.karukka,katja.kytokorpi}@gmail.com

Abstract. Young children are emotionally dependant on their parents. Sometimes they have to be apart from each other, for example, when a parent is travelling. Current communication technologies are not optimal for supporting the feeling of presence. Our goal was to explore the design space for remote communication between young children (4-6 years) and their parents. More specifically, we aimed at gaining user feedback to a variety of non-verbal interaction modalities using augmented everyday objects. We developed the Teddy Bear concept and created an embodied mock-up that enables remote hugging based on vibration, presence indication, and communication of gestures. We conducted a user study with eight children and their parents. Our qualitative findings show that both children and parents appreciated Teddy Bear for its non-verbal communication features, but that some aspects were not easily understood, such as gestures for strong emotions. Based on our findings, we propose design implications for mediated presence between young children and their parents.

Keywords: Child-parent communication, presence, interaction modalities, embodied interaction, augmented everyday objects, concept design, user study.

1 Introduction

Communication is a fundamental part of human life, and the ability to communicate remotely within families has become essential due to mobile and often hectic life-styles. This is especially significant with young children, whose communication needs are often bound to emotion, embodied interaction, and a feeling of care from their parents. Existing commercialized communication technology is designed primarily for grown-up users. The special needs of young children communicating with their parents over technology have not yet been largely addressed in the existing solutions. Mobile phones and PCs support talking and sending of pictures and video, for example through Skype. Still, these media lack specific means to mediate presence and

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 81–96, 2012.

subtle or spontaneous emotional communication. Using people's senses and conveying bodily expressions in a richer way than what is possible with the current tools could offer novel, pleasant ways to be in touch with the loved ones who are away.

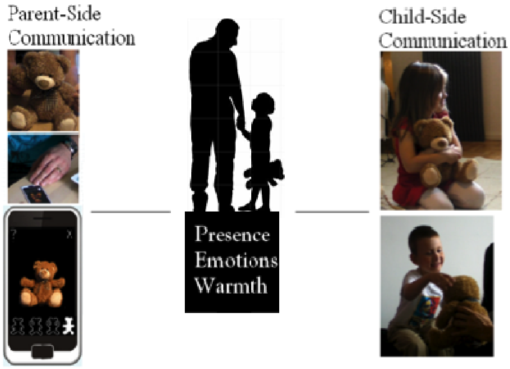


Fig. 1. The core concept of the Teddy Bear, a smart, embodied everyday object for communication of care and presence between young children and their parents

When studying children as users of technology, the four developmental stages described by Jean Piaget in his child development theory can be considered: 1) sensorimotoric (ages 0-2), 2) preoperational (ages 2-6), 3) concrete operational (ages 6-12) and 4) formal operational (12-adulthood) [16]. According to Chen et al. [7], children become mobile phone users during their concrete operational stage. For younger children, i.e. in the preoperational stage, children become increasingly adept at using symbols, and increasingly use pretence in play. For example, a child is able to use an object to represent something else. Skills of understanding symbolic representations open opportunities for new types of communication tools.

Non-verbal interaction plays an important role in communicating with young children [7]. As they have undeveloped linguistic skills, small children naturally express themselves by actions. Non-verbal gestures, movement, and play are therefore more suitable means for young children in using technology-based communication media [2]. Embodied interaction which combines tangible and social interaction [9] has been recognized as a potential design direction for parent-child interaction [22].

In this study, our goal was to explore alternative ways to support the feelings of presence between children aged four to six years and their remote parents. By presence, defined as "the subjective experience of being there" ([4], p. 2), we mean *the feeling or experience of the other person being in connection with oneself* – in this case the connection between the child and the parent. Our focus was to explore subtle communication through various embodied interaction modalities and related augmented everyday objects. Our research is based on Research through Design approach [28]. Our study consisted of the concept design and a two-phase user study of an augmented Teddy Bear mock-up for child-parent communication (see Figure 1). We suggest design implications based on our findings. This work contributes to the understanding of the user needs for mediated child-parent communication and the design of smart devices supporting presence between these parties.

2 Related Research

Communication and keeping in touch with family members is a fundamental issue in people's everyday lives. Previous research has investigated concepts that could enhance communication between parents and children. Wayve [14] is a situated messaging device for the home that incorporates handwriting and photography. Wayve was studied in households for three months and the findings emphasise the role of play in social family relationships. Whereabouts Clock [4] focused on presenting family members' approximate location and activities to enhance the feeling of presence. FAMEX [9] was developed to support family members, including teenaged children, to create a common family history communicated face-to-face. Interactive Carpet prototype [24] was designed to support family connectedness and presence in a subtle way. The emphasis in these systems is on family presence and cooperation but not specifically on remote communication between young children and parents.

Several studies of family communication systems have focused on young children. In their research about video play, Follmer et al. [9] strive for creating the feeling of togetherness through a shared play in a video call. By exploring different mechanisms for playfulness and engagement, their iterative design moved towards book reading as a scaffolding for a joined, engaging activity for parents and children. A portable prototype, Story Play, developed around the concept is presented by Raffle et al. [18]. The prototype was targeted especially for small children (2-3 years), and focused on supporting family reading rituals such as reading in the bed in the evening.

Using wearable technology and embodied interaction for emotional family communication is demonstrated with Huggy Pajama [23], through which parents can hug their children remotely. Mobile input for this hug system allows parents to express their emotions. Dalsgaard et al. [7] studied mediated intimacy between parents and children through cultural probes and contextual interviews. Physical interaction as a token of intimacy was one important factor that was highlighted in the findings.

Tangible and embodied interfaces provide a good potential for communication and collaboration technologies targeted at children. Antle [1] studies tangible systems based on an understanding of why and how tangible interaction can support cognitive development of children. She concludes that successful tangible systems will incorporate adaptive, body-based styles of interaction which support children's developing and existing repertoire of physically-based actions. Interaction will be learned through exploration with real-time feedback of how things work. Furthermore, tangible systems might be well suited to help children communicate abstract themes.

Toys provide easily approachable and tangible interfaces that children are accustomed to, and their effective use for computing and communications has been demonstrated in earlier research. Bonanni et al. [2] developed PlayPals which are wireless dolls intended to be used for playful remote communication, sharing of multimedia experiences, and virtual co-presence between children aged 5-8 years. Their user study reveals that embedding digital communication into existing play patterns and objects enhances both remote play and communication. Bruikman et al. [4] developed a system called Lali which allows very young children (9-42 months) to play with each other using plush toys which can be manipulated from both ends. Freed et al. [7]

studied remote communication between children with a tangible system – a dollhouse – containing small-scale interfaces for the dolls with a variety of multimodal communication functions. Based on their study of 5-12 years old children, they point out that the toy perspective and manipulable toy elements are particularly helpful in supporting play and successful use of communication technologies.

Raffle et al. [18] present a toy-like design form factor was used in a networked device Pokaboo, and the experiences collected from in situ use trial revealed that the design was perceived compelling and attractive. With Pokaboo devices, a child could press a button and make his/her self-portrait to appear in another, corresponding device. The child's communication partner (or rather, playmate) could response by pressing a button, which caused the button in another device to pop up. Yonezawa and Mase [26] introduce a stuffed toy interface that translates a user's interaction with the toy to musical output in face-to-face communication situations.

Shimizu et al. [18] presented a robotic interface called RobotPHONE based on a teddy bear, intended for interpersonal, synchronous communication. When the child moves the teddy bear's hand, the other person's teddy bear mimicked the movements, conveying motions to the other side. However, no user study is presented for this concept. Toy-like UIs have been explored also by Marti and Schmandt [16], where toy animals were used as conversational agents (to alert of incoming calls, for example). However, the concept presented was developed for adult use, not for children.

Modlitba and Schmandt [16] present an interview study of parents and their children aged 4-10 years. They found several experiences which may take place when parents are traveling, such as separation anxiety, parents' guilt while being away, need to maintain family routines and the wish to communicate at least once a day. Furthermore, they found that the communications needs of parents and children are asymmetric. Whereas children are more concerned of the initial moment of separation, parents experience a continuous feeling of separation anxiety, and want to stay in touch with their children on a more continuous basis. Modlitba and Schmandt suggest designing a system that, based on affective objects that the child cares for, encourages the child to interact with the remote parent. In addition, the system should increase the overall opportunity for synchronous communication, for example by alerts, allow the parent to know when the child is receptive to communication, and make it easier for the child to initiate interaction.

Our research explores the use of tangible devices and social interaction, i.e. embodied interaction [9] for the communication of presence between young children and their parents. After the initial literature research and brainstorming phase, a familiar, tangible toy form factor, teddy bear, was chosen for its rather universal and common presence in families with young children, and by its nature as a warm and emotional character. Unlike Shimizu et al. [18], who focus on the implementation and do not present any user studies or evaluations, we chose to focus on exploring and refining the communication concept with a user study with children and parents. In our study we furthermore explore various communication modalities by which presence could be communicated in a remote parent-child situation. Our research extends the related work by addressing the younger target group than most of the previous studies and by the focus on the non-verbal interaction modalities for communication of presence.

3 The Teddy Bear Concept Design

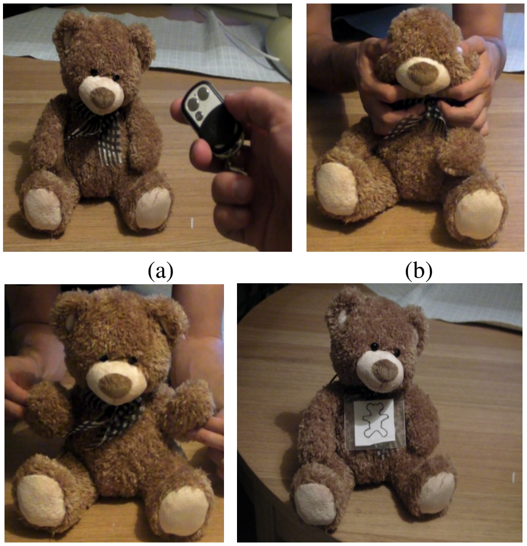
Our research approach was Research through Design (Zimmerman et al. 2007). The literature review was followed up by the concept design phase. The main findings from the literature review were used as design drivers for concepting. These included children's need to initiate communication [16], the potential of various non-verbal communication modalities to support embodied interaction [1], [22], the application of playful, affective interaction objects [18], [26] and the opportunity to utilize familiar, tangible and toy-like objects in the home environment [2], [4], [18].

The concept development began as a day-long brainstorming workshop (see Figure 2) with four participants (three females and one male) who were user experience researchers and designers with 3-12 years of professional experience. In this session, the *Random Words* creativity method was used: categories were *situation* (for example, at home or in a car), *interaction modality* (for example, speech, gestures or light), and targeted type of *experience* (such as caress or excitement). Different combinations of categories were randomly picked up to inspire creation of concepts with different forms of devices, usage scenarios, and modalities to communicate presence.

Fig. 2. A brainstorming session

In the brainstorming, several concepts were created for remote child-parent communication. For example, *Flashlight* allows the child to create light patterns with a hand-held device which makes a similar device in a remote location to shoot out the same light patterns. In *Nurture Swing*, a child can go inside a big cradle or swing which includes an audio and/or video system, and the child can be in communication with the parent. With *Travelling Screen*, a small display that provides information about the parent's travel location, the child can look at pictures that the parent has taken along the travel. The *Teddy Bear* concept was chosen for further development because of its familiarity to young children and for its inherent nature to act as an embodied object for emotion and care. It also seemed to offer a high potential for multimodal, embodied interaction techniques and communication feature variations.

In the Teddy Bear concept, the child has a bear and a parent has either a bear or a mobile application representing the bear. Each Teddy Bear reflects the status and actions of the other party. The core features of the Teddy Bear are sending of hugging and indicator of availability. In addition, other features extend the communication modalities (see Figure 3 and Table 1).



(a) (b)

(c) (d)

Fig. 3. (a) The Teddy Bear core concept has a hugging feature mocked-up by a remotely operated vibrator inside the bear. (b) The core concept also has an mocked-up indicator for availability/non-availability: Ears down and hands on eyes means that the parent (or the child) is not available for communication. (c) Example of an emotional gesture for "happy" (Variation 1). (d) Variations 2 and 3 have a necklace with a display to show symbols or the video stream.

The hugging feature was intended to enable communication of the feeling of presence and care. If the parent hugs their bear, the child's bear vibrates and vice versa. Communication always takes place synchronously, and thus the presence of the remote party is felt by both users. The bear also uses an indicator of the availability of the other person: When the bear is ready for communication, its eyes are open and ears up, and when the bear is not ready for communication, it has its hands on eyes and ears down.

As our goal was also to explore suitable interaction modalities for the communication of presence, a rich variety of communication features using different input and output modalities were designed. Table 1 presents the variations of the concept and how the features are supported with different interaction modalities.

The variations 1-4 are not mutually exclusive but they can extend the core concept in any combination. Even though speech was not the focus of the study it was included in the first user study round to gain users' feedback of its necessity. However, the non-verbal modalities were primary in the set of concept variations.

Table 1. The core concept of Teddy Bear and its 4 variations

Variation	*Features and modalities*
Teddy Bear core concept	Hugging initiated by the user and presented as vibration of the Teddy Bear to the remote party (Figure 1a). An indicator of availability shown with eyes and ears (Fig. 3b).
Variation 1: Gestures to enhance presence and enrich communication	Gestures in communication. The remote person's bear imitates the movements that the user makes using the bear's arms (see Fig. 3c).
Variation 2: Display to enhance communication of emotions	Bear wears a necklace with a display. On the display simple symbols (e.g. a heart, smileys) sent by the other user can be shown (Fig. 3d).
Variation 3: Video stream to enhance presence	Video stream from the child's bear's eyes is shown to parents on their mobile phones or display on bear. The child's bear can also show the parent's bear or the mobile device. (Fig. 3d)
Variation 4: Speech to support explicit (verbal) communication	Phone-like communication through the bear; talking with the parent.

4 User Study in the Family Homes

Appropriate research setups and evaluation methods play high role in child-computer research [13]. For example, the familiar test setting and presence of a friend help children feel at ease [14]. The study context should also represent an area where children are likely to encounter the evaluated product (ibid). Children are typically enthusiastic and playful [16]. Generally children also enjoy situations in which adults observe and give them attention, and even give them rewards [14]. Deriving from this background, we decided to run user study sessions at children's homes, and have their parents present during the sessions. The main aim of the user study was to gain feedback about the Teddy Bear concept and its interaction modalities from the target user group of 4-6 years old children and their parents.

There were two rounds in the user study. In the first round, initial feedback related to the core concept and its variations was gathered to determine whether the concept was generally acceptable, and to identify the modalities of Teddy Bear interaction that should be focused on. We also explored how different gestures of the bear were interpreted, and whether the parents would prefer their own physical Teddy Bear or a mobile device as a user interface (UI) to a virtual Teddy Bear.

In the second round, Variation 1 of the Teddy Bear was selected for gathering of more feedback. In addition, a paper prototype of the touch-based mobile phone UI was created for the parents to determine how they would interact with the bear and how they recognized the gestures and icons presented in the mobile UI.

The ages of the child participants are shown in Table 2. In Round 1, there were twins in one family (a girl and a boy, 5 years). The present parents were four mothers and four fathers.

Table 2. Child participants of the two-round user study

	Gender	*Age*		*Gender*	*Age*
Round 1	Girl	4	**Round 2**	Boy	4
	Girl	5		Boy	6
	Boy	5		Girl	4
	Girl	6		Boy	4

4.1 The Study Setup

All study sessions were conducted in the homes of the children (N=8), in their own room or in a separate playroom. In the sessions, one parent of the child, the child and the researcher were in the room. The parent was in the room to ensure the comfort of the child and also to support the communication between the child and the researcher (see Figure 4, left). The scenario was explained wherein a parent would be absent and the child would like to communicate with that parent. Before the session, the parent was asked to identify the usual situation when a parent is away, and explain how they usually communicate in that situation. This information was used in the scenario. The researcher was always at the same (or lower) height level than the child during the tests to avoid communicating status differences and to create a good rapport (Figure 4, right).

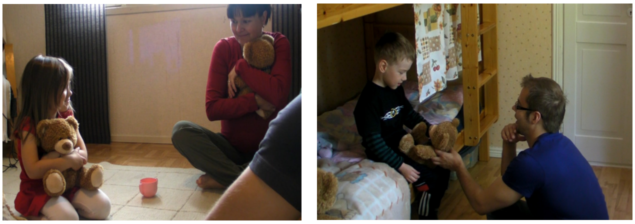


Fig. 4. The user study situation: On the left, the child, parent and researcher; on the right, with the researcher

The study was conducted using the *Wizard of Oz* method. For the hugging feature, the Teddy Bear mock-up had a vibrating mechanism inside it which was controlled by the moderator with a remote controller (see Figure 3 a). The other features were mocked up by the researcher mimicking Teddy Bear's gestures (Variation 1) and speech (Variation 4) and by a paper prototype hung in the neck of the Teddy Bear (Variations 2 and 3).

In round one of the user study, the sessions included the following steps:

Step 1. Explanation of the scenario and the Teddy Bear concept with its hugging feature. This was done by giving a parent one bear and another one to the child. When the parent hugged their bear, the vibration of the child's bear was activated with a remote controller. Reactions towards the bear, hugging, and vibration were observed.

Step 2. Feedback about the gestures (Variation 1). The gesture feature was explained and demonstrated with two bears. Then, the gestures "hi", "hug me", "happy", "angry" and "sleeping" (not available) were shown to the child and the child was asked: "What if bear does this, what would it mean?" After demonstrations of different gestures, the child was asked to show how they would gesture with Teddy Bear.

Step 3. Feedback about the symbolic display (Variation 2). Different paper prototype pictures – symbols of "a heart", "a greeting teddy bear" and "a hugging teddy bear" – were then put on the display and the child was asked how he or she would interpret them.

Step 4. Feedback about the video stream (Variation 3). An explanation was given that the bear's eyes allowed the parent to see what child was doing. Then, the child was asked which things they would like to show to the parent with the Teddy Bear.

Step 5. Discussion about the possibility of speaking with the parent via the Teddy Bear (Variation 4). The children and parents were asked if they would like to use a phone or the bear for speaking.

Step 6. Comparison questions about the variations: "Did you prefer this one or this one?" and "Which one was the best?" Even though a strict, measurable comparison was not the aim of this task, we wanted to get an impression of which interaction modalities the children would prefer.

Variation 1 was chosen for a further study in round two, as the children preferred this over other variations. (See below for the results of the first round.) The Steps 1 and 2 of round two were the same as the previous round. Following these, a brief paper prototype test of the mobile UI was done by the parent (see Figure 5). The parent was interviewed to obtain feedback.

4.2 Data Gathering and Analysis

Sessions were recorded with a video camera for analysis of participants' comments and reactions. Children's reactions were divided into positive (smiles, laughs, and excitement), negative (negative faces, ignoring, attention going elsewhere) and neutral (showing interest, but no strong reactions). With some children, the positive reactions were clear (laughter, smiling, and animated speech), whereas with others there were fewer perceivable reactions. The qualitative data analysis was performed by transcribing the comments and reactions, and by thematically grouping those items (written in altogether 130 notes) to major themes on an affinity wall. Two researchers conducted the analysis.

5 Findings of the User Study

5.1 The First Round with the Teddy Bear Concept and Its Variations

Hugging with vibration feedback (the core concept) received positive feedback from every child and parent. The hugging feature and gesture communication of the bear (Variation 1) were preferred over other features (such as communication with speech or streamed video and pictures). Speech was seen as effective for communication, but the participants – both children and parents – preferred using phones to talking via the Teddy Bear. As video communication was readily available via Skype, the parents commented that the novel, non-verbal communication modalities were perceived as the best for Teddy Bear.

The gestures of the Teddy Bear were interpreted in various ways. For example, children would recognize Teddy Bear's gesture of both hands up (intended as "happy") as "doing gymnastics", "it's happy", or "bear wants to move hands up." Movement was important for interpretation of the meaning of the gesture. The non-emotional gestures such as "sleeping" and "hi", as well as the "hug me" gesture were recognized more easily than stronger emotions of "happy" or "angry". When asked what the bear does when it is doing the gesture for not being available (hands on eyes, ears down), a six year old girl responded: *"it's sleeping."*

To design better gestures for the second round, the children were asked to show how they would express "hi", "happy", "angry", and "hug me" with the bear. "Hi" was gestured with either a sideways movement or an up and down movement with Teddy Bear's hand. For "happy", gestures from the children were mostly playing (throwing Teddy Bear into the air) or raising both hands up. Often the child used voice to express the message, rather than the pure gesture (saying, for example, *"yay!"* when expressing happiness). The expression for "hug me" did not differ from the originally planned gesture of rounded arms. The gesture for "angry" was difficult for the children: For one child, the Teddy Bear's arms crossed represented anger; for another, anger was expressed with the bear's hands on the hips.

Feedback about the symbols on the display and the contents of the video stream to show to the parent were very hard to obtain. Some children were able to recognize the symbols but could not verbalize what they could mean in the child-parent communication situation. The variation where the child could show video stream to the parent was clearly too difficult to explain by this age group.

A positive finding for this concept was that when the bear was given to the children they usually wanted to hold it all the time. Two of the children kept on hugging the Teddy Bear and said that they wanted to keep it. *"(Laughing) it felt nice (laughing)"*, girl, 6, while hugging the Teddy Bear. The embodied interaction with the tangible, soft toy provided positive experiences for the young children.

Three of the four parents preferred to use a mobile phone application over having their own Teddy Bear. They said this is due to the availability of the mobile phone at all times and the expected inconvenience of carrying a toy around while travelling.

Based on the results of the first study round, Variation 1 (with vibration feedback and gestures) was chosen to be studied in the round two. Also, a simple mobile UI paper prototype was developed for the second test round as parents preferred a mobile UI over carrying their own bear.

5.2 The Second Study Round with Variation 1 and the Mobile UI for Parents

5.2.1 Evaluation of Teddy Bear with the Children

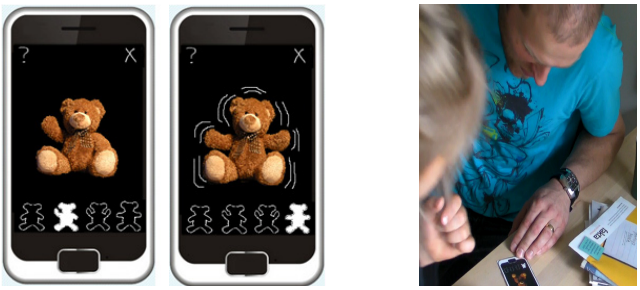
Reactions to Teddy Bear (Variation 1) were in general more neutral in the second round than in the first round, although children still seemed to like the idea. "*What is this doing, this bear is funny*", boy, 4, while hugging Teddy and it was vibrating back).

The second round of the study revealed more clearly a need for developing the gestures of the bear. As in round one, gestures for "hi" and "hug me" were better recognized than emotional gestures. The emotional gesture of "happy" (hands up) was not recognized at all, nor was any clear movement for "happy" suggested by the children. *"It's doing gymnastics"*, girl, 6 years, when asked what the gesture of lifting Teddy's hands up would mean.

5.2.2 Parents' Feedback to the Mobile UI

Whereas some earlier research concepts (e.g. [18], [18]) use special form factors for both communication parties, we provided the parents an alternative user interface through the mobile phone. This was considered as a more realistic solution for grown-up family members to use while travelling e.g. on business trips. This solution is also in line with the earlier finding of the need for asymmetric communication between young children and parents [16], where parents are in more frequent need to feel the presence of their children. The mobile phone is more easily available for the parents in various usage situations.

The mobile UI was based on Variation 1 of the Teddy Bear (see Table 1), focusing on hugging and gesture communication. Hugging feature was based on the vibration feedback. When the child hugs the bear the mobile UI would provide vibration and visual feedback to the parent (see Figures 6a and 5b). The parent answers or initiates hug by clicking the bear on the mobile UI. When the bear in mobile UI is touched the vibration feedback is given to both child's bear and the mobile device. The gestures of the mobile UI bear were controlled by two ways: 1) either by having direct manipulation (holding finger on the bear's hand and moving it) or 2) choosing one of the template gestures activated by clicking the gesture icon (see figure 5a). The template gestures and icons were based on the findings of the first round of the study. In the paper prototype direct manipulation and vibration was only described to the parent.



(a) (b) (c)

Fig. 5. Mobile UI paper prototype and its testing situation

In the paper prototype test (see Figure 5c), the parents intuitively recognized the basic interaction of the mobile UI. Vibration of the phone was easily interpreted by the parents as the child hugging the bear. However, they had difficulty recognizing some of the icons and gestures. Direct manipulation of the bear through touch screen was central to interaction and the parents suggested various ways to interact with the bear (such as petting).

Parents also expressed that they wanted clear status indicators for the bear for opening and closing communication such as awake/sleep. The parents also recognized that the concept would fit into the shared routine moments where presence was important, such as hugging the child at bedtime.

The Teddy Bear should have the general impression of warmth as it should be recognized more as way of communicating than a communication *device*. Parents commented several times that a too technical device with too many functions and buttons in it could turn Teddy Bear too "cold". Thus, simplicity was an important characteristic to aim at with the concept.

Parents who have experiences with using Skype video connection or phone as the communication medium with children brought up that children under four years have difficulties in understanding the communication situation in general. With the Teddy Bear, the main open question of the parents was if the child comprehends the communication and who is he communicating with.

6 Design Implications

In this section we present the main findings as design implications which can be taken as grounds for designing applications and smart devices for child-parent interaction with the aim of supporting presence in situations where the parent is temporarily remotely located. The design implications are based on the qualitative analysis of the findings of both evaluation rounds. The affinity wall (see section 4.2) revealed the main themes (see Figure 6). All themes were connected to how presence is created, when it is important, and how it can enhanced with mediating technology. Based on the found themes, the main design implications are described below.

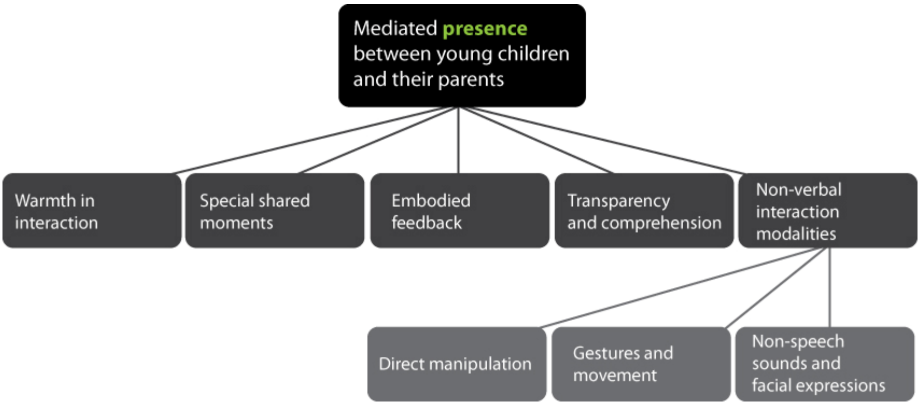


Fig. 6. Themes of the affinity diagram based on the two user study rounds

Warmth Is Essential for Interaction in Mediated Communication between Young Children and Parents. The communication system should not be too technical and it should not have too many functions – otherwise it may become "cold". The emotional side of interaction is important and the device should not be seen as a machine for multimedia messaging but a medium for emotional presence and affective communication.

Support the Special Shared Moments. The system should be a medium for sharing the special moments of presence between the young child and parent. These moments are usually related to basic daily routines in the family such as the moment just before going to bed or sharing experiences or feelings after arriving at home from day care. Simple gestures such as hugging can provide support for these moments. Both children and parents should be able to initiate these interactions.

Provide Embodied Feedback. Getting information about the communication partner's feelings and actions is essential. This is important to both the child and the parent. Symbolic interaction is hard for this age group, and thus tangible means of interacting are a viable option to convey presence. In addition, especially the parent needs to be able to provide embodied feedback to the child, for example hugging or other movement-based responses. Even relatively simple feedback such as vibration can work for these purposes, as long as it provides pleasant alerts of attention.

Support Transparency and Comprehension. Comprehension of the communication situation is critical for the suitability of the communication service or device for the task. With young children the challenge is to make them feel that the parent is in contact with the child. In addition, there should be clear function for starting and ending the communication, as well as availability and unavailability.

Offer Non-verbal Interaction Modalities. When a communication device is intended to support the feeling of presence and care, it may need to be supported by various interaction modalities. The following aspects related to communication modalities may be used as design considerations in the development of systems for supporting presence between young children and their parents:

Direct manipulation. Even though the parents preferred a mobile phone UI over a real Teddy Bear, they wanted to have direct interaction with the child's Teddy Bear. For example, they wanted to be able to pet and poke the bear, thus transferring direct manipulation actions to the child's device, in order to the child to experience the parent's presence. A communication system should support such direct interaction of the other party's device, thus transferring their actions remotely.

The importance of gestures and movement. Movement is an important element for raising curiosity and attention. In a device which takes a shape of a human (such as a doll), an animal, or an imaginative creature, gestures may be used as a central modality communicate presence. They can also be used as comforting elements in communication (such as hugging and waving). Movement helps understanding the meaning of gestures. For better understandability, the children should be able to define the used gestures themselves.

Non-speech sounds and facial expressions. Non-speech sounds and facial expressions are important additional cues in communication. For example, children could send laughter or growling when communicating happiness or anger with the device.

Mediating facial expressions with a teddy-like device could ease the communication of the emotions and increase the sense of presence.

7 Discussion

We have presented the design and evaluation of the Teddy Bear concept, which is a smart object to be used as a communication device between young children (4-6 years) and their remote parents with the aim of supporting the feeling of presence between them. In two-round user study with altogether eight children and their parents, the concept variation which supports hugs and gestures received positive reactions. The children had problems in recognizing some of the gestures, especially for strong emotions, which leads to the consideration of customized gestures. Overall, the non-verbal, embodied interaction modalities with the tangible device were found attractive for the child-parent interaction.

Parents preferred a mobile phone UI with a touch screen as a direct manipulation interface to the child's Teddy Bear. This way the parent has easily accessible means to be in touch with the child. Even though the parents' viewpoint was not the focus of our study, their positive feedback support the results of the interview study of Modlitba and Schmandt [16] which pointed out the parents' need to stay in touch with their children on a more continuous basis.

Communication of presence in a positive and warm manner is very important for young children. If the device is too technical, the positive feeling may be lost. For presence to be communicated in a satisfying way, it is essential that the child comprehends the communication situation and that both parties receive recognizable feedback. Tangibility of the smart device may help in motivating the children to initiate the communication. Availability indicators – such as the Teddy Bear's "sleeping" and "awake" gestures – can be a significant feature to support transparent communication of presence.

The findings which we have presented in form of design implications can provide practical guidance for developers of related tools. In the future, the found design implications could be developed into design guidelines by validating and iterating them in more extensive user studies. Overall, a smart, toy-like communication device has an opportunity to become a mediator of presence in special shared moments of young children and their remote parents.

With regards to ethics considerations of this study, our university rules or country legislation do not require ethics permission for basic HCI studies with children, as long as the parents give their written consent. We attempted to be very considerate with the participating children and for example, asked their parents to be present in the study at all times.

Our study set-up had some limitations. First, the fact that the parent stayed in the same room and the low-fidelity nature of the Teddy Bear mock-up did not enable evaluating the children's actual responses to the presence of a remote parent. Second, the symbolic interaction and video streaming could not be comprehended with the mock-up by the participating children. These limitations could probably be overcome by a more functional prototype and a more realistic study situation. As is pointed out by Antle [1], children slowly develop an understanding of a tangible object and its referent, and a longer-term study would be needed for the evaluation of experiences of presence. Within these limitations, our main aim of exploring the embodied Teddy Bear concept and its non-verbal interaction modalities was well supported by the presented study setup.

As an important step of further work, it would be interesting to run a long-term user study with a fully functional version of the Teddy Bear. The implementation could be based on actual sensor and actuator technologies to support the selected modalities. Alternatively, a prototype could utilize simple add-on technologies for making objects interactive, such as the PINOKY rings presented by Sugiura et al. [22] for animating plush toys. Non-verbal communication with the Teddy Bear could also be developed further to include human-like cues such as facial expressions (e.g. smiles) and sounds (e.g. laughter). It would also be interesting to explore alternative augmented objects for communicating presence remotely.

References

1. Antle, A.N.: Designing Tangibles for Children: What Designers Need to Know. In: Proc. of CHI 2007 EA, pp. 2243–2248. ACM (2007)
2. Ballagas, R., Kaye, J., Ames, M., Go, J., Raffle, H.: Family Communication: Phone Conversations with Children. In: Proc. of IDC 2009, pp. 321–324. ACM (2009)
3. Bonanni, L., Vaucelle, C., Lieberman, J., Zuckerman, O.: PlayPals: Tangible Interfaces for Remote Communication and Play. In: Proc. CHI 2006 EA. ACM (2006)
4. Bostan, B.: Requirements Analysis of Presence: Insights from a RPG Game. In: Computers in Entertainment, vol. 7(1). ACM (February 2009)
5. Brown, B., Taylor, A.S., Izadi, S., Sellen, A., Kaye, J.J., Eardley, R.: Locating Family Values: A Field Trial of the Whereabouts Clock. In: Krumm, J., Abowd, G.D., Seneviratne, A., Strang, T. (eds.) UbiComp 2007. LNCS, vol. 4717, pp. 354–371. Springer, Heidelberg (2007)
6. Bruikman, H., van Drunen, A., Huang, H., Vakili, V.: Lali: Exploring a Tangible Interface for Augmented Play or Preschoolers. In: Proc. of IDC 2009, pp. 174–177. ACM (2009)
7. Chen, Y., Yong, R., Li, L.: Interaction Design of Children's Mobile Phone for Enhancing Cognitive Ability. In: Proc. of ICNC 2010. Sixth International Conference on Natural Computation pp. 2117–2120 (2010)
8. Dalsgaard, T., Skov, M.B., Stougaard, M., Thomassen, B.: Mediated Intimacy in Families: Understanding the Relation between Children and Parents. In: Proc. of IDC, pp. 145–152. ACM (2006)
9. Dourish, P.: Where The Action Is: The Foundations of Embodied Interaction. MIT Press, Cambridge (2004)
10. Jarusriboonchai, P., Väänänen-Vainio-Mattila, K.: Using Mobile Technology to Bring Families Together: The Design of a Family History Concept to Motivate Face-to-Face Communication. International Journal of Mobile HCI 4(2), 1–17 (2012)
11. Follmer, S., Raffle, H., Go, J., Ballagas, R., Ishii, H.: Video Play: Playful Interactions in Video Conferencing for Long-Distance Families with Young Children. In: Proc. of IDC 2010, pp. 49–58. ACM (2010)

12. Freed, N., Burleson, W., Raffle, H., Ballagas, R., Newman, N.: User Interfaces for Tangible Characters: Can Children Connect Remotely through Toy Perspectives? In: Proc. of IDC 2010, pp. 69–78. ACM (2010)
13. Hanna, L., Risden, K., Alexander, K.: Guidelines for usability testing with children. In: Interactions, pp. 9–14. ACM (September 1997)
14. Lindley, S.E., Harper, R., Sellen, A.: Designing a technological playground: a field study of the emergence of play in household messaging. In: Proc. of CHI 2010, pp. 2351–2360. ACM (2010)
15. Markopoulos, P., Read, J., MacFarlane, S., Höysniemi, J.: Evaluating children's interactive products. Principles and practices for interaction designers. Morgan Kaufman, Amsterdam (2008)
16. Marti, S., Schmandt, C.: Physical Embodiments for Mobile Communications Agents. In: Proc. of UIST 2005, pp. 231–240. ACM (2005)
17. Modlitba, P.L., Schmandt, C.: Globetoddler: Designing for Remote Interaction between Preschoolers and their Traveling Parents. In: Proc. of CHI EA 2008, pp. 3057–3062. ACM (2008)
18. Raffle, H., Ballagas, R., Revelle, G.: Family Story Play: Reading with Young Children (and Elmo) Over a Distance. In: Proc. of CHI 2010. ACM (2010)
19. Raffle, H., Mori, K., Ballagas, R., Spasojevic, M., Pokaboo: A Networked Toy for Distance Communication and Play. In: Proc. of IDC 2011, pp. 201–204. ACM (2011)
20. Shimizu, N., Koizumi, N., Sugimoto, M., Nii, H., Sekiguchi, D., Inami, M.: A Teddy-Bear-Based Robotic User Interface. Computers in Entertainment (CIE) 4, 3, 4–14 (2006)
21. Siegler, R.: Children's Thinking. Prentice Hall, Upper Saddle River (1998)
22. Sugiura, Y., et al.: PINOKY: A Ring That Animates Your Plush Toys. In: Proc. of CHI 2012, pp. 725–734. ACM (2012)
23. Teh, J., Cheok, A., Peiris, R., Choi, Y., Thuong, V., Lai, S., Huggy Pajama: A Mobile Parent and Child Hugging Communication System. In: Proc.of IDC 2008, pp. 250–257. ACM (2008)
24. Tharakan, M.J., Sepulveda, J., Thun, W., Cheok, A.D.: Poetic Communication: Interactive Carpet for Subtle Family Communication and Connectedness. In: Keyson, D.V., Maher, M.L., Streitz, N., Cheok, A., Augusto, J.C., Wichert, R., Englebienne, G., Aghajan, H., Kröse, B.J.A. (eds.) AmI 2011. LNCS, vol. 7040, pp. 335–339. Springer, Heidelberg (2011)
25. Yarosh, S., Cuzzort, S., Muller, H., Abowd, G.D.: Developing a Media Space for Remote Synchronous Parent-Child Interaction. In: Proc. IDC 2009, pp. 97–105. ACM (2009)
26. Yonezawa, T., Mase, K.: Musically Expressive Doll in Face-to-Face Communication. In: Proc. of ICMU 2002, pp. 417–422. IEEE (2002)
27. Zaman, B., Abeele, V.V.: Laddering with Young Children in User eXperience Evaluations: Theoretical Groundings and a Practical Case. In: Proc. IDC 2010, pp. 156–165. ACM (2010)
28. Zimmerman, J., Forlizzi, J., Evenson, S.: Research through Design as a Method for Interaction Design Research in HCI. In: Proc. CHI 2007, pp. 493–502. ACM (2007)

Automatic Behavior Understanding in Crisis Response Control Rooms

Joris Ijsselmuiden[1], Ann-Kristin Grosselfinger[1], David Münch[1], Michael Arens[1], and Rainer Stiefelhagen[1,2]

[1] Fraunhofer IOSB, Karlsruhe, Germany
{joris.ijsselmuiden,ann-kristin.grosselfinger, david.muench,michael.arens}@iosb.fraunhofer.de
[2] Karlsruhe Institute of Technology, Karlsruhe, Germany
rainer.stiefelhagen@kit.edu

Abstract. This paper addresses the problem of automatic behavior understanding in smart environments. Automatic behavior understanding is defined as the generation of semantic event descriptions from machine perception. Outputs from available perception modalities can be fused into a world model with a single spatiotemporal reference frame. The fused world model can then be used as input by a reasoning engine that generates semantic event descriptions. We use a newly developed annotation tool to generate hypothetical machine perception outputs instead. The applied reasoning engine is based on fuzzy metric temporal logic (FMTL) and situation graph trees (SGTs), promising and universally applicable tools for automatic behavior understanding. The presented case study is automatic behavior report generation for staff training purposes in crisis response control rooms. Various group formations and interaction patterns are deduced from person tracks, object information, and information about gestures, body pose, and speech activity.

Keywords: automatic behavior understanding, smart environments, rule-based expert systems, fuzzy metric temporal logic, situation graph trees.

1 Introduction

In recent years, there has been great progress in computer vision and other areas of machine perception, for example in person tracking and body pose estimation. However, high-level systems using multiple machine perception modalities and combining multiple objects have not progressed at the same pace. We are developing a toolkit for automatic behavior understanding that deploys multimodal machine perception for multiple objects, fuses everything into a world model with a single spatiotemporal reference frame, and generates semantic descriptions about the observed scene. The current system uses a dedicated annotation tool instead of multimodal machine perception as shown in Figure 1.

The presented case study is situated at the State Fire Service Institute (Institut der Feuerwehr) Nordrhein-Westfalen, during one of their staff exercises

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 97–112, 2012.

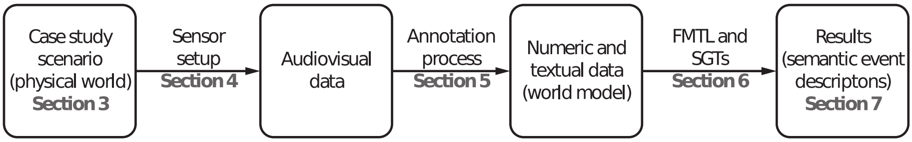


Fig. 1. System overview

for crisis response control room operations (see Figure 2). The task is to automatically generate behavior reports from multimodal machine perception during staff exercises and actual crisis management. These reports about staff behavior in the control room can be used for training purposes, evaluations, and audit trails. For instance, given the identity, position, orientation, and speech activity of the staff members over time, and information about objects in the room, these reports can contain descriptions and visualizations of group formations and interaction patterns, i.e. who was doing what with whom, using which support tools. This can be combined with audiovisual recordings and visualizations, and with the corresponding developments in cyberspace, i.e. field unit status, crisis dynamics, and other context information. Such a system would provide a rich information source, conveniently searchable for specific events.

The presented reasoning process is domain independent because it is separated from any machine perception it might use. The annotation process too is designed to be customizable for other application domains. Possible application domains include other behavior understanding applications, multimedia retrieval, robotics, ambient assisted living, intelligent work environments, intelligent user interfaces, indoor and outdoor surveillance, and situational awareness and decision support for military and civil security. Applied machine perception can range from video to radar, and from person and vessel tracking to body pose estimation, speech recognition, and activities in cyberspace. Other uses include camera control, sensor deployment planning, future event prediction, information exchange between system components, and top-down knowledge for machine perception to guide its search and improve outputs.

This paper is organised as follows. After discussing related work in Section 2, we explain the applied processing chain step by step as depicted in Figure 1. Section 3 describes the case study scenario: automatic behavior report generation for training purposes in crisis response control rooms. A staff exercise was recorded using multiple cameras and microphones with appropriate postprocessing as described in Section 4. The next step is turning the recorded audiovisual data into a world model consisting of numeric and textual data. Ultimately, this should be accomplished using machine perception and multimodal fusion, but we currently use a different approach. Section 5 describes how the postprocessed audiovisual data was manually analysed and annotated using a tool specifically developed for such purposes. The resulting world model forms the input for the reasoning engine based on fuzzy metric temporal logic (FMTL) and situation graph trees (SGTs) presented in Section 6. It delivers semantic descriptions about staff behavior, which can be compared to ground-truth results annotated

Fig. 2. Case study scenario from the State Fire Service Institute (Institut der Feuerwehr) Nordrhein-Westfalen: automatic behavior report generation for training purposes in crisis response control rooms. Several events we aim to recognize are visible here: conversation, discussion with document, and editing a display.

using the annotation tool. Section 7 presents some initial results, Section 8 explains how we can handle imperfect input data, and Section 9 concludes the paper.

The ultimate goal is an integrated system performing all these steps, using multiple machine perception components and multimodal fusion instead of manual annotation. Such a system should run in real time with synchronous visualizations of sensor data, machine perception, and resulting semantic descriptions. In application domains such as robotics and intelligent user interfaces, appropriate embodiment and action generation would be required. Our research is situated in a work environment that focuses on machine perception (especially computer vision) and human-machine interaction, which facilitates the progress toward such an online system. In the meantime, the presented approach improves high-level reasoning processes without the need for corresponding progress in machine perception. And even though behavior reports and visualizations cannot be generated fully automatically yet, our current and future data, observations, and reasoning results can improve understanding of control room operations. The novel contributions of this paper are as follows. Several steps toward an integrated development toolkit were completed, including a new dataset and a new tool for data analysis and annotation. The presented case study is of general interest because of its unique character and its large amount of perception modalities and objects. A newly developed FMTL/SGT knowledge base for this case study is contributed that is also applicable to other domains, along with corresponding experimental results. And we explain how to handle imperfect input data using FMTL and SGTs.

2 Related Work

Surveys on automatic behavior understanding are provided by [1–3]. In [1], a distinction is made between single-layered approaches operating directly on sensor data and hierarchical approaches applying machine perception first and using its output to generate semantic descriptions. In hierarchical systems, semantic event descriptions are usually generated from machine perception outputs using either the statistical approach, the syntactic approach, or the description-based approach. In statistical approaches, event likelihoods are computed by (derivations of) hidden Markov models, (dynamic) Bayesian networks, propagation networks, or similar models [4–6]. In syntactic approaches, atomic events are combined into complex events using formal (stochastic) grammars, mapping spatiotemporal changes in image sequences to events for instance [7–9]. And description-based approaches use formal languages such as logics and and-or graphs for representing and reasoning about spatiotemporal dynamics [10–14]. Statistical and description-based approaches are combined in Markov logic networks by [15–17]. Similarly, Bayesian compositional hierarchies are combined with rule generation from ontologies in [18]. Related studies on smart environments, surveillance, and other applications are found in [19–21]. And [22, 23] present two relevant studies from the field of crisis management.

Our own hierarchical description-based approach to automatic behavior understanding uses fuzzy metric temporal logic (FMTL) combined with situation graph trees (SGTs). In [24, 25], FMTL and SGTs are used to monitor road traffic scenes, [26–29] apply them to human behavior understanding and surveillance, and [30] uses them for intelligent robot control. Preliminary work on the case study presented in this paper is included in [31]. The models we use for representation and reasoning are based on expert knowledge rather than learned from training data. Compared to other approaches, expert-knowledge-based representation and reasoning in FMTL and SGTs is intuitive, convenient, flexible, and easily controllable. The clear boundary between machine perception and reasoning makes it easier to improve one without the other. Furthermore, deductions are understandable by humans and completely provable, and existing rules can be adapted to new settings with relatively little effort. Especially the ability to understand the reasoning process is essential to the presented case study. FMTL/SGT expert systems are suitable for knowledge intensive problems with heterogeneous search spaces such as the one presented here.

3 Case Study Scenario

The presented case study is situated at the State Fire Service Institute (Institut der Feuerwehr) Nordrhein-Westfalen during one of their staff exercises for crisis response control room operations (Figure 2). The exercise is a six hour role playing effort where the participants take on the roles of a full control room staff and others stage the outside world; simulating field units, crisis dynamics, distress calls, and radio communications. The simulated crisis for this exercise was a collision between a passenger train and a cargo train carrying hazardous material. The staff inside the control room is organized as follows. Each first officer is responsible for a functional area: unit management (S1), situation assessment (S2), strategy (S3), and supplies (S4). The first officers answer to the director of operations, and each first officer as well as the director of operations have one or two additional staff members answering to them. Furthermore, there is some supporting staff for maintaining displays (e.g. maps and unit tables), editing documents, and managing incoming and outgoing messages. Several instructors are offering assistance, the director of operations being one of them.

What follows is a description of the typical workflow in such control rooms, corresponding to the recorded data. Once the control room is fully occupied, the director of operations introduces the staff to the current crisis situation. Everybody stops working and returns to their seats to listen. After the introduction, the director of operations tells his staff to continue their preparations and asks his first officers to join him at the table at the central table for strategic planning. When this is done, the director of operations addresses the whole room, announcing that everybody must attend to their tasks until the next briefing. This is when their behavior becomes highly dynamic. Director of operations, first officers, their subordinates, and supporting staff scatter across the room, attending to their displays, documents, and messages. Groups are constantly forming and breaking, and there is a lot of discussion going on. In due time, the director of operations calls the next briefing and everybody returns to their seats. After an introduction by the director of operations, each of the first officers stands in front of the appropriate wall display to give a status report on their own functional area. Everybody listens quietly, except for the director of operations who is occasionally asking the presenter questions, sometimes involving one of the other first officers in the discussion. The director of operations concludes the briefing by summarizing the current action plan and everybody gets back to performing dynamic control room operations.

We aim to model and recognize the different types of person-person interaction and person-object interaction in various group formations. Besides dynamic behavior, we also aim to recognize the more structured events during briefings. The recorded data consists of five briefing / dynamic behavior cycles, each lasting around 70 minutes. The first cycle containing the described introductory phase was analyzed thoroughly and two four minute fragments and two ten minute fragments were selected for the annotation process described in Section 5.

4 Sensor Setup

The staff exercise was recorded with four normal cameras and one fisheye, providing complete and redundant coverage from various angles as shown in Figure 3. Four microphones were installed across the room to provide complete audio coverage. To make the data analysis and annotation easier, we used the raw video data to generate one synchronized five-pane image per second. One of them is shown in Figure 3. A sampling rate of *1fps* is sufficient, because no machine

Fig. 3. Five-pane image showing the cameras' viewing angles. To simplify the annotation process, such images are generated at *1fps* from the raw video data recorded at the State Fire Service Institute (Institut der Feuerwehr) Nordrhein-Westfalen.

perception is performed on the data, and because the events we aim to recognize do not have fast dynamics. Higher sampling rates could be obtained with corresponding annotation effort, or an interpolation algorithm applied to the $1fps$ numeric and textual data (world model). The audio data is used to better understand what is going on in the control room (context information), and to annotate the participants' speech activity.

5 Annotation Process

Two four minute fragments and two ten minute fragments from the first 70 minutes of the exercise were selected for annotation with hypothetical outputs from machine perception and ground-truth for corresponding semantic event descriptions. A PyQt annotation tool was specifically developed for this purpose. Its main component is an interactive birdseye view allowing the user to manipulate the modeled objects. The tool is used to create a birdseye view and underlying XML data (hypothetical machine perception and semantic ground-truth results) for each second of recorded data. This is exemplified by the screenshot in Figure 4, displaying the same data as the five-pane image in Figure 3.

Before the annotation process can begin, the user has to edit an XML stage file using a custom XML scheme, specifying which dynamic objects can be added to the scene; in this case people, notepads, and messages. The stage file specifies their static attributes: name, type, subtype, and size. The file also determines which semantic event types should be recognized, so that the corresponding ground-truth results can be annotated using the provided interaction elements. Event types are specified in terms of their names, arities (number of arguments), and argument domains (allowed object types). Finally, the stage file describes

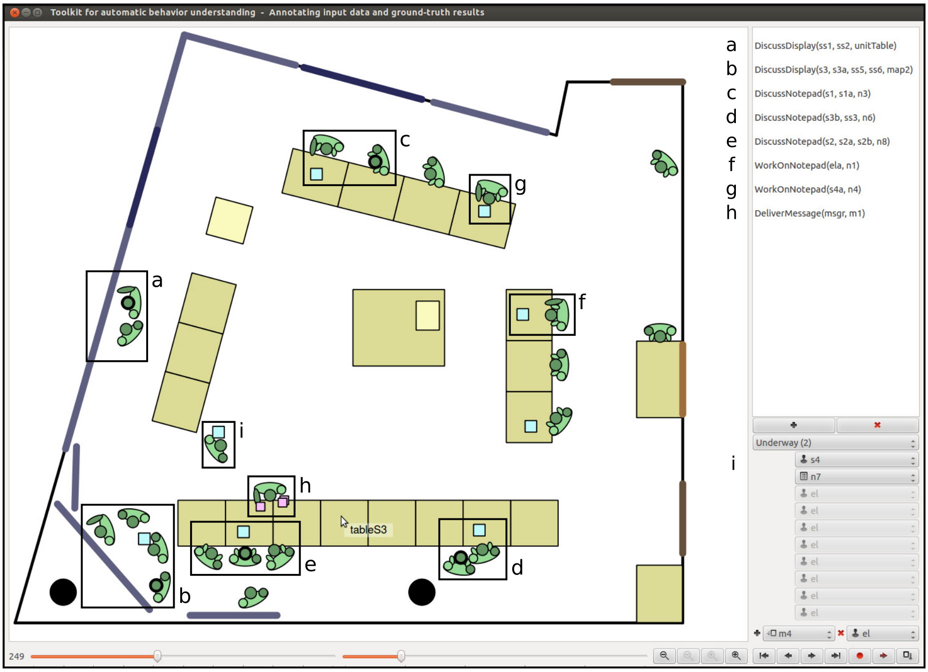


Fig. 4. Tool for annotating audiovisual data with hypothetical machine perception and semantic ground-truth. It provides an interactive birdseye view for manipulating modeled objects. The displayed data and ground-truth results correspond to Figure 3: *a)* two people discussing the field unit status table, *b)* similar *c)* two people discussing a notepad, *d,e)* similar, *f)* person working on a notepad, *g)* similar, *h)* delivering a message, and *i)* underway with notepad.

the static objects in the room in terms of name, type, subtype, location, and size. In this case, the static object types are: wall, table, display, door, hatch, and device. The static objects are visualized as in Figure 4. And the dynamic object specifications (in this case for people, notepads, and messages) and the ground-truth event types to choose from, are used to fill the corresponding interaction elements. Upon loading the stage file, no dynamic objects are present, they are added and removed through user interaction.

The attributes of the dynamic objects are manipulated using mouse interaction. Each person can be moved and rotated, and their body pose, gesture activity, and speech activity can be set. Speech is indicated by a rim around the head, speech-supporting gesticulation by a rim around the right hand. An extended and optionally rotated arm indicates pointing and interaction with displays, notepads, and messages, an extended head indicates looking down, and extended legs indicate sitting (see Figure 4). Notepads and messages can only be moved around. After the dynamic objects have been manipulated to reflect the audiovisual data, semantic event descriptions can be annotated by selecting the required event-argument-combinations (see Figure 4, i). This process is repeated for each second of data, i.e. for each of the images exemplified by Figure 3. The interface includes elements for recording, playing back, and navigating through the data, and data files can be saved to be reloaded at a later time. The resulting XML data contains a description of the recorded dymanic objects and ground-truth events for each second. People, notepads, and messages possess the attributes name, type, subtype (strings), presence (boolean), x-coordinate, y-coordinate, width, and height (integers). In addition, people have orientation, gesture (integers), speech, looking down, and sitting (booleans). The recorded ground-truth events have their name and list of arguments specified in XML.

To further improve the annotation process, still images, video streams, and audio streams should be displayed in sync with the birdseye view visualizing the XML data. Furthermore, results and ground-truth (semantic event descriptions) should be visualized in the birdseye view, and ideally also in the audiovisual data. The bounding boxes in Figure 4 were added manually. Other ideas to improve the annotation tool include a more sophisticated set of data objects, 3D data functionality, and convenience/data-quality improvements through AI and physics laws that operate on the data model. The presented tool can also visualize real machine perception outputs, and it can be customized for any of the application domains described in Section 1.

6 Fuzzy Metric Temporal Logic (FMTL) and Situation Graph Trees (SGTs)

Once input data is available, the actual generation of semantic event descriptions can begin. The XML data from the annotation process and the handwritten XML stage file are fed into a reasoning engine based on fuzzy metric temporal logic (FMTL) and situation graph trees (SGTs) [24–31]. We use F-Limette: a reasoning engine for FMTL written in C, and the SGT-Editor: a Java application for editing and traversing SGTs. The FMTL language is a first order logic extended with fuzzy evaluation and temporal modality. Fuzzy evaluation allows for reasoning about inherently vague concepts such as distance categories (e.g. close, far) as well as reasoning about uncertainty in the input data. The latter will be addressed in Section 8. Temporal modality allows for reasoning about temporal developments using rule conditions grounded in points along the time axis corresponding to past, current, and future states of the world.

Each reasoning process starts at the root node of an SGT, which is then traversed as described in [28]. From each traversed node, FMTL rule conclusions are queried that initiate Prolog-like rule execution processes (i.e. F-Limette uses the logic programming paradigm). Each rule execution process returns a truth value between *0.0* and *1.0* depending on the rule conditions that were directly or indirectly evaluated after querying the rule conclusion in the SGT node, and ultimately on the atomic facts from the input data. The returned truth values are carried down to the next SGT node where they are used as base truth value (instead of *1.0*). Semantic event descriptions with corresponding truth values but also actuator commands can be generated from any SGT node.

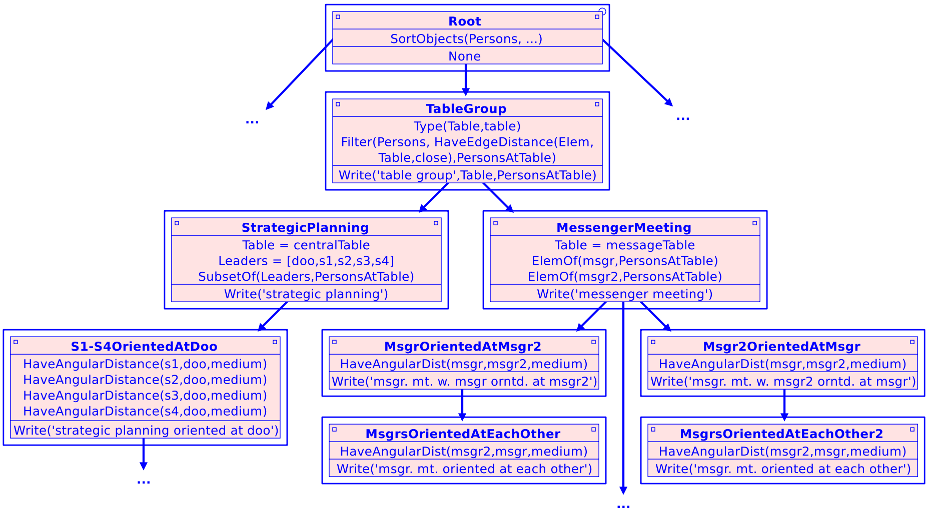


Fig. 5. Part of a situation graph tree (SGT) from the presented case study. It is used to detect groups around tables and conceptual refinements thereof.

SGTs are hypergraphs consisting of situation graphs (see Figure 5). Each situation graph contains one or more situation schemes, and each situation scheme possesses a name, one or more preconditions, and zero or more postconditions (i.e. semantic event descriptions and/or actuator commands). To model temporal dynamics and and events consisting of multiple phases, situation schemes can be interconnected through temporal edges within each situation graph. This feature is not used in Figure 5, as its situation graphs (visualized by thick boxes) contain only one situation scheme each. Its only temporal edge is the reflexive one on the *Root* situation scheme, causing the reasoning process to continue over time. Conceptual refinement is visualized by a thick edge between a situation scheme and a situation graph below it. FMTL rules are largely domain independent and typically about spatiotemporal relations, whereas SGTs are more domain specific as they usually constitute abstract relations between the FMTL rules they deploy. Once an FMTL rule base has been established it stays relatively fixed and it can be used by different SGTs within the same application domain or across different application domains. We now provide a detailed description of some of the formal knowledge that was developed for the presented case study. Figure 5 depicts an example SGT, Equations 1–10 and Figure 6 show some of the applied FMTL rules, and Table 1 lists the available atomic fact types.

$$\begin{aligned} &EdgeDistanceIs(p,q,\delta) \wedge AssociateEdgeDistance(\delta, category) \\ &\quad \rightarrow \boldsymbol{HaveEdgeDistance}(p,q,category) \end{aligned} \tag{1}$$

$$\begin{aligned} &Position(p,x_p,y_p) \wedge Position(q,x_q,y_q) \wedge Size(q,w_q,h_q) \wedge Orientation(q,\theta_q) \\ &\quad \wedge DistancePointToPlane(x_p,y_p,x_q,y_q,w_q,h_q,\theta_q,\delta) \\ &\quad \rightarrow \boldsymbol{EdgeDistanceIs}(p,q,\delta) \end{aligned} \tag{2}$$

$$AngularDistanceIs(p, q, \delta) \wedge AssociateAngularDistance(\delta, category) \rightarrow \boldsymbol{HaveAngularDistance}(p, q, category) \quad (3)$$

$$Position(p, x_p, y_p) \wedge Orientation(p, \theta_p) \wedge Position(q, x_q, y_q) \wedge Angle(x_p, y_p, \theta_p, x_q, y_q, \delta) \rightarrow \boldsymbol{AngularDistanceIs}(p, q, \delta) \quad (4)$$

$$Position(p, x_p, y_p) \wedge AbsoluteArmAngle(p, \theta_{arm_{abs}}) \wedge Position(q, x_q, y_q) \wedge Angle(x_p, y_p, \theta_{arm_{abs}}, x_q, y_q, \delta) \wedge AssociateAngularDistance(\delta, \boldsymbol{close}) \rightarrow \boldsymbol{ExtendingArmToward}(p, q) \quad (5)$$

$$Orientation(p, \theta_p) \wedge ExtendingArm(p, \theta_{arm}) \wedge AngularSum(\theta_p, \theta_{arm}, \theta_{arm_{abs}}) \rightarrow \boldsymbol{AbsoluteArmAngle}(p, \theta_{arm_{abs}}) \quad (6)$$

$$\delta_y = y_q - y_p \wedge \delta_x = x_q - x_p \wedge Atan2(\delta_y, \delta_x, \theta_{yx}) \wedge AngularDifference(\theta_p, \theta_{yx}, \delta) \rightarrow \boldsymbol{Angle}(x_p, y_p, \theta_p, x_q, y_q, \delta) \quad (7)$$

$$Speaking(p) \vee Gesticulating(p) \vee ExtendingArm(p, \theta) \rightarrow \boldsymbol{Interacting}(p) \quad (8)$$

$$\Diamond_{-1}Interacting(p) \vee Interacting(p) \vee \Diamond_{1}Interacting(p) \rightarrow \boldsymbol{InteractingInInterval}(p) \quad (9)$$

$$\Diamond_{-1}InteractingInInterval(p) \wedge \Diamond_{1}InteractingInInterval(q) \rightarrow \boldsymbol{InteractingTogether}(p, q) \quad (10)$$

Root in Figure 5 sorts the modeled objects into lists according to arbitrary FMTL sort criteria, in this case objects p with *Type(p, person)*. The situation scheme *TableGroup* selects objects *Table* with *Type(Table, table)*. For each of them, the list containing all persons is filtered into a list containing only persons that are close to that table. *HaveEdgeDistance(Elem, Table, close)* calculates the distance between a person's center and an object's closest edge and then associates this distance with fuzzy categories (see Equations 1 and 2 and Figure 6, left). *Filter(InputList, RuleToApply(Elem, ...), OutputList)* applies an arbitrary rule (in this case *HaveEdgeDistance(...)*) to each element *Elem* in *InputList* and adds each *Elem* with truth value *V[RuleToApply(Elem, ...)]* $>$ *0* to *OutputList*. *V[Filter(InputList, RuleToApply(Elem, ...), OutputList)]* is the average over all *V[RuleToApply(Elem, ...)]*.

The situation scheme *TableGroup* is refined into *StrategicPlanning* if *Table* = *centralTable* and the director of operations (*doo*) and S1 through S4 are close (determined by fuzzy evaluation in *HaveEdgeDistance(...)*). This can be further refined into *S1-S4OrientedAtDoo* if *HaveAngularDistance(sX, doo, medium)* applies to S1 through S4, i.e. if they have the director of operations in their fuzzy fields of vision (see Equations 3, 4, and 7, and Figure 6, right). The other side of Figure 5 shows how *MessengerMeeting* and its refinements can be deduced once *TableGroup* has been established. *Table* needs to be instantiated as *messageTable* and the staff handling incoming and outgoing messages (*msgr* and *msgr2*) need to be close (*HaveEdgeDistance(...)*). Then, *HaveAngularDistance({msgr, msgr2}, {msgr, msgr2}, medium)* is used to deduce whether they are oriented at each other.

Figure 5 depicts one branch from the current SGTs. Other branches recognize events centered around persons, notepads, messages, and displays. Furthermore,

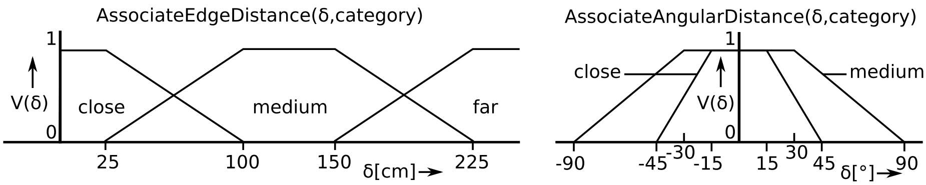


Fig. 6. Visualization of FMTL rules associating distances to distance categories

Table 1. Atomic facts from the input data (static and dynamic object attributes)

$Present(p)$	$Orientation(p, \theta_1)$	$Speaking(p)$	$Sitting(p)$
$Position(p, x, y)$	$Type(p, \tau_1)$	$Gesticulating(p)$	$LookingDown(p)$
$Size(p, w, h)$	$Subtype(p, \tau_2)$	$ExtendingArm(p, \theta_2)$	$OwnerOf(p, q)$

all branches can contain further conceptual refinements for describing interaction patterns. Equations 5–7 and Figure 6 (right) for example can be used for display centric events, calculating a fuzzy truth value for *ExtendingArmToward(person, display)*. Equations 8–10 can be used in a conceptual refinement for Figure 5. Here, interactivity is checked at the previous, current, and next frame using $\Diamond_{-1}$ and $\Diamond_1$. And if two persons in a group are interacting simultaneously or in short succession, they are probably interacting together, provided that they are facing the same object or facing each other.

7 Results

Figure 7 shows some experimental results generated by the SGT in Figure 5 (black lines displaying truth values V as a function of time t in s), and the corresponding ground-truth that was annotated using the tool depicted in Figure 4 (black dots). The top-left graph shows that the system correctly recognizes when the director of operations and his first officers are gathering around the central table. The bottom-left graph shows that it also succeeds in detecting that the first officers are oriented at the director of operations. The system correctly drops this deduction when the director of operations is referring to a display and the first officers turn to look at it. The top-right graph shows the successful recognition of the supporting staff in charge of message handling (*msgr1* and *msgr2*) meeting at the message table, and *msgr2* briefly stepping away from it twice. And the bottom-right graph shows that the system correctly classifies their orientations.

Table 2 provides an overview of the events that are recognized by the current system (top) and the ones that are still under development (bottom). Note that this list is by no means final and that each event can have multiple refinements where the staff members' roles and object names are taken into account for example. The presented results were obtained in real time, but as the number of involved objects, the predicates' arities, and the complexity of the FMTL rules and SGTs increase, runtime needs to be improved by applying better parallellization, more computer resources, and heuristics about which objects to consider.

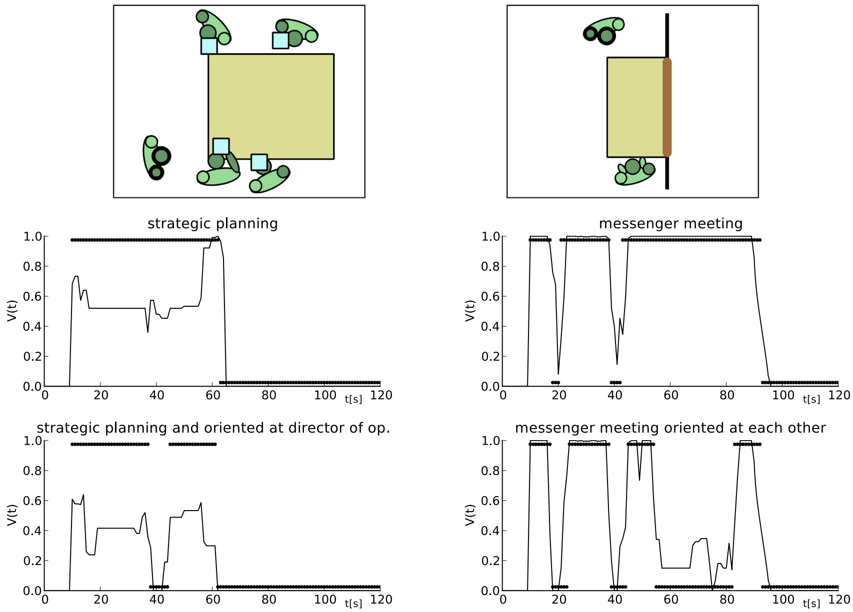


Fig. 7. Experimental results generated by the SGT in Figure 5

Table 2. Events currently recognized (top) and still under development (bottom)

Person alone
Group around {person, notepad, message, table}
Person {joining, leaving} group
Person underway
Person underway with {notepad, message}
Group underway (similar speeds)
Group or person {observing, editing, discussing} {notepad, message} (uses $LookingDown(p)$)
Group or person {observing, editing, discussing} display
Person {talking, listening} to someone
Everybody at own seat (uses $Sitting(p)$ and $OwnerOf(p, q)$, truth value = $\frac{sitting}{all}$)
Briefing phases: introduction, S1, S2, S3, S4, conclusion
Discussions during briefing
Message handling and its phases
Fetching someone to join a group

At the time of writing, a quantitative evaluation was not yet possible. We are currently performing one to evaluate the presented system.

8 Handling Imperfect Input Data

Algorithms for automatic behavior understanding must be able to handle gaps and uncertainty in their input data. Incomplete data handling is important because of possible occlusions in the sensor data, areas without sensor coverage, and technical problems with machine perception components. High-level events can not be detected if some of their rule conditions are not fulfilled due to missing data. Uncertainty handling is important because machine perception components often provide confidence values that should be incorporated into the reasoning process so that uncertainty in perception outputs is reflected in the high-level results as well.

Related problems include various types of noise in the input data as well as wrong data, typically in the form of outliers. Our approach can inherently handle noisy data through FMTL rules applying temporal filtering and fuzzy evaluation. Such rules are also helpful against outliers. Additionaly, outlier detection can be applied to the input data during preprocessing. Outliers could also be detected by the reasoning process itself, using rules about the data's expected dynamics, potentially even providing machine perception with top-down knowledge about this to improve its outputs or guide sensor and resource deployment.

In logic reasoning, the effects of incomplete data can be countered to a certain degree using abduction, where intermediate conditions that can not be deduced are hallucinated instead so that reasoning can continue and certain events can be detected despite their missing conditions. Furthermore, interpolation can be applied to incomplete input data to counter such effects early in the processing chain. It can be applied as preprocessing, independently of the chosen high-level methods. This effectively turns the missing data problem into an uncertainty problem, because interpolated data should have appropriate confidence values associated with them that depend on the confidence in the surrounding data used for interpolation as well as on the temporal distances between new data points and the ones they were calculated from. Confidence should increase towards an interpolated gap's edge, and large gaps should cause ever lower confidence values as you move to the center. In [27], we describe how to apply abduction and interpolation to the FMTL/SGT framework.

Uncertainty in the input data (from interpolation or other causes) can be handled in the FMTL/SGT framework as follows. Each perception output i can have a confidence value $P[i]$ between 0.0 an 1.0. Let a and b be two points on a plane with confidence values $P[a]$ and $P[b]$. The Euclidian distance between a and b is calculated by an appropriate FMTL rule as $\delta_{ab} = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}$. Confidence values are usually combined through multiplication and δ_{ab} depends on a and b, so $P[AssociateDistance(\delta_{pq}, category)] = P[\delta_{pq}] = P[a]P[b]$. Vague truth values for $V[AssociateDistance(\delta_{pq}, category)]$ are calculated from δ_{pq} as in Figure 6, regardless of these confidence values. This means that uncertainty and vagueness are represented separately. Using appropriate FMTL conjunction semantics, each $P(f)$ and $V(f)$ can be condensed into $V'(f)$, a truth value reflecting both uncertainty and vagueness.

9 Conclusion

The presented toolkit for automatic behavior understanding generates semantic event descriptions from machine perception using fuzzy metric temporal logic (FMTL) and situation graph trees (SGTs). It was applied to a case study on automatic behavior report generation for training purposes in crisis response control rooms. Instead of machine perception and multimodal fusion we used a newly developed annotation tool to provide the reasoning engine with input. The paper contains several novel contributions: a new dataset, a new tool for data analysis and annotation, a unique case study with a large amount of perception modalities and objects, a newly developed FMTL/SGT knowledge base for this case study (also applicable to other domains), corresponding experimental results, and an explanation on how to handle imperfect input data.

This forms the basis for our future work. First and foremost, exhaustive quantitative evaluations will be performed on the case study data, comparing the results to ground-truth in a precision/recall-fashion. Our annotation tool currently generates binary ground-truth, but we would like to expand this to n-valued or fuzzy ground-truth because of the fuzzy nature of the results. Second, we will keep improving the FMTL rules and SGTs to recognize more sophisticated events from the presented case study data, exploiting the full power of fuzzy evaluation and temporal modality. Third, such experiments will be performed on various types of imperfect input data as described in Section 8 to evaluate the system's robustness. Fourth, the annotation tool shall be developed further as described at the bottom of Section 5.

We are also starting to involve end-users, human science experts, and software developers. We currently focus on the physical attributes of the people and objects in the room (hypothetical machine perception outputs), but the system can be improved by taking into account more domain specific attributes, i.e. context information (unit status, crisis dynamics, staff roles, and more object information). This would allow us to model more sophisticated expert knowledge in FMTL and SGTs in order to deduce a richer set of semantic event descriptions that is of greater use to potential end-users. To achieve this, we plan to organize a seminar with participants from the State Fire Service Institute (Institut der Feuerwehr) Nordrhein-Westfalen and participants from the research group that was involved in the data recording. In addition to the audiovisual data, they gathered and analysed the messages, documents, and context-related developments of the staff exercise. We are also investigating the alternative application domains listed in Section 1. Our research is situated in an environment that focuses on computer vision and other forms of machine perception, which facilitates the progress toward an online system. The ultimate goal is an unsupervised real-time system containing multiple machine perception components and multimodal fusion instead of manual annotation, with embodiment and action generation, and synchronous visualization of sensor data, machine perception, and semantic event descriptions.

Acknowledgements. This work is supported by the Fraunhofer-Gesellschaft Internal Programs under Grant 692 026. We thank the State Fire Service Institute (Institut der Feuerwehr) Nordrhein-Westfalen for providing the opportunity to record a staff exercise and for their valuable expert knowledge and feedback.

References

1. Aggarwal, J.K., Ryoo, M.S.: Human Activity Analysis: A Review. Computing Surveys 43(3), 16:1–16:43 (2011)
2. Ye, J., Dobson, S., McKeever, S.: Situation Identification Techniques in Pervasive Computing: A Review. Pervasive and Mobile Computing 8(1), 36–66 (2011)
3. Turaga, P., Chellappa, R., Subrahmanian, V., Udrea, O.: Machine Recognition of Human Activities: A Survey. IEEE Transactions on Circuits and Systems for Video Technology 18(11), 1473–1488 (2008)
4. Kosmopoulos, D.I., Doulamis, N.D., Voulodimos, A.S.: Bayesian Filter Based Behavior Recognition in Workflows Allowing for User Feedback. Computer Vision and Image Understanding 116(3), 422–434 (2012)
5. Fischer, Y., Beyerer, J.: Defining Dynamic Bayesian Networks for Probabilistic Situation Assessment. In: Conference on Information Fusion, Internat, FUSION (2012)
6. Shi, Y., Huang, Y., Minnen, D., Bobick, A., Essa, I.: Propagation Networks for Recognition of Partially Ordered Sequential Action. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 862–869 (2004)
7. Aloimonos, Y., Guerra-Filho, G., Ogale, A.: The Language of Action: A New Tool for Human-Centric Interfaces. In: Human Centric Interfaces for Ambient Intelligence, pp. 95–131. Elsevier (2009)
8. Kitani, K.M., Sato, Y., Sugimoto, A.: Recovering the Basic Structure of Human Activities from Noisy Video-Based Symbol Strings. International Journal of Pattern Recognition and Artificial Intelligence 22, 1621–1646 (2008)
9. Ivanov, Y., Bobick, A.: Recognition of Visual Activities and Interactions by Stochastic Parsing. IEEE Tran. on Pattern Analysis and Machine Intell. 22(8), 852–872 (2000)
10. van Kasteren, T.L.M., Englebienne, G., Kröse, B.J.A.: Hierarchical Activity Recognition Using Automatically Clustered Actions. In: Keyson, D.V., Maher, M.L., Streitz, N., Cheok, A., Augusto, J.C., Wichert, R., Englebienne, G., Aghajan, H., Kröse, B.J.A. (eds.) AmI 2011. LNCS, vol. 7040, pp. 82–91. Springer, Heidelberg (2011)
11. Filippaki, C., Antoniou, G., Tsamardinos, I.: Using Constraint Optimization for Conflict Resolution and Detail Control in Activity Recognition. In: Keyson, D.V., Maher, M.L., Streitz, N., Cheok, A., Augusto, J.C., Wichert, R., Englebienne, G., Aghajan, H., Kröse, B.J.A. (eds.) AmI 2011. LNCS, vol. 7040, pp. 51–60. Springer, Heidelberg (2011)
12. Yao, B.Z., Yang, X., Lin, L., Lee, M.W., Zhu, S.C.: I2T: Image Parsing to Text Description. Proceedings of the IEEE 98(8), 1485–1508 (2010)
13. Ryoo, M., Aggarwal, J.: Semantic Representation and Recognition of Continued and Recursive Human Activities. International Journal of Computer Vision 82, 1–24 (2009)
14. Gupta, A., Srinivasan, P., Shi, J., Davis, L.: Understanding Videos, Constructing Plots, Learning a Visually Grounded Storyline Model from Annotated Videos. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2012–2019 (2009)
15. Sadilek, A., Kautz, H.: Location-Based Reasoning about Complex Multi-Agent Behavior. Journal of Artificial Intelligence Research 43, 87–133 (2012)
16. Morariu, V., Davis, L.: Multi-Agent Event Recognition in Structured Scenarios. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 3289–3296 (2011)

17. Kembhavi, A., Yeh, T., Davis, L.: Why Did the Person Cross the Road (There)? Scene Understanding Using Probabilistic Logic Models and Common Sense Reasoning. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part II. LNCS, vol. 6312, pp. 693–706. Springer, Heidelberg (2010)
18. Bohlken, W., Neumann, B.: Generation of Rules from Ontologies for High-Level Scene Interpretation. In: Governatori, G., Hall, J., Paschke, A. (eds.) RuleML 2009. LNCS, vol. 5858, pp. 93–107. Springer, Heidelberg (2009)
19. Augusto, J.C., Nugent, C.D. (eds.): Designing Smart Homes, The Role of Artificial Intelligence. Springer (2006)
20. Gottfried, B., Aghajan, H. (eds.): Behaviour Monitoring and Interpretation, Smart Environments. IOS Press (2009)
21. Gong, S., Xiang, T.: Visual Analysis of Behaviour, From Pixels to Semantics. Springer (2011)
22. Ley, B., Pipek, V., Reuter, C., Wiedenhoefer, T.: Supporting Improvisation Work in Inter-Organizational Crisis Management. In: ACM Annual Conference on Human Factors in Computing Systems (CHI), pp. 1529–1538 (2012)
23. Toups, Z.O., Kerne, A., Hamilton, W.A., Shahzad, N.: Zero-Fidelity Simulation of Fire Emergency Response: Improving Team Coordination Learning. In: ACM Annual Conference on Human Factors in Computing Systems (CHI), pp. 1959–1968 (2011)
24. Nagel, H.H.: Steps Toward a Cognitive Vision System. AI Magazine 25(2), 31–50 (2004)
25. Arens, M., Gerber, R., Nagel, H.H.: Conceptual Representations between Video Signals and Natural Language Descriptions. Image and Vision Computing 26(1), 53–66 (2008)
26. Bellotto, N., Benfold, B., Harland, H., Nagel, H.H., Pirlo, N., Reid, I., Sommerlade, E., Zhao, C.: Cognitive Visual Tracking and Camera Control. Computer Vision and Image Understanding 116(3), 457–471 (2012); Special Issue on Semantic Understanding of Human Behaviors in Image Sequences
27. Münch, D., IJsselmuiden, J., Grosselfinger, A.-K., Arens, M., Stiefelhagen, R.: Rule-Based High-Level Situation Recognition from Incomplete Tracking Data. In: Bikakis, A., Giurca, A. (eds.) RuleML 2012. LNCS, vol. 7438, pp. 317–324. Springer, Heidelberg (2012)
28. Münch, D., Jüngling, K., Arens, M.: Towards a Multi-Purpose Monocular Vision-based High-Level Situation Awareness System. In: International Workshop on Behaviour Analysis and Video Understanding @ ICVS (2011)
29. González, J., Rowe, D., Varona, J., Roca, F.X.: Understanding Dynamic Scenes Based on Human Sequence Evaluation. Image and Vision Computing 27(10), 1433–1444 (2009); Special Sec. on Comp. Vis. Meth. for Ambient Intelligence
30. Bellotto, N.: Robot Control Based on Qualitative Representation of Human Trajectories. In: AAAI Symposium on Designing Intelligent Robots: Reintegrating AI (2012)
31. Münch, D., IJsselmuiden, J., Arens, M., Stiefelhagen, R.: High-Level Situation Recognition Using Fuzzy Metric Temporal Logic, Case Studies in Surveillance and Smart Environments. In: 2nd IEEE Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams (ARTEMIS) @ ICCV (2011)

Combining Implicit and Explicit Methods for the Evaluation of an Ambient Persuasive Factory Display

Ewald Strasser, Astrid Weiss, Thomas Grill, Sebastian Osswald, and Manfred Tscheligi

ICT&S Center, University of Salzburg, Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria
{Ewald.Strasser,Astrid.Weiss,Sebastian.Osswald, Manfred.Tscheligi}@sbg.ac.at

Abstract. Research in ambient intelligent systems faces a challenging endeavor, namely the evaluation of user experience of ambient displays. Due to the fact that ambient displays should be unobtrusive, it is hard for users to appraise them on a reflective level (i.e. interviews and questionnaires). In this paper we present a methodological approach that combines an implicit (the Affect Misattribution Procedure (AMP)) and an explicit measurement technique (questionnaire for the persuasive effect (PeQ)) to tackle this problem. We used this approach in a study of an interface (Operator Guide) that provides information to operators in a semiconductor factory. Results show that the implicit technique is better suited to assess fine attitudinal differences on how users experience the display than explicit questionnaires. However, explicit measures are valuable to gain suggestions for improvements and thus it is concluded that this method triangulation adds value for the research on ambient persuasive interfaces.

Keywords: Method triangulation, Methodology, implicit, Affect Misattribution Procedure, AMP, persuasion, ambient, automatic attitude.

1 Introduction

Ambient intelligent systems refer to the notion of natural interaction with ubiquitous computing systems surrounding us [1]. The vision of ambient intelligent environments describes the pervasion of the everyday world with digital technology. This allows to anticipate the needs of users and to support them when interacting with such systems [2]. Ambient intelligent systems are embedded into the user's environment, adapt to their requirements, and react to their presence. Therefore, the aim of ambient intelligent system is that users do not have to proactively interact with the system, but are supported by it an unobtrusive way [3]. Unobtrusive ambient intelligent systems offer the opportunity to persuade people on a subtle or subconscious level [4]. Preliminary research on this aspect demonstrated the potential of the usage of ambient technology for a persuasive purpose [5-7]. However, subtle AMI systems design and user experience aspects play also a dominant role [8] in acceptance and rejection of a new interface. User experience factors, such as emotion and attitude are still difficult to

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 113–128, 2012.

measure [9] and therefore difficult to apply in the evaluation of ambient intelligent system. User experience is often measured on the reflective level by the means of interviews and surveys [e.g. 10], however, this approach contains several open questions: (1) What exactly are the limitations of cognitive portability of emotional experience? (2) Does the act of reporting change the experience? (3) How should we try to assess different modalities, types, dimensions, and relational aspects of an emotional experience [11]. The assessment of emotions is currently enabled either by explicit approaches that include mostly physiological measurements like EEG [e.g. 12]. Physiological measurements with their electrodes and cables have a very intrusive character and are questionable regarding the fact that the results are influenced by and contain artifacts from the technique itself. Implicit measurement techniques are less intrusive and do not ask for conscious ratings. Although implicit measurement techniques have some weaknesses, they are supposed to reliably measure automatic attitudinal responses [e.g. 13].

In this paper we address how a method triangulation of explicit and implicit measurement techniques can be used to appraise ambient intelligent systems on a reflective and on an automatic or subliminal level. Implicit measurement techniques have a high incremental validity which means they measure different aspects of the same topic and hence complement each other [14]. Roughly summarized persuasiveness is the potential of an interface to explicitly or implicitly induce a change in behavior or attitude. Thus we conducted a case study with an ambient display, the Operator Guide (OG) [15], in which we used a post-study survey as an explicit measurement and the Affect Misattribution Procedure (AMP) as implicit measurement technique to assess the attitude and the persuasive potential of the system. It was not the authors aim to directly compare implicit and explicit measurement techniques because it is already shown that these two approaches do only moderately correlate with each other and are better regarded as complementary [16].

In the following section we investigate the state of the art of user experience measurement (for a more thorough survey see [17]) and proceed with a presentation of the challenges of a semiconductor factory for ambient intelligent systems, followed by a description of the OG as an ambient intelligent guidance tool for the factory. Subsequently, we present the OG study, a laboratory-based experiment. The study is analyzed and the results are discussed in terms of applicability and utility. We conclude the paper with a summary of the findings as well as an outline of the limitations of the work.

2 Related Work

As [18] summarizes, usability and functionality becomes standard when an industry becomes mature and hence user experience is often the differentiating factor that decides if a product succeeds over a competitor's product or not. In the case of ambient intelligence this statement is especially true because the usability of an ambient intelligent system must reach a level where technology becomes unobtrusive and therefore the influence of usability on user experience becomes marginal. Factors like appearance become the main influence on user experience and therefore the feature that decides if users accept a product. According to [18] evaluation methods for user experience are still inadequate. This becomes crucial when we need to link these results with private and individual experiences [19].

Looking at the current attempts of user experience evaluation we found two main streams: On the one hand there is the "classical" way stemming from the area of cognitive sciences to measure user experience with questionnaires or ask explicitly for specific user experience factors. Such self-report measurements include strategies like aspirations and expectations, task observations which focus on specific elements of the design, or post experience feedback [20]. In literature these techniques are frequently criticized for their reconstructive nature [21-26], their selective bias of attention [27, 28], their demand characteristics [29], and the distracting effects of cultural embedded theories about causes of feelings and experiences. However, in our research experience on the assessment of user experience aspects of ambient persuasive displays self-reporting measures have proven their usefulness to address the challenges that are arising when a detailed assessment of specific user experience aspects is necessary [30].

The use of questionnaires is often combined with log data and more qualitative data like incent diaries and wish lists [e.g. 31]. We believe that this approach represents a shift towards that part of user experience that is not measureable on a reflective level. Especially the assessment of logging data is already a way to implicitly assess and test assumptions about the subconscious user experience. In the context of evaluating ambient intelligent systems this seems to be an appealing solution, as the systems often have to log user data as basic functionality. The weakness of this approach is that the evaluated product has to reach a very high degree of matureness, which further contradicts with the wish to evaluate user experience in a very early stage of the design process in order to take them into account in future iterations. Even if there is the possibility to gain log data at an early stage it is difficult to causally match design features to a certain measure. [11]

The second methodical approach for assessing user experience is the measurement with sensorial equipment that assesses physiological sensations of users to causally link them with the perceived user experience of a certain product. EEG [e.g.: 12] and eye-tracking [e.g.: 32] are techniques that are commonly used to assess users experience while interacting with a computing system. Based on their quantitative nature, the results of such measurements are good candidates to provide generalizable and comparable outcome. The main problem with this approach is its intrusiveness. The proper preparation of an EEG study lasts for each participant at least half an hour, meanwhile participants cannot move freely and the EEG cap is omnipresent. It is clear that the interpretation of user experience is difficult as such measurement approaches already affect the user's state and user experience with the study object under observation. This is in particular true for ubiquitous systems with their need for unobtrusiveness.

To our conviction the approach of using implicit measurement methods like the AMP [33] could help to deal with these various challenges. As shown by Bassili [34] such operative measures which are thought to be direct manifestations of the information processing of attitudinal judgments are less in danger to be influenced by extraneous influences than meta-attitudinal measures which require reflection on the attitude like many questionnaire approaches do. These measures are certainly not a general solution in the difficult field of user experience evaluation, but they cover a part of user experience that was up to now difficult to assess and a combination of explicit and implicit measurement methods will be an economic solution for the assessment of user experience on a standardized measurement basis. Before introducing our measurement approach we describe the research context in which we applied it: A semiconductor factory, which uses an ambient persuasive display (the Operator Guide) to support operators in their daily working routines. The characteristic of this context and the experience it triggers underlines our claim for broad user experience assessment.

3 Research Context: A Ubicomp Factory

While the term "factory context" may refer to the context of the various types of factories (e.g. steel works, car factory or food factory), we focus on the specific context of semiconductor factories. A semiconductor factory is a complex interplay of several entities. Its overall purpose is to manufacture as many error free integrated circuits (wafer) as possible. The central part of a semiconductor factory is the cleanroom, an area where the environment is controlled to eliminate all dust, since even a single speck can ruin a microcircuit with its features being much smaller than dust. Contrary to offices, where employees sit on a desk in front of a single computer, operators (the workers in a semiconductor factory) within a cleanroom have to move between several kinds of interfaces to gain all the information they need. The interaction is challenging as the work has to be done quickly and exactly because human errors can lead to high costs.

The semiconductor fabrication plant where we are investigating has dedicated its European factories towards the development and production of new technologies. This requires short and more complex production cycles, which demands a high flexibility within the whole production system. This is often contradictory to a high level of automation. The intention behind our research is not to have a fully automated factory without human support, but to foster a synergetic relationship between human operators and the surrounding technologies within the context. Therefore, we started to investigate ambient technologies (the factory is also rarely tackled in the research field on ambient and ubiquitous systems, e.g. [35]), which are integrated into the cleanroom environment to give the operators the possibility to interact with various machines (so called equipment) and intelligent automated systems in an intuitive way and transport important information subliminal and without additional cognitive load to the operators [4]. The idea of smart automation has already been implemented by the semiconductor manufacturer to a certain extend. Within the so-called "intelligent factory project" the logistic process has been modified by introducing a combination of different radio technologies with ultrasound technologies. Innovative hardware (e.g. RFID) and software (e.g. message bus architectures) technologies build the basis for the integrated smart automation approach and made the factory and especially the cleanroom to an ambient intelligent environment. However, to reach the goal of “zero defect” production smart automation does not imply full automation. The concept is rather based to support the operators within their activities in the manufacturing process.

The workflow in a semiconductor factory implies a huge amount of distributed information an operator needs to perceive and use. The information is spread over multiple entities like machines, displays, measurement devices, etc. in the context and influences the user's workflow as well as his performance when trying to comply with the main goal of achieving the “zero-defect-factory” paradigm in a 24/7/365 production process.

We developed the OG to support especially the factory as a ubiquitous context. We conducted ethnographic studies to collect requirements for the particular system. The OG was designed as a persuasive ambient interface, with the aim of serving as an additional information system to support the operators to make better-informed decisions [15] by visualizing the upcoming working steps with regard to the next machine for processing wafers, rather than the next lot box (box with silicon wafers of microchips) to process. In this way, the OG should foster a higher compliance to management orders by guiding the operator through the optimal next working steps, improve the efficiency, and reduce errors by simplifying the interface and therefore reducing the information overload.

Information overload can cause errors and therefore we were interested in investigating the effect of different alerting modes combined with the OG to enhance the perception of critical situations. To asses which modality was perceived best in terms of user experience we used the AMP technique as implicit and the PeQ questionnaire as explicit measurement technique, which are in the following described in detail.

4 Measurement Techniques

Hofmann [16] showed that implicit and explicit measurement methods measure different psychological constructs. Even if they are well aligned to each other on the object they want to measure, the correlation between these two concepts is not very high. When research on implicit methods started, the missing link to explicit methods was criticized and interpreted as proof against the validity of these methods. Later it was shown [14, 36-38] that these methods are valid in the sense that they are able to predict behavior. In particular they are suitable to predict behavior that is automatically activated without conscious awareness. Therefore, we consider them as an ideal completion for explicit measurement where participants have to reflect about their behavior.

4.1 “Implict” AMP

Implicit measurement methods like the AMP are used to assess automatic attitude toward the object of interest. Users are asked to fulfill tasks that have seemingly nothing to do with the issue, but implicitly these ratings are influenced by the automatic attitude of the participant. In our study we used the Affect Misattribution Procedure (AMP) [33] to measure the overall automatic attitude towards the operator guide. In the AMP participants are seated in front of a computer screen and they are instructed to rate whether they perceive abstract patterns, such as Chinese characters, as beautiful or ugly. They are told that this is an experiment about how people make quick decisions. Rating is done over the keyboard and normally the “E” and the “I” key are used to rate if they find the Chinese letter pleasant or unpleasant (see Fig. 1). After a training phase in which participants learn how to rate in the AMP they see a screen where on the left side either the word “Pleasant” or “Unpleasant” is written and on the right side there is also the word “Pleasant” or “Unpleasant”. On which side of the screen the categories “Pleasant” or “Unpleasant” are placed is ordered randomly and the mix is counterbalanced for the whole sample to control for sequence effects. If the word “Unpleasant” is on the left side of the screen, participants have to press “E” to rate the picture as unpleasant. The actual procedure consists of three short sequences. First, a pictures of the object that should be evaluated or the contrast category (an object that is the opposite or a good contrast to the object that should be evaluated) or the neutral stimuli (white noise picture) is presented (primed) for 150ms, second a randomly selected Chinese letter is presented for 200ms and at last a masking stimuli (a white noise picture that is not associated with any of the used pictures) is laid over the Chinese letter till the participant rates the picture. The participants are instructed to disregard the first object, but they are more inclined to perceive the Chinese character as pleasant if the object before is favorable for them. By showing either the OG or the blank screen in a random order a score for the bias towards the OG and the blank screen is gathered. From the AMP several scores are calculated, but in this study we are only interested in the difference of “Pleasant” ratings for the OG before and after the interaction. Changes in these ratings which are meant to represent automatic attitude towards the systems are caused by the interaction with the interface and mean that there is change in automatic association between the concept pleasantness and the OG.

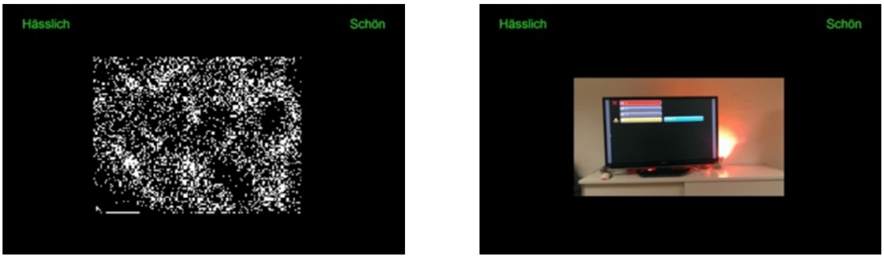


Fig. 1. The Affect Misattribution Procedure (AMP). The left picture shows the AMP with a mask, this picture is seen by the participants when they have to rate if they liked the Chinese letter or not. The right picture shows the priming of the OG.

4.2 "Explicit" Persuasion Questionnaire (PeQ)

The PeQ is an enhanced version of a structured interview from a previously conducted study. This study was conducted to evaluate a persuasive display for a shop window [30]. Regarding the valuable results the authors gathered during their study, we decided to alter this interview in a way that it fits the factory context and to explore its potential as a generalizable persuasion measuring method. Since there is to our knowledge also no other measurement that measures persuasion the way we present it here, the PeQ could be also a valuable contribution for a standardized measurement of the persuasive quality of interfaces.

The PeQ consists of three parts: The first part should assess if participants noticed that there was a display and what alerting method exactly made them aware of the display. The reason for this is that especially ambient displays do not necessary catch so much attention that participants consciously recognize the persuasive display. The second part (in the following referred to as persuasion part) consists of multiple-choice questions, which focus on the persuasive effect of the evaluated interface. The third part of the questionnaire targets the areas of the display the participants would like to customize. This part also includes an open question where participants can write down wishes, which are not listed in the questionnaire.

Our aim for the PeQ was to design the questionnaire in a manner can be used in other studies as well. Therefore questions were designed in a way that it will be possible to replace the name of the ambient persuasive interface and the type of modalities it uses without changing the meaning of the question. However it is clear that reusability is limited and therefore it's necessary to carefully think about the use of this questionnaire in every context and to do further validation studies. Therefore the presented study was used as a first validation study for the PeQ. In all questions we used a five point likert scale to express agreement or disagreement. From the gathered data we calculated values for the reliability and looked at the factor structure of the PeQ.

Regarding reliability we found for the persuasion part a Cronbach's Alpha of .838 and for the wishes part .761. In psychology, an Alpha of .7 and higher is considered acceptable [39].

Further we did a factor analysis as a principal component analysis with a varimax rotation on part two and three of PeQ. The analysis revealed that the originally proposed seven psychological factor structure was not useful. Instead we decided to go for a three factor solution (see Table 1). We interpreted the first factor as the participants reflections about their work with the interface (1), the second factor as a general work attitude factor (2) and in this special case the third factor was the reflection about the work with the vibration wristband (3), which was used in this study as one of the alarming conditions.

For the wish list part of the PeQ, originally one factor was predicted. However, the analysis showed that the three factor solution describes the data best. Our interpretation for the first factor is that it reflects the "wish to interactively adapt the system", the second factor loads on the "modalities of the interface" and the third factor is interpreted as a "general attitude towards intrusive feedback" modalities.

Table 1. Rotated Principal Component Analysis, using varimax rotation for the PeQ questionnaire. Questions are marked according to their highest loading for factor 1 (participants reflections about their work with the interface), factor 2 (general work attitude) and factor 3 (reflection to work with the vibration wristband).[1]

	1	2	3
The operator guide (OG) influenced my attitude towards this work positive.	.916	.080	.057
The design of the OG made me curious and so I worked with it.	.813	.027	-.056
The OG caught my attention.	.738	-.382	.310
I imagine that it is useful to have an OG for this kind of work.	.675	.205	.084
The OG reminded me of the next relevant work step.	.643	-.052	-.368
Light at the OG made me curious and so I worked with it.	.580	-.561	.271
I can imagine working with the OG.	.534	.338	.445
Light at the OG caught my attention.	.465	-.660	.223
I like to work.	.350	.543	.325
I enjoy working more when I have operating instructions (paper or electronic).	.256	.811	-.067
Operating instructions are useful for the work.	.048	.799	-.009
I prefere operating instructions written on paper.	-.102	.644	.380
The vibrating wristband made me curious and so I worked with the OG..	.197	-.120	.855
The vibrating wristband caught my attention and so I worked with the OG..	-.136	.090	.933

The scale "Reflections about the work with the interface" consists of seven questions with a minimum score of 7 and a maximum score of 35, whereas the scale "General work attitude" has 5 questions with a minimum score of 5 and a maximum score of 25. The data gathered with the PeQ was used to measure the explicit persuasive effect of the interface on our participants.

5 The Operator Guide Study

The goal of the study was to compare findings from explicit and implicit measurement techniques and show the unique potential to assess user experience. In the following we describe the experimental design of the study, which was conducted in our user experience laboratory.

6 Study Design

Our sample consisted of 25 participants with 13 male and 12 female persons which were recruited on the website of the federal students union. None of them had experience in the semiconductor production and they were asked if they are willing to participate in a series of lab studies. The average age was 27.5 (SD=9). The oldest person was 52 and the youngest 20 years old.

Our first goal, was to assess how the alerting modalities ambient light and vibration wristband affect how participants experience the OG.

The original intended version of the OG was designed as a completely “passive” ambient display, which should only inform the operator about the next steps in the working progress. However, this initial design did not take into account intensive warning for suddenly occurring problems, such as equipment failures. Thus an investigation about how to alert operators was considered a reasonable next step. Equipment failure cannot be manipulated in the real cleanroom context because this would heavily disturb the production process, the study was conducted as laboratory evaluation with a simulated cleanroom scenario.

The study was set-up as a repeated measures within-subject experiment with three experimental conditions. The plain OG should serve as a baseline condition (first condition, strong ambient character) compared to two other conditions, which differ in their degree of intrusiveness, namely additional ambient light (second condition - visual perception channel; more salient, not very strong ambient character) and a vibrating wristband (third condition - haptic channel; most salient, weak ambient character). Fig. 2 shows the OG with the ambient light (light when error) and the vibrating wristband (vibrates when error).

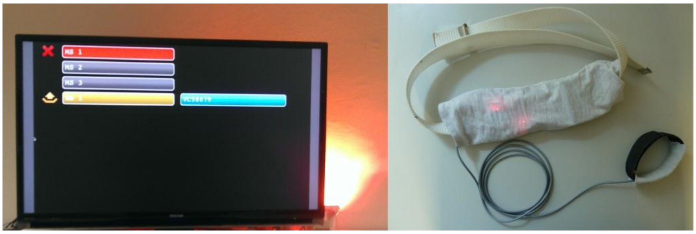


Fig. 2. Left picture: Ambient light condition of the OG, on the screen the machines of the simulation are listed and the status is shown by the color of the items on the left of the screen is a suggestion which lot to do next; Right picture: Vibrating wristband condition

Due to the experience differences that could occur between real operators we decided to ”contextualize” normal participants, which means that we instructed the participants (approximately 45 minutes) with training material applied also for real operators. The order of the experimental conditions was counterbalanced in order to controll carryover and learning effects in our measures.

6.1 Study Procedure and Hypothesis

On the basis of the described study design we wanted to investigate the following two hypotheses, regarding the two measures we used:

1. Depending on the alerting condition the user perception of the persuasive potential differs (PeQ measure).

2. Independent of the alerting condition an automatic attitude change caused through the interaction with one of the OG versions can be identified (AMP measure).

Instruction. Participants were told about the purpose of the study and were instructed which tasks they have to conduct in the defined scenario. To "contextualize" them with a cleanroom experience they were shown a video about the cleanroom and they also had to wear a cleanroom suit, gloves, and a mask. Before our participants could work in the simulated cleanroom scenario they needed to be trained on the common workflow of an operator for wafer production to understand the meaning of the different elements appearing during the study. The main focus of these instructions was on the simulated workflow and the "contextualization" with the cleanroom experience. From previous studies in which the researchers were working as trainees in the cleanroom, we know which information is relevant for novice users to perform simple wafer production tasks. After the instruction which lasted approximately 45 minutes, participants had to conduct the first AMP, which served as a baseline for the second AMP that was conducted after the simulation. We used the AMP only in one condition because it was conceivable that further measures would be a mixed attitude of all conducted conditions. Therefore we decided to make the study less stressful for participants and thus gather better data quality. Through the counterbalancing of the experimental conditions the AMP was used equally in every experimental condition, but for a comparison the number of participants per condition was too small.

Task Conduction. Participants had to work in the simulated cleanroom scenario in each of the three experimental conditions. For the simulated cleanroom scenario it was not only necessary to build a physical environment that was similar to the real cleanroom and support the cleanroom experience, we also needed to prepare the formal and informal working rules of the semiconductor production. We reduced this complex workflow to the basic knowledge, which is also applied in a real cleanroom to introduce novice users. The overall task of an operator in the cleanroom is to load equipment as efficient as possible. Therefore the following rules were explained to the participants:

- Try to avoid idle times on the equipment. This means that empty equipment has to be loaded as fast as possible and finished lots have to be replaced as fast as possible.
- Try to debug equipment that is out of order. This can be done by typing in the error code that is shown on the screen next to the equipment.
- Finished or empty working equipment always outranks the debugging of broken equipment. This means that debugging must not be started when other equipment is idle or finished and that debugging has to be stopped when the equipment finished processing.

After participants had done this first run in the simulation they conducted the AMP again. The score gathered in this run of the AMP was compared with the baseline from the beginning and differences between these two values are caused by the interaction. The AMP was followed by demographic questions and the PeQ.

Post Questionnaire. After each condition participants had to fill in the PeQ questionnaire.

The study was conducted with 25 participants and every person had to perform each of the experimental conditions. Fig. 3 shows the simulated cleanroom in which participants had to work.

Fig. 3. Recording of participant working in the simulated factory environment. Picture shows an operator taking a lot from a shelf before bringing it to the equipment in the background.

7 Results

In the following we present our results relating to our hypotheses followed by a discussion of the particular results.

Hypothese 1: Depending on the alerting condition the user perception of the persuasive potential differs (PeQ measure).

Regarding our first question if different alerting conditions of an ambient intelligent interface have different persuasive potential we found in fact a difference between the three alarming conditions regarding the scale "vibra wristband" in the PeQ. An ANOVA for the scale "vibra wristband" in the PeQ showed that the conditions were significantly ($F(2, 75)=6.807$; $p=.002$; $f=.149$) differently appraised. A Kolmogorov-Smirnov test showed that the data is normal distributed. A post hoc comparison with the Bonferroni correction showed that only the wristband alarming condition in the simulation differed significant from the other conditions ($p=.004$ for ambient light; $p=.009$ for standard condition) regarding their values in the PeQ for the scale "vibra wristband". Looking at the means of the conditions it becomes clear that the participants did not perceive the wristband as very persuasive. With 5.84 (SD=3.13) the vibra wristband condition had the lowest rating followed by the standard OG condition with 8.14 (SD=2.61) and the ambient light condition was rated best with 8.40 (SD=2.38).

Regarding the other factors of the PeQ, an ANOVA found no differences between the conditions. So neither in the scale "participants reflections about their work with the interface" ($F(2, 75)=.314$, $p=.732$) nor in the "general work attitude" ($F(2, 75)=.775$, $p=.464$) there was a difference between the experimental alerting conditions. Table 2 shows means and standard deviations for the scale "participants reflections about their work with the interface" and "general work attitude" for the standard, ambient light and vibra wristband condition.

Table 2. Mean and standard deviation for the scale "Reflections about the work with the interface" and the scale "General work attitude" in the PeQ for the different alarming conditions

		Mean	Std. Deviation
Reflections about the work with the interface	Standard OG	16.43	4.71
	Ambient light	15.60	3.38
	Wristband	15.80	3.69
General work attitude	Standard OG	10.00	1.93
	Ambient light	10.00	1.66
	Wristband	9.48	1.53

The "wish list" part of the PeQ revealed further details of the explicit appraisal by our participants.

The statement that the OG should react on the mood of the user scored highest but at the same time the statement that the user should be able to adapt the elements (fields and notifications) of the OG scored low. This indicates the wish for an automatic adaptation. The question "Would you say that you feel restricted through the interface to make your own decisions?" scored high and indicates that participants could work with the OG.

Hypothese 2: Independent of the alerting condition an automatic attitude change caused through the interaction with one of the OG versions can be identified (AMP measure).

We addressed the second part of hypotheses by analyzing the AMP scores participants reached before and after the interaction with the Operator Guide. The AMP score that was gathered before the interaction was compared with the score after the interaction. Any significant change in the automatic attitude is caused by the interaction. A paired sample t-test revealed that participants significantly ($t=2.144$; $p=.042$, $d=.429$) liked the OG interface more after they interacted with it. The mean values of both conditions show that the OG reached 9.08 (SD 2.72) of 16 possible score points, when participants interacted with it and only 7.76 (SD 3.93) points when they had no contact with the interface.

As the first run took place before participants interacted with the interface we interpret this result as an automatic preference for the OG, caused through interaction and not by a simple cosmetic attraction. Automatic attitude as a part of user experience benefits in this case from the interaction.

8 Discussion

This paper presented an approach to assess the persuasive potential and the "implicit" attitude towards an ambient intelligent interface of the cleanroom in a simulated context. The AMP showed that interaction with the system improves the "implicit" attitude towards the system, which means that the automatic attitude towards the OG improves through the interaction with it.

Thus it can be assumed that the AMP is capable of assessing subtle produced user experience towards ambient persuasive interfaces. Amongst others the most crucial advantage of such techniques is that scores that are assessed with implicit measurement methods predict automatic behavior better than questionnaires [14], which will be especially useful to evaluate ambient intelligent systems that persuade on a subliminal basis [4]. Due to the limitations of the simulation approach it will not be possible to present ecological valid behavior changes but we showed that the AMP is capable of assessing changes in automatic attitude even before users interacted with the system to assess a baseline of its perception.

The possibility to perform the AMP before participants interacted with the system is especially useful for the assessment of ambient intelligent systems because these systems often supposed to work at the background and persuasion can happen on a unconscious level. If we would further like to evaluate ambient intelligent systems with explicit measures we would change their characteristic and destroy the subtle built user experience by making them more visible then they potentially should be.

The PeQ in contrast was more useful for the assessment of conscious wishes of the users then finding differences between the conditions. We assume that these differences were too small to be measured on a reflective level. Since the scale "vibra wristband" in the PeQ is very specifically tailored on the wristband condition and the modality itself was very conspicuous it was easy for the participants to reflect on their opinion towards the persuasive effect of the modality. The other two factors of the PeQ were much more general and it was difficult for participants to reflect on the relating questions.

Not surprisingly it turned out that the wristband condition had significantly the lowest persuasion score in the scale "vibra wristband" and the results from the analysis on basis of questions of the "wish list" part of the PeQ, show that the item which refers to the intrusiveness of the interface has the lowest mean. This can be interpreted as a wish for more intrusive feedback. Combining these results with the result from the first ANOVA for the persuasion part of the PeQ which showed that the "vibra wristband" was not perceived as persuasive is further evidence for the wish for a more intrusive feedback.

Combined results from the AMP and the PeQ show that the "implicit" attitude towards the interface improves through the interaction with it. The interface could be improved by making it adaptable to the customers wishes and especially by an automatically adaptation function, which was only deducible from the PeQ.

The combination of these two measurement methods had the advantage that we could shed light on user experience of our system not only from explicit point of view but also from a more automatic attitude towards the system. Using only the AMP we would have concluded that the interface is fine and we do not need to redesign it, from the perspective we gained from the PeQ we probably would have considered a complete redesign of the system. The more differentiated picture we got from the combination of these two methods will help to save time and will foster an iterative design process.

It is clear that simulation studies have several limitations regarding their generalizability on the factory context. Especially the noisy context, the feeling of dealing with real equipment, the social context and the daily work pressure are not easy to reproduce. Hence it was not the goal of the study to find concrete behavior changes but general tendencies of behavior in cleanroom setting. Although results from lab studies cannot be generalized on the real context exactly as they are, we are confident that this approach will ease our work in the real context of the cleanroom with a more focused knowledge about potential issues to deal with. Since the semiconductor production is a very competitive and time critical industry it is not possible to replicate this study in the real factory. Instead we will use the insights from this study to enhance the current system and evaluate later how the new system performs in the sense of user-experience compared to the old system.

9 Conclusion and Future Work

In this paper we presented an approach that combines implicit and explicit measurement techniques to improve efficiency regarding the evaluation of user experience with ambient intelligent systems. Both measurement methods show good incremental validity to each other in the sense that the AMP is able to assess fine differences in the automatic attitude towards a system and the PeQ is able to assess conscious available thoughts that are useful for concrete improvements of the system. We know that simulation studies have several limitations, especially when it comes to predicting behavior changes. Thus our focus was on the evaluation the perceived persuasive effect and the "implicit" attitude towards this display with implicit and explicit measurement methods. Further evaluations of the OG will therefore be aligned on the useful comparison of behavioral data with our measurements.

Acknowledgements. The financial support by the Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged (Christian Doppler Laboratory for "Contextual Interfaces"). We also wish to thank Alexander Meschtscherjakov, Florian Phör, and Wolfgang Reitberger for their support in the development of the Operator Guide.

References

1. Weiser, M.: The Computer for the 21st Century. Scientific American 265, 94–104 (1991)
2. Aarts, E.: Ambient Intelligence: A Multimedia Perspective. IEEE MultiMedia 11, 12–19 (2004)
3. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proceedings of CHI 1997, pp. 234–241. ACM Press, Atlanta (1997)
4. Riener, A.: Subliminal Persuasion and Its Potential for Driver Behavior Adaptation. IEEE Transactions on Intelligent Transport Systems 13, 71–80 (2012)
5. Rogers, Y., Hazlewood, W.R., Marshall, P., Dalton, N., Hertrich, S.: Ambient influence: can twinkly lights lure and abstract representations trigger behavioral change? In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, pp. 261–270 (2010)

6. Ham, J., Midden, C., Beute, F.: Can ambient persuasive technology persuade unconsciously?: Using subliminal feedback to influence energy consumption ratings of household appliances. In: Proceedings of the 4th International Conference on Persuasive Technology, pp. 29–35 (2009)
7. Balaam, M., Fitzpatrick, G., Good, J., Harris, E.: Enhancing interactional synchrony with an ambient display. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 867–876 (2011)
8. de Ruyter, B.: Social interactions in Ambient Intelligent environments. Journal of Ambient Intelligence and Smart Environments 3, 175–177 (2011)
9. Varela, F.J.: Neurophenomenology: A methodological remedy for the hard problem. Journal of Consciousness Studies 3, 330–349 (1996)
10. Vastenburg, M., Keyson, D., De Ridder, H.: Measuring User Experiences of Prototypical Autonomous Products in a Simulated Home Environment. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4551, pp. 998–1007. Springer, Heidelberg (2007)
11. Nielsen, L., Kaszniak, A.W.: Conceptual, theoretical, and methodological issues in inferring subjective emotion experience. In: Handbook of Emotion Elicitation and Assessment, pp. 361–375 (2007)
12. Mikhail, M., El-Ayat, K., El Kaliouby, R., Coan, J., Allen, J.J.B.: Emotion detection using noisy EEG data. In: Proceedings of the 1st Augmented Human International Conference, pp. 7–14 (2010)
13. Greenwald, A.G., Poehlman, T.A., Uhlmann, E.L., Banaji, M.R.: Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. Journal of Personality and Social Psychology 97, 17–41 (2009)
14. Gawronski, B., Conrey, F.R.: The Implicit Association Test as a measure of automatically activated associations: Range and limits. Psychol. Rundsch. 55, 118–126 (2004)
15. Meschtscherjakov, A., Kluckner, P., Pöhr, F., Reitberger, W., Weiss, A., Tscheligi, M., Hohenwarter, K.H., Oswald, P.: Ambient persuasion in the factory: The case of the Operator Guide. In: Advanced Semiconductor Manufacturing Conference, pp. 1–6. IEEE (2011)
16. Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., Schmitt, M.: A meta-analysis on the correlation between the implicit association test and explicit self-report measures. Pers. Soc. Psychol. B 31, 1369–1385 (2005)
17. Law, E.L.C.: The measurability and predictability of user experience, pp. 1–10. ACM (2011)
18. Väänänen-Vainio-Mattila, K., Roto, V., Hassenzahl, M.: Now let's do it in practice: user experience evaluation methods in product development. In: Extended Abstracts on Human Factors in Computing Systems, CHI 2008, pp. 3961–3964 (2008)
19. Coan, J.A., Allen, J.J.B.: Handbook of emotion elicitation and assessment. Oxford University Press, USA (2007)
20. Vyas, D., van der Veer, G.C.: Rich evaluations of entertainment experience: bridging the interpretational gap. In: Proceedings of the 13th Eurpoean Conference on Cognitive Ergonomics: Trust and Control in Complex Socio-Technical Systems, pp. 137–144 (2006)
21. Bartlett, F.C.: Remembering. Scientia, pp. 221–226 (1935)
22. Hasher, L., Griffin, M.: Reconstructive and reproductive processes in memory. Journal of Experimental Psychology: Human Learning and Memory 4, 318–330 (1978)
23. Kahneman, D., Diener, E., Schwarz, N.: Well-being: The foundations of hedonic psychology. Russell Sage Foundation Publications (2003)
24. Loftus, E.F.: Make-believe memories. American Psychologist 58, 867–873 (2003)
25. Mitchell, K.J., Johnson, M.K.: Source monitoring: Attributing mental experiences. The Oxford Handbook of Memory, pp. 179–195 (2000)

26. Roediger III, H.L., McDermott, K.B.: Tricks of memory. Current Directions in Psychological Science 9, 123–127 (2000)
27. McNally, R.J.: Cognitive bias in the anxiety disorders. In: Nebraska Symposium on Motivation, vol. 43, pp. 211–250. University of Nebraska Press, Lincoln (1996)
28. Wells, A., Matthews, G.: Attention and emotion: A clinical perspective. Lawrence Erlbaum Associates, Inc. (1994)
29. Marlow, D., Crowne, D.P.: Social desirability and response to perceived situational demands. Journal of Consulting Psychology 25, 109–115 (1961)
30. Meschtscherjakov, A., Reitberger, W., Mirlacher, T., Huber, H., Tscheligi, M.: AmIQuin - An Ambient Mannequin for the Shopping Environment. In: Tscheligi, M., de Ruyter, B., Markopoulus, P., Wichert, R., Mirlacher, T., Meschterjakov, A., Reitberger, W. (eds.) AmI 2009. LNCS, vol. 5859, pp. 206–214. Springer, Heidelberg (2009)
31. Lee, H., Smeaton, A.F., O'connor, N.E., Smyth, B.: User evaluation of Físchlár-News: An automatic broadcast news delivery system. ACM Transactions on Information Systems (TOIS) 24, 145–189 (2006)
32. Aula, A., Majaranta, P., Räihä, K.-J.: Eye-Tracking Reveals the Personal Styles for Search Result Evaluation. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 1058–1061. Springer, Heidelberg (2005)
33. Payne, B.K., Cheng, C.M., Govorun, O., Stewart, B.D.: An inkblot for attitudes: Affect misattribution as implicit measurement. Journal of Personality and Social Psychology 89, 277 (2005)
34. Bassili, J.N.: Meta-judgmental versus operative indexes of psychological attributes: The case of measures of attitude strength. Journal of Personality and Social Psychology 71, 637–653 (1996)
35. Hynes, G., Monaghan, F., Thai, V., Strang, T., O'Sullivan, D.: SAGE: An Ambient Intelligent Framework for Manufacturing. In: 9th IFAC Symposium on Automated Systems Based on Human Skill and, Nancy, France (2006)
36. Steffens, M.C., Konig, S.S.: Predicting spontaneous big five behavior with implicit association tests. Eur. J. Psychol. Assess. 22, 13–20 (2006)
37. Cunningham, W.A., Preacher, K.J., Banaji, M.R.: Implicit attitude measures: Consistency, stability, and convergent validity. Psychol. Sci. 12, 163–170 (2001)
38. McConnell, A.R., Leibold, J.M.: Relations among the implicit association test, discriminatory behaviour, and explicit measures of racial attitudes. Aust. J. Psychol. 53, 121 (2001)
39. DeCoster, J., Claypool, H.M.: A meta-analysis of priming effects on impression formation supporting a general model of informational biases. Personality and Social Psychology Review 8, 2–27 (2004)

Context Awareness in Ambient Systems by an Adaptive Multi-Agent Approach

Valérian Guivarch, Valérie Camps, and André Péninou

Institut de Recherche en Informatique de Toulouse
{Valerian.Guivarch,Valerie.Camps,Andre.Peninou}@irit.fr
http://www.irit.fr

Abstract. In the field of ambient systems, the dynamic management of user context is needed to allow devices to be proactive in order to adapt to environmental changes and to assist the user in his activities. This proactive approach requires to take into account the dynamics and distribution of devices in the user's environment, and to have learning capabilities in order to adopt a satisfactory behaviour. This paper presents *Amadeus*, an Adaptive Multi-Agent System (AMAS), whose objective is to learn, for each device of the ambient system, the contexts for which it can anticipate the user's needs by performing an action on his behalf. This paper focuses on the *Amadeus* architecture and on its learning capabilities. It proposes some promising results obtained through various scenarios, including a comparison with the Multilayer Perceptron (MLP) algorithm.

Keywords: Context, evolution, adaptation, learning, ambient intelligence, self-organization.

1 Introduction

Initially confined to the use of a computer, informatics has evolved significantly in recent years with the growth of mobile and integrated technologies. The information is not transmitted to a single unit under the control of a user, but over many physical devices distributed in the real environment that provide information or services to the user. Ambient systems are characterized by their dynamic and their complexity: they are composed of many heterogeneous devices distributed in a large physical environment, and these devices can appear and disappear at run-time. So, ensure that ambient systems have the capability to be proactive towards the user is a topical and an important issue. When saying "to be proactive", we mean to give the system the ability to self- adapt the user's needs and to anticipate them in order to provide him a relevant service (to perform an action, to provide assistance, recommendation, information, etc.). Such an objective requires deciding by using dynamic learning the relevant behaviour that has to adopt by the system in particular situations corresponding to precise contexts. In the framework of our ongoing work, the goal is that the system learns what actions the user does in what contexts in order to be able to do these actions later on behalf of the user.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 129–144, 2012.

The rest of this article relies on this view of proactiveness. It focuses on the context awareness in ambient systems and, especially on learning of contextual situations where a proactive action is relevant. After having proposed our definition of "context" and of "context-aware", we give a brief overview of existing learning methods including the Adaptive Multi Agent System (AMAS) approach, with which we designed our system *Amadeus*. We then present the *Amadeus* architecture and functionalities. Finally, we illustrate and evaluate the operation of *Amadeus* through three scenarios, including a comparative study with the Multilayer Perceptron.

2 Context and Learning

The large number of devices involved as well as their heterogeneity, the numerous interconnections between these devices, their dynamics (appearance or disappearance of devices at run-time, with the displacement of the user for example) make very difficult to precisely define the system, and therefore to centralize its management. Making an exhaustive list of all the situations that the whole system may be faced with is impossible. That is why it becomes necessary to make such systems able to adapt by themselves to the user's context. Such systems are called "context-aware". [2] [11]

Due to the specific dynamics of ambient systems, the environment in which ambient systems are immersed is [10] (i) inaccessible: for each device of an ambient system, only a part of all the information can be collected; (ii) continuous: the number of possible actions and perceptions is not discrete in a real environment; (iii) non deterministic : the result of an action performed in a real environment cannot be determined in advance with certainty and (iv) dynamic: the environment may change under the action of the system, but it can also evolve under the action of other processes, in particular users.

Our system aims to continually interact with its environment, to establish by itself the most satisfactory functionality, and to adapt itself using its environment feedbacks. Lemouzy [7] defines autonomous adaptation as *"the ability for system to dynamically change in real time and without the intervention of an external entity, its way of acting according to the observed behaviour in the environment in response to its actions and in order to provide a service, a stable functionality through time, despite changes in the environment"*. To obtain this result, we are interested in different existing learning algorithms, to assess their relevance in the field of ambient systems.

2.1 Existing Learning Methods

Supervised Learning. Supervised learning processes a set of objects having n dimensions, each of these objects being associated with an output (a class, value, etc.). The objective is to use these objects as the basis of examples to determine a function that can deduce the output value for each new perceived object. To achieve this learning, the system has therefore, in each received case during the learning phase, a feedback from its environment that explicitly expresses the value of the expected output. One of the most classical supervised learning algorithms is the Multilayer Perceptron (MLP) [9]. This is an algorithm based on neural network modeling. It operates in two phases: the learning phase, during which the algorithm determines the desired function, then the use phase, where it applies this function on new input cases. The quality of this learning depends on the number of samples received during the learning phase, and this learning cannot be changed at runtime with new examples, unless to restart learning from the beginning.

Unsupervised Learning. The unsupervised learning [12], such as the supervised learning, aims to treat a set of objects having n dimensions. However, these objects are not associated *a priori* to an expected output. The goal is to search for natural structures in the data. Most often, the goal is to gather together objects into clusters such as each object (i) is as similar as possible to other objects in its cluster, and (ii) is as different as possible to objects belonging to other clusters.

Reinforcement Learning. Reinforcement learning algorithms [13], whose the most classical example is the Q-Learning [15], are algorithms aiming to learn, for all possible states of the system, the best action among feasible actions. This type of learning requires only a feedback from its environment that expresses the usefulness of its activity. A reinforcement learning algorithm then attempts to establish optimal control over its environment in order to maximize the perceived utility. The main difficulty of this type of learning is to find the right ratio, for a given situation, between exploration of new actions and performing of yet-known actions with known effect.

Discussion. In ambient systems framework, by combining user's actions with contextual situations indicating where/when these actions took place, it is possible to use supervised learning algorithms to create a cases base that will be used to learn the expected functionality. However, this approach has some limitations. More precisely, in ambient systems, it is necessary to be able to self-adapt to unexpected situations, such as the appearance or disappearance of devices, the user's preferences evolution, etc. An algorithm such as the MLP can achieve the functionality to be learned only if it restarts its learning from the beginning.

If we consider these limitations, reinforcement-learning algorithms seem to be a good solution. Several variants have been proposed to overcome the issue of the long time required to learn. For example the context management system proposed by Zaidenberg [17] uses an indirect reinforcement-learning algorithm. However, a learning algorithm based on a "trial/error" process does not seem appropriate for the context-aware control of an ambient system, because it requires the exploration of new solutions that may be bad solutions (inappropriate action in such contexts). And in this application framework, any error may disrupt the user who may finally reject the system.

Finally, unsupervised algorithms seem to suit the problem requirements because they do not require explicit feedback from their environment, but they are generally applied to classification problems. Combining contextual situations will not help us to determine what action needs to be done.

For our problem, it is necessary to design a system able to combine some characteristics of these different algorithms. In particular, we seek to establish a system able, for each perceived context, to determine the action to perform. This system must be capable of introspection, allowing it to self-assess at any time without regular and necessary explicit feedback of its environment, and to adapt when necessary (if a new context occurs for example) in order to modify its learning at runtime. This objective leads us to use the Adaptive Multi Agent Systems (AMAS) approach.

2.2 The AMAS Approach

Adaptive Multi Agent Systems (AMAS) [4] are based on a local approach to design complex systems solving problems for which an *a priori* known solution does not exist. This approach splits any systems into agents and focuses on defining the local behaviour of agents to be truly adaptive, while ignoring the purpose of the overall system, but ensuring that the collective behaviour is the one expected, i.e. the system is "functionally adequate". To this end, agents must have a local cooperative behaviour. Our definition of cooperation is not a conventional one (resource sharing or the fact of working together). It is based on three local meta-rules that have to be locally checked by every agent and that the designer has to instantiate depending on the problem to be solved: (c_{per}) any signal perceived by an agent must be understood without ambiguity; (c_{dec}): each information coming from its perceptions have to be useful to its reasoning, (c_{act}): its reasoning has to lead the agent to perform actions useful for the others and the environment.

In the AMAS approach, the agents above all have to anticipate, to prevent or to repair any Non Cooperative Situations (NCS). A NCS appears when an agent does not locally check at least one of the three previous meta-rules. Several generic NCS have been highlighted: *incomprehension* and *ambiguity* if (c_{per}) is not verified, *incompetence* and *unproductivity* if (c_{dec}) is not checked and, finally, *concurrency*, *conflict* and *uselessness* where (c_{act}) is not verified.

This approach has important methodological implications: as a matter of fact, designing an AMAS needs to define and to give cooperation rules to the agents. In particular, the designer has, for a given problem, (i) to define the nominal behaviour of an agent, then (ii) to deduce the NCS to which the agent can be faced with, and finally (iii) to define the actions that the agent has to carry out in order to come back to a cooperative state.

This approach has been used to solve several types of problems related to different areas: the real-time profiling [6], the bioprocesses control [14], etc. The AMAS capability to solve complex, dynamic and distributed problems makes them relevant to solve in ambient system the problem of proactive behaviour learning according to the user.

3 Proposed Approach : *Amadeus*

We want to design a system that is able to make any ambient system context-aware, especially regarding the adaptation of the system itself to dynamically adjust its functionality at runtime. We suppose that an ambient system consists of a set of devices, themselves consisting of sensors and effectors. The traditional approach to design a context manager for ambient systems is to design a tool that is able to collect data from various sensors, then to structure them to make them usable by applications that will manage these data to act correctly on the effectors. Our approach aims to give to ambient systems the ability to perceive and directly use contextual data. To determine the best behaviour without constantly requiring the user's approval, these systems must be based on actions performed by the user. Indeed, we consider that when a user performs an action in a particular situation, he performs it in order to increase his satisfaction. The user's satisfaction here is considered as the user's requirements regarding the state of the system (the devices).

We designed *Amadeus*, a multi-agent system that can be seen as an "ambient" application in the sense that it has to be distributed on several devices composing the ambient system. Thus, we create an *Amadeus* instance for each device. Each instance is responsible for collecting the local context of the associated device, and to adapt its behaviour depending on its context. For this, *Amadeus* first observes and learns (without *a priori* knowledge) how the user uses each device and in what contexts, in order to do any recurring action on behalf of the user. Any new action of the user can modify or improve the learning process because learning is continuous during the system functioning.

Each instance of *Amadeus* collects directly, through the sensors of the associated device, several data called "contextual data". However, for each device, contextual data are not necessarily limited to those it directly perceives. So, the first research topic for the design of *Amadeus* focuses on the sharing of data between instances of *Amadeus*. This research topic includes a very important point: the study of the relevance of a piece of data for each instance of *Amadeus*. As a matter of fact, in a realistic ambient system, the number of possibly perceived data is very important, whereas only a part of this data is included in the device context. Moreover, a great part of these data may be useless for the learning process of the device. So, in order to avoid damaging the learning with useless data, it is important to perceive all necessary information and no more information than these. However, this point will be study later in our work, and we consider for now only simple case studies where only relevant data will be available.

Then, in possession of these data, *Amadeus* must determine the user's satisfaction. We define it as a data included in [0,1] that represents the user's satisfaction regarding the state of the device effectors and according to the context. The second research topic focuses on the assessment of the user's satisfaction according to the state of the device and according to the context. One way to evaluate this value is to use a user profile explicitly defined by the user. We limit ourselves to this solution for now. However, this second research topic will be studied later too in our work, because we consider a static profile as a weakness in our system, that has to be able to adapt its knowledge about the user.

Finally, *Amadeus* has to learn, for each context, what action it has to perform in order to maintain the user's satisfaction as high as possible. This paper focuses on this last point, and presents only a part of the multi-agent system that we designed. This part has to learn the best action to perform according to the context. We therefore consider that the agents in charge of distributing and evaluating data and those in charge of evaluating the user's satisfaction have already been implemented and are functional.

3.1 Case Study

To illustrate *Amadeus* functioning, we describe a simple case study. *Amadeus* was of course designed independently of this case study that is just one example of application that we will use in order to present the first obtained results. We consider an apartment in which the living room is equipped with a lamp and an electric shutter. The tenant can turn on or turn off the lights and open or close the shutter. Moreover, the living room is equipped with a light sensor and a presence sensor.

On the one hand, evaluating our system in real ambient system situations implies considerable effort and time, because it requires to find a volunteer user during many days and to install *Amadeus* on real devices meanwhile we need to tune it regularly. Although a real experiment is planned for the final version of *Amadeus*, we need a faster solution in order to evaluate our system while it is under development. On the other hand, because our system needs to interact with its environment in order to perform its processing, using a real data set of people in a living lab (e.g. MavHome project [1]) does not allow observing the consequences of the *Amadeus* actions (since the data set is off-line). Finally, we choose to use a simulator, in order to simulate a user in an ambient system connected with our system. This simulator allows us : (i) to work with dynamic scenarios where *Amadeus* actions can be faced with the behaviour of the environment (user and other devices), (ii) to "accelerate" time up to play scenarios of about 50 days in a few minutes under programmer control with a graphical interface, (iii) to tune *Amadeus* and retest it very quickly.

The simulator allows describing some rooms of a house, devices inside them, and values returned by sensors. Moreover, this simulator allows the description of the simple behaviours of a simulated user in a virtual ambient system; for example, entering and leaving the room, opening shutters or turning on light when necessary, etc.

For our study, we described a single user with a simple behaviour: he walks in the apartment from one room to another while making sure that the luminosity is suitable when he is in the living room and taking care of energy consumption. In practice, this means that when the user is in the living room, if the luminosity is too low he will open the shutter; then he will turn on the lamp if the luminosity is not enough. Conversely, if the luminosity is too high, the user will first turn off the lamp if it is on, and close the shutter otherwise. Finally, when the user is not in the living room, he does not care about the state of the shutter, but he has to ensure that the lamp is not uselessly turned on. A random displacement of the user enables to change the moments and the duration where he is present (or not) in the living room. To improve the realism of the simulation, conditions associated to the user's actions are not strict but rather stochastic; for example, if the theoretical condition making the user turn on the light is luminosity greater than 55, then the real conditions will be randomly fixed between 50 and 60.

3.2 Required Proprieties

Many studies have led to the development of software tools able to "sensitize" the ambient systems to user's context. We summarize the properties that such software has to respect according to us:

Genericity: It has to necessarily be generic and integrate any type of data, either directly or by expanding some categories of contextual data if necessary. Indeed, a too strict categorization will exclude all contextual data that are not in these categories.

Distribution: A centralized management of the system, aggregating all the information to process them at best, seems to be the most natural solution. However, as explained by Euzenat [3], "with such a system, the scope of context management would be only efficient in a limited physical area", which contradicts the concept of ambient system. So, context data and management have to be distributed.

Openness: Because of the strong dynamics of ambient systems, it has to take into account the appearance or disappearance of devices at runtime, without having to be (re) configured.

Adaptation: A system whose behaviour is fully defined in advance according to different provided contexts will not be able to adapt to new situations, and will therefore be potentially inadequate. Moreover, it is impossible to list in advance all use cases of a dynamic and open system. The system has to be able to self adapt its behaviour during its functioning. This adaptation necessarily requires introspection and self-assessment capabilities in order to request the least as possible the user; the system has to assess itself if its behaviour is correct or not.

3.3 *Amadeus*

An instance of *Amadeus*, associated with a device of the ambient system (Figure 1), consists of several types of agents. The *data* agents represent contextual data that are either collected from local sensors of the device or received from other instances (other devices). The *user* agent has to estimate at any time the user's satisfaction level (defined as a value in [0, 1]) with regard to the state of the device. For this study, this *user* agent is simulated and can be considered as using a user profile in order to determine, at any time, the user's satisfaction regarding the state of the device. The behaviours of the *context* and *controller*

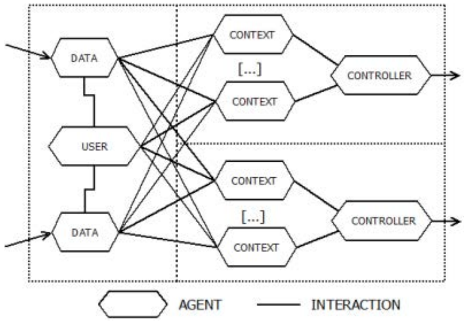


Fig. 1. General structure of the system

agents that ensure the device adaptation in order to maintain a user's satisfaction as high as possible are presented in this paper.

General Functioning. A *controller* agent is associated with each effector of the device. It aims at determining, at every moment, what is the best action to apply on this effector in order to increase as much as possible the user's satisfaction. For this, it has a set of *context* agents that provides information on the actions it can perform and the consequences of these actions on the user's satisfaction. These *context* agents aim at determining the effects of a particular action on the user's satisfaction, but also at identifying situations where these predictions are indeed correct.

***Context* Agent.** The objective of a *context* agent is to propose an action for a given context and a forecast about the consequences of this action on the user's satisfaction: the action may increase the satisfaction up to a certain value. A *context* agent has, for each contextual data, a range of values representing its validity range. Each range of values has a validity state that is considered as valid if the current value of the data is included within the bounds of this range of values. This requires that all perceived data are numeric values. This point will be discussed in section 4.4. Each bound of ranges of values is implemented with an Adaptive Value Tracker (AVT) [7]. An AVT is a software component for finding, we can say learning, the value of a dynamic variable in a given space through successive feedbacks.

The *context* agent has a validity status that is valid when all its ranges of values are valid (invalid otherwise), and a selection state that can take the status of selected or unselected by the *controller* agent. Finally, a *context* agent owns an action proposition that is composed of the proposed action itself (i.e. the state to affect to the associated effector), a predictive value on the effect of this action on the user's satisfaction that is implemented by an AVT, and a value included in [0, 1] representing the confidence level of the prediction.

Algorithm 1. *context* agent behaviour

```
Require: ListValuesRanges, ListDataUpdate, ContextState ∈ {valid, invalid}, ProposedAction ∈ ℝ, Forecast ∈ [-1;1], Confidence ∈ [0;1], ContextSelection ∈ {selected, unselected}
ListValuesRanges ← udpate_values_ranges(ListValuesRanges,ListDataUpdate)
if (∀ vr ∈ ListValuesRanges, state(vr) EQUAL valid) then
   ContextState ← valid
   send_proposition_at_controller(ProposedAction,Forecast,Confidence)
end if
ContextSelection ← update_Context_Selection()
if ContextSelection EQUAL selected then
   UserSatisfaction ← actual_User_Satisfaction()
end if
if ContextSelection EQUAL unselected then
   if UserSatisfaction + Forecast EQUAL actual_User_Satisfaction() then
      Confidence ← increase_confidence(Confidence)
   else
      Confidence ← decrease_confidence(Confidence)
      ListValuesRanges ← adaptation_values_ranges(ListValuesRanges)
   end if
end if
```

The algorithm 1 proposes the nominal behaviour of a *context* agent. After the update of its ranges of values, it checks if it became valid, and if it is the case, it sends its action proposition to the *controller* agent. If it is then selected, it saves the current value of the user's satisfaction. Then, when it becomes unselected, it can compare its forecast with what actually happened. Thus, if the value of the user's satisfaction is equal to the forecast, the agent increases its confidence. Otherwise, it reduces its confidence, and changes its ranges of values; thus if the same situation occurs again, the *context* agent will not be valid and will not make its proposition wrongly.

We consider that the confidence of a *context* agent informs about the relevance of the agent's behaviour. This confidence is calculated according to a function that evaluates a confidence level T_{t+1} at time t+1 based on its previous confidence level T_t t time t, a feedback F between 0 and 1, and a parameter λ that represents the impact of the feedback to calculate the new confidence level. This function is the following:

$$T_{t+1} = T_t * (1 - \lambda) + F * \lambda$$

To increase the confidence value of the *context* agent, we use a feedback close to 1, whereas to decrease it, we use a feedback close to 0. The λ parameter has been experimentally determined (after several simulations), and we obtained good results with a value equal to 0.1. This value means that the last feedback influences 10% of the confidence value. Nevertheless, we consider a fixed parameter as a weakness for an adaptive system; that is why we plan to make this parameter able to self-adapt.

***Controller* Agent.** The *controller* agent starts its life cycle by collecting proposals from *context* agents. Each proposal contains a description of the proposed action, an estimation of the impact that this action will have on user's satisfaction, and the confidence value that the *context* agent gives to its proposal.

Then, the *controller* agent has to estimate which proposal is the most interesting to increase the user's satisfaction. For that, it first ensures that two *context* agents do not offer the same action at the same time with different forecasts. If this happens, the *controller* agent solves this conflict by considering only the most confident agent. Once this choice is made, the *controller* agent evaluates what is the best action among those proposed by the remaining *context* agents. This action can be to change the status of an effector or to maintain this effector in its current state, because a *context* agent can propose the action "preserve the current state". The *context* agent associated with the best action proposal is then selected, while the current *context* agent is "deselected".

If a user action occurs on the device, which means that he wants the current state of the device to be changed in order to fit his satisfaction, the *controller* agent creates a new *context* agent. In the same way, if no *context* agent is selected and the users does not change the state of the device, which means that the user wants the current state of the device to be maintained in order to fit his satisfaction, the *controller* agent creates a new *context* agent.

In all cases, this new created *context* agent takes the action performed by the user, or the action to change nothing in the device, to create its action proposition. Its forecast value is equal to the difference between the observed user's satisfaction before and after this action, and the confidence value is equal to 0.5.

To summarize, the learning of the user's behaviour is based on the interactions between *context* and *controller* agents. It is initiated through the creation of *context* agents (when no *context* agent takes into account current state of the device), and it is refined through the self-adaptation of ranges of values (adaptive value trackers) and the adjustment of the confidence value of the *context* agent's (with feedbacks on the current situation). This allows *Amadeus* to further act on behalf of the user.

4 Results and Analysis

We propose to evaluate the performance of *Amadeus* by using the case study described in section 3.1. We have implemented *Amadeus* with SpeADL/MAY [8], which allows the definition and implementation of execution platforms of agents enabling the exploitation of reusable components.

We simulate a simplified ambient environment, in which we describe various elements and their evolution: the time, the outside light, the living room brightness, the coordinates and dimensions of the living room, a lamp and an electric shutter. The simulator also includes a user with his coordinates indicating his position in the environment. Moreover, we add a light sensor recovering the brightness of the living room, and a presence sensor able to determine if the user's coordinates are included in the living room coordinates. The simulated user's behaviour is described in section 3.1.

The objective of this evaluation is to show the ability of our system to determine, by observing and learning from user's actions, how the user maintains his highest satisfaction. Our system acts and learns continuously. By assigning a quite regular behaviour to our virtual user, *Amadeus* will gradually perform actions on the behalf of the user.

4.1 Learning User's Actions

Initially, we ran our simulator without connecting it to *Amadeus*. We generated a simulation of 50 days, in order to observe how the learning is realized over time according to different levels of luminosity. One cycle of the simulator is equal to one minute. The user can achieve on average ten actions per day in different light conditions. Because the user's movements are randomly simulated, and despite the static rules associated to his behaviour, we observed some differences in the number of actions made per day, varying from 4 to 16. The figure 2 shows the number of user's actions during a simulation of 50 days.

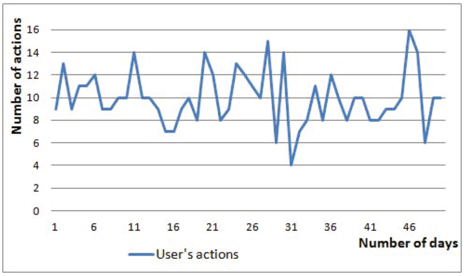


Fig. 2. Simulation without *Amadeus*

The simulator is able to generate the same simulation several times. This allows us to compare the results with and without *Amadeus*. We associate an instance of *Amadeus* to the lamp and an instance of *Amadeus* to the shutter. A *controller* agent is created and associated to each of these instances during the instantiation. As said before, a *user* agent is simulated by a function giving the user's satisfaction (calculated by the simulator). Figure 3 shows the user's actions during the same simulation, but with *Amadeus* instances connected to the devices of the virtual ambient system, and the actions performed by *Amadeus*.

The first day, *Amadeus* did not act at all while the user has performed 9 actions. The second day, the number of user's actions has sharply decreased, most of them being carried out by *Amadeus*. If we compare the number of actions with or without *Amadeus*, we notice a decrease from 53 to 13 user's actions for the first 5 days, so a decrease of about 75% of user's actions. If we make the comparison over a period of 50 days, the number of user's action varies from 499 to 29, so a decrease of about 95%. The only actions performed by the user

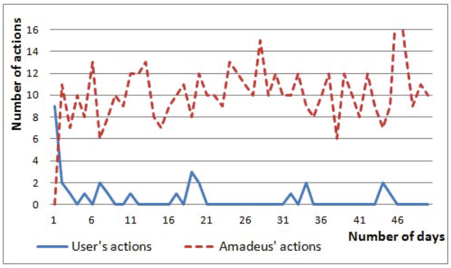


Fig. 3. Result of the learning of user's actions by *Amadeus*

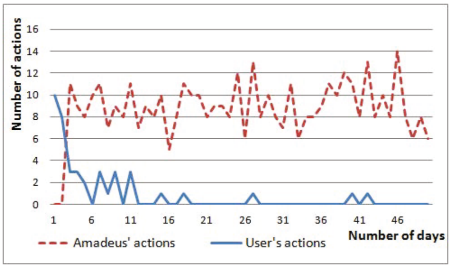


Fig. 4. Result of learning of user's actions by MLP

are special situations not yet encountered by the system; the recurrent user's actions tend to gradually disappear.

The performances of *Amadeus* learning capabilities can be compared with those of a more classical algorithm, such as the Multilayer Perceptron (MLP). Indeed, this algorithm allows to process non-linear phenomena (non-linearly separable classification), which is the general the case of our application.

We used the Java library *Weka* [5] to implement the MLP. This algorithm works in two phases: the learning phase and the operation phase. To assess accurately the performance of this algorithm, it is therefore necessary to establish, for a given simulation, the number of required days for the learning phase. During this phase, each situation is associated with the user's action and stored. If the allocated number of days for learning is too low, the learned functionality is not relevant. However, if the allocated number of days is too important, besides the fact that we want to learn as fast as possible, we can end up with an overlearning effect. In this case, the learned functionality is too specific, and is no longer able to generalize correctly; a bad behaviour results in a new situation. For our case, an empirical study has shown that two days of simulation are enough to learn a proper functionality.

Figure 4 illustrates the results produced by applying the MLP in our simulation. The first two days, the user performs his actions while the MLP observed all situations that it has associated with user's actions. Then, the third day it comes into the operating phase; so it takes control of effectors and performs the actions on behalf of the user, so the user's actions greatly decrease.

If we compare the results obtained with the MLP algorithm and those obtained with *Amadeus*, we can see similar results with regard to the learning time. The learning with the MLP only took the first two days while the learning with *Amadeus* is more gradual, but occurs mainly during the first 5 days.

Let us notice that, if we strictly compare day by day, the total number of actions of the user and the system (either MLP or *Amadeus*, figures 3 and 4) is equal or greater than the total number of actions when the user is alone (figure 2). This is due to two effects. Firstly, the system may do some "wrong" action that the user will himself cancel. This will happen when learning is incomplete or may not cover the situation. Secondly, the system may do some action that the user would not have done in such situation but he nevertheless accepts the action. This will happen because user's preferences are not strict (see section 3.1).

Since the user has a stochastic behaviour (with random movement), we can generate different simulations with the same user's behaviour rules. So, we apply *Amadeus* on 20 simulations and, with the replay method of the simulator, we apply the MLP algorithm at the same simulations. Then, we compare the performances of these two algorithms. We consider the number of human actions as a relevant variable in order to evaluate these algorithms. As a matter of fact, lower is this number, better is the ability of an algorithm to act on behalf of the user. The table 1 shows the number of human's actions in each simulation, depending on the applied algorithm.

Table 1. Human's actions for 20 simulations, depending on the applied algorithm

Amadeus	22	24	23	23	23	23	24	25	24	23	24	22	24	24	23	22	24	21	25	24
MLP	30	25	30	29	22	26	23	52	33	20	31	34	20	23	22	35	25	29	39	34

In order to compare the performances of the two algorithms, we apply the test of Wilcoxon [16] that allows detecting significant differences in two sets of data. In our case, this test gives a p-value of 0.0056 which is lower than the classical threshold of 0.05. So, the test of Wilcoxon shows a statistically significant difference leading us to consider the interest of our approach. Indeed, using *Amadeus*, the overall average number of user's actions is smaller than with MLP.

4.2 Adaptation to Changing Preferences

We conducted a second study, based on a similar scenario to the first study, but in which we introduce a change at runtime. Thus, at the end of the fiftieth day, whereas *Amadeus* has stabilized itself, we perturb the system by changing the user preferences. His behaviour is still similar (moving in the room), but his preferences in terms of brightness change. He then wants a luminosity globally lower, which should encourage him to close the shutter or turn off the light sooner, and open the shutter and turn on the lamp later.

Figure 5 shows the effects of this change on the user actions and those of *Amadeus*. The fiftieth day, we can see that the user begins to act again, because the functioning of *Amadeus* does not answer any more its expectations. Consequently, *Amadeus* almost does not act because the actions it knows efficient in order to satisfy the user were no longer suitable. But quickly, we can see that *Amadeus* behaviour adapt itself at the new user requirements, and the number of user's actions decreases and aims again towards zero.

4.3 Appearance of a New Device

In this third study, we repeated the simulation with a user looking for a low light, because this simulation supports the regular use of the shutter. Indeed,

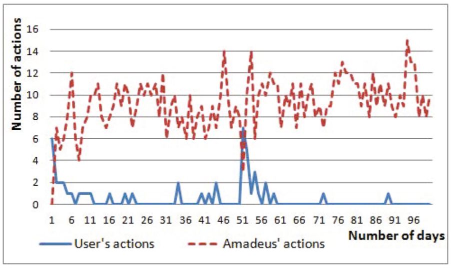


Fig. 5. Result of the adaptation to the changes of preferences

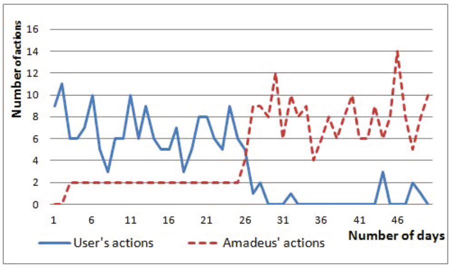


Fig. 6. Result of the integration of a new device

in this study, we begin our simulation by instantiating *Amadeus* at the shutter device, but instantiate *Amadeus* at the lamp device only after 25 days, for a total simulation of 50 days. The objective here is to add a device in *Amadeus* at runtime. We can see in the figure 6 that *Amadeus* has quickly learned when to open and to close the shutter, independently of the lamp (which it does not see). The user actions are then performed only towards the lamp. But once an instance of *Amadeus* is added to the lamp, the two *Amadeus* instances can learn the correct behaviour of the lamp and the shutter in order to decrease progressively user's actions.

4.4 Discussion

These three scenarios allowed us to make a first evaluation of *Amadeus*, and to ensure that it satisfies the properties identified in Section 3.2.

Genericity Property: Modelling the contextual data by means of ranges of values, we respect the property of genericity. Indeed, any new data being modelled in the form of its only numerical value, it can be directly used to establish the ranges of validity, without associating it a semantic or category. It is just necessary to convert some types of values in numeric format: a list of string (for example, a list of rooms where the user can be) can be considered as a set of booleans, and a boolean can be encoded as value 0 or 1. Our theory is that it is not necessary to make a pre-processing on perceived data in order to interpret them, for example to detect the activity of the user or to differentiate types of sensors. The correlation between data values and user actions can be detected without considering any semantic.

Adaptation Property: The first scenario was used to check, on still limited but promising cases, the *Amadeus* ability to learn without initial knowledge, the relevant behaviour of the ambient system. The second and third scenarios also showed its ability to adapt itself to changes at runtime. *Amadeus* therefore respects the property of adaptation when applied to our type of case study. The comparison between the MLP and *Amadeus* shows some better performance (while staying limited benefit), but our system is able to adapt at runtime.

Although it is possible to use a more classical algorithm, this implies to make processing in order to detect when the learning is degraded enough to reboot the learning phase. Moreover, it becomes then necessary to save previous data, but more we have data, more it takes time to make the learning. Filtering saved data raises other problems: how to decide which data can be deleted? An older data does not mean that it is less relevant, and user preferences changing can cause contradiction between data. It seems very complex to respond to all this questions independently of a specific use case, whereas *Amadeus* is designed in order to learn without saving previous data, only by constant adaptation to current situation at run-time.

Openness Property: The third scenario shows that when adding a new device in the perceptions of *Amadeus*, by associating it with a new instance, it was able to learn the existence of this new device dynamically without having to repeat the learning associated with other devices.

Distribution Property: The fourth property is partially respected because each device has its own instance of *Amadeus*. So, the system is distributed. However, for now, all the instances being aware of all of the information system, this property remains to be explored.

5 Conclusion and Future Work

In this paper, we presented the architecture and functioning of the adaptive multi-agent system *Amadeus*. In the field of ambient systems, *Amadeus* is able to learn, for each device of the ambient system, the contexts for which it can anticipate the user's requirements by performing an action on his behalf. We also presented an evaluation of this system through various case studies. This study shows us similar performances between a classical learning algorithm and our system, but it also shows us the performances of our system to adapt when faced to environment changes without having to reinitialize its learning.

However, several problems have not yet been resolved. First, we must complete our experiments on more complex case studies, especially when considering a larger number of devices and users, or users with more complex behaviour. In particular, we should study the impact of irrelevant information on the learning performed by an instance of *Amadeus*. Indeed, in a system as opened as an ambient system, data collected are far more numerous than those necessary for the proper functioning of each device. It is therefore necessary to learn, among the data collected, those belonging to the context of each device and those that do not belong to. Moreover, we should study the performance of *Amadeus* when it is applied in real cases with several users.

References

1. Cook, D.J., Youngblood, M., Heierman III, E.O., Gopalratnam, K., Rao, S., Litvin, A., Khawaja, F.: Mavhome: An agent-based smart home. In: Proceedings of the First IEEE International Conference on Pervasive Computing and Communications (PerCom 2003), pp. 521–524. IEEE (2003)
2. Dey, A., Abowd, G., Salber, D.: A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. Human-Computer Interaction 16(2), 97–166 (2001)
3. Euzenat, J., Pierson, J., Ramparany, F.: Dynamic context management for pervasive applications. The Knowledge Engineering Review 23(01), 21–49 (2008)
4. Georgé, J.-P., Gleizes, M.-P., Camps, V.: Cooperation. In: Di Marzo Serugendo, G., Gleizes, M.-P., Karageorgos, A. (eds.) Self-organising Software. Natural Computing Series, pp. 193–226. Springer, Heidelberg (2011)
5. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. ACM SIGKDD Explorations Newsletter 11(1), 10–18 (2009)
6. Lemouzy, S., Camps, V., Glize, P.: Real Time Learning of Behaviour Features for Personalised Interest Assessment. In: Demazeau, Y., Dignum, F., Corchado, J.M., Pérez, J.B. (eds.) Advances in PAAMS. AISC, vol. 70, pp. 5–14. Springer, Heidelberg (2010)
7. Lemouzy, S.: Systémes interactifs auto-adaptatifs par systémes multi-agents auto-organisateurs: application á la personnalisation de l'accés á l'information. Thése de doctorat, Université Paul Sabatier, Toulouse, France, juillet (2011)
8. Noel, V., Arcangeli, J.-P., Gleizes, M.-P.: Between Design and Implementation of Multi-Agent Systems: A Component-Based Two-Step Process (regular paper). In: European Workshop on Multi-Agent Systems (EUMAS), Paris (France), 16/12/2010-17/12/2010, page (electronic medium), Université Paris Descartes (Décembre 2010), http://www.univ-paris5.fr/
9. Rosenblatt, F.: Principles of neurodynamics. perceptrons and the theory of brain mechanisms. Technical report, DTIC Document (1961)
10. Russell, S.J., Norvig, P.: Artificial intelligence: a modern approach. Prentice hall (2010)
11. Schilit, B.N., Theimer, M.M.: Disseminating active map information to mobile hosts. IEEE Network 8(5), 22–32 (1994)
12. Sejnowski, T.J.: Unsupervised learning: foundations of neural computation. The MIT Press (1999)
13. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction, vol. 1. Cambridge Univ. Press (1998)
14. Videau, S., Bernon, C., Glize, P., Uribelarrea, J.L.: Controlling bioprocesses using cooperative self-organizing agents. In: Advances on Practical Applications of Agents and Multiagent Systems, pp. 141–150 (2011)
15. Watkins, C.J.C.H., Dayan, P.: Q-learning. Machine Learning 8(3), 279–292 (1992)
16. Wilcoxon, F.: Individual comparisons by ranking methods. Biometrics Bulletin 1(6), 80–83 (1945)
17. Zaidenberg, S.: Apprentissage par renforcement de modeles de contexte pour l'informatique ambiante (October 2009)

Towards Fuzzy Transfer Learning for Intelligent Environments

Jethro Shell and Simon Coupland

De Montfort University,
Gateway House, Leicester, United Kingdom
JethroS@dmu.ac.uk,
SimonC@dmu.ac.uk
http://www.dmu.ac.uk

Abstract. By their very nature, Intelligent Environments (IE's) are infused with complexity, unreliability and uncertainty due to a combination of sensor noise and the human element. The quantity, type and availability of data to model these applications can be a major issue. Each situation is contextually different and constantly changing. The dynamic nature of the implementations present a challenging problem when attempting to model or learn a model of the environment. Training data to construct the model must be within the same feature space and have the same distribution as the target task data, however this is often highly costly and time consuming. There can even be occurrences were a complete lack of labelled target data occurs. It is within these situations that our study is focussed. In this paper we propose a framework to dynamically model IE's through the use of data sets from differing feature spaces and domains. The framework is constructed using a novel Fuzzy Transfer Learning (FuzzyTL) process.

The use of a FuzzyTL algorithm allows for a source of labelled data to improve the learning of an alternative context task. We will demonstrate the application of an Fuzzy Inference System (FIS) to produce a model from a source Intelligent Environment (IE) which can provide the knowledge for a differing target context. We will investigate the use of FuzzyTL within differing contextual distributions through the use of temporal and spatial alternative domains.

Keywords: Fuzzy Logic, Transfer Learning, Intelligent Environments, Ambient Intelligence, Context-Aware.

1 Introduction

Intelligent Environments have emerged in a number of different implementations across applications and domains changing the way in which we approach and use our world. Military[3,19], domestic[11] and healthcare[24,23] applications are all emerging. The hugely dynamic implementations and fluid environments can cause issues as they yield inherently uncertain datasets. The dominant techniques of data mining[6,12] and supervised learning encounter issues when modelling

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 145–160, 2012.

these complex environments. Both methodologies focus on the use of training and target data from the same feature space and distribution, and additionally require the reconstruction of the model if that distribution alters. In applications such as IE's which are often dynamic and transient in nature, such data sourcing can be extremely difficult. The implementation of Transfer Learning[16,14], a process whereby information from one context can be learnt and used within another approach, could have large benefits.

To address the issues of modelling environments in the presence of uncertainty and noise, we propose a fuzzy logic based system. The use of fuzzy logic allows for the incorporation of approximation and a greater expressiveness of the uncertainty within the data[25]. Using fuzzy logic as a base, we propose a transfer learning framework to dynamically model target tasks within intelligent environments using a labelled source task. This process will hereby be referred to as Fuzzy Transfer Learning (FuzzyTL). Through the application of the framework on a real world data set we will show that FuzzyTL is able to model context variations, both spatial and temporal. Overall the contributions of this work are:

– A generic knowledge transfer framework that allows information from contextually differing environments to be incorporated into the modelling process.
– A novel adaptive fuzzy process enabling the absorption of the inherent uncertainty and dynamic nature of Intelligent Environments (IE's).

The paper is constructed as follows: Initially Section 2 gives an overview of Transfer Learning (TL) within the sphere of Intelligent Environments which is followed by Section 3 which offers an in depth description of the framework used within the Fuzzy Transfer Learning process. Section 4 looks at the experimentation carried out within this research through the application of the framework to a real world IE data set. The paper is concluded with Section 5 which offers a conclusion of the findings and a summary of future work.

2 Transfer Learning

Predominantly Intelligent Environments (IE's) are constructed using a large number of varying types of sensor ranging from temperature and humidity sensors within environmental monitoring [10] to Passive Infra-red Sensor (PIR Sensor) within smart home structures. The implementations of such networks result in a wide array of dynamic data sources. The quantities of sensors used, possibly in excess of 100pcs in a single deployment, can produce large quantities of data, and the uncertain and dynamic form of the environments make the construction of models extremely difficult. How ever, there is a need within the majority of machine learning techniques that the data used for training and testing to come from the same feature space. The collecting of data is often expensive, in both terms of time and cost, and within some applications this can be extremely difficult. Assisted Living (AL) domains exemplify such challenging environments. AL environments are especially difficult to model as each is extremely unique due to the requirements of the occupants and the services that are offered. The needs of each occupant result in varied sensor construction and placement within the residence. As AL becomes more sophisticated clients are being offered a multitude of services based on fall detection [15,20], medication monitoring[23] and activity recognition [24] with each needing its own unique specific requirements. The application of TL techniques within AL's and similar such environments can be seen to be highly beneficial.

2.1 Implementations of Transfer Learning within Intelligent Environments

The main focus of Transfer Learning (TL) and Knowledge Transfer research within the area of Intelligent Environments (IE) has been around the subject of Activity Recognition applications. The study of context-aware-systems that focus on human activity has been an increasing area of interest. Inroads have been made into the recognition of a number of basic activities. However, as the activities increase in complexity the inherent variability and uncertainty in the execution of tasks by differing people over differing durations increases. This equally increases the difficulty in recognising the activity[2]. Within this sphere, machine learning techniques have been applied to model the highly complex activities. As with many other applications of machine learning, the use of these techniques requires a large quantity of labelled data and the resource that is associated with acquiring such information.

Rashidi and Cook[17] discuss the use of TL within smart homes to improve the accuracy of activity recognition. They introduce a methodology called Multi Home Transfer Learning (MHTL) which is based upon a location mining method for target activity recognition. The approach incorporates a multitude of information from structural, temporal and spatial features of the activities. The basis of the methodology is to mine the data accrued to produce the target activities[17].

Blanke and Schiele[2] also approach the problem of activity recognition using TL through the concept of a partonomy of relationships between activities and composites of activities. In their approach concept activities are referred to as the underlying activity entities, and composite activities as the high level entities for which these are made of. The partonomy is used to transfer activity events across domains to recognise composite activities.

Within the area of computer vision, Farhadi and Tabrizi[8] use TL to learn a discriminative model of activity in one view and transfer the features of that view into another that is lacking labels. The process employs a methodology to build an activity model using labelled examples and subsequently apply this to unlabelled data to model how appearance changes with aspect.

3 Fuzzy Transfer Learning Framework

The Fuzzy Transfer Learning (FuzzyTL) method that we present is contained within a framework structure. The key components can be seen in Fig. 1.

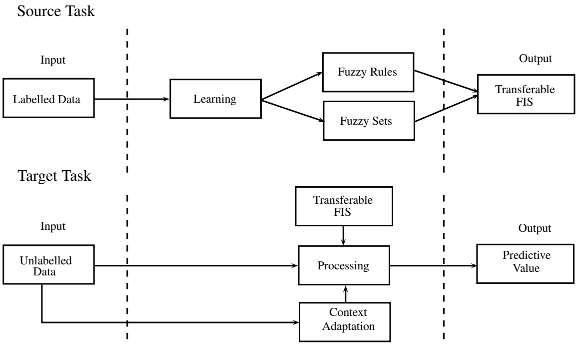


Fig. 1. Overview of the Fuzzy Transfer Learning Framework

The framework consists of three main elements:

1. Learning: Fuzzy rules and sets are created from numerical data.
2. Transfer: The fuzzy rules and sets created are transferred to the required task.
3. Adaptation: The FIS that has been created is adapted to the contextual changes in the target task.

3.1 Learning

The learning process employed within the FuzzyTL is based on an Ad-Hoc Data Driven Learning (ADDL) approach. ADDL is a learning procedure that uses the structure of the data to form the basis of the learning parameters. This is as opposed to the process of using experience to form the basis of the model. Within dynamic environments such as IE's this form of learning has been employed as it is able to model varying forms of time-series data[7]. A prominent form of ADDL that has emerged is Data Driven Fuzzy Modelling (DDFM). DDFM has found uses in a number of applications[5] in adapted forms and across varying domains. The qualitative nature of the modelling structure allows for the representation of knowledge in a linguistic, humanistic form along with an ability to approximate non-linear models with simple forms[5].

The first stage of the FuzzyTL process is the formation of the elements that constitute the FIS (Fuzzy Inference System). Through the use of numerical source data, fuzzy rules and fuzzy sets are formed via the use of an ADDL process. The employed method uses numerical data to form the sets and rules, a procedure based on an algorithm proposed by Wang and Mendel[22] and later expanded by Wang[21]. There are numerous benefits for using this type of method in the extraction of a model from numerical data. Its simplicity makes it easily understandable and the nature of the low computation required allows for a greater speed of implementation. The swiftness in the execution within the early stages of the preliminary fuzzy modelling process allows for subsequent adaptation of the model by other methods [4]. The framework builds upon the method by adding a novel rule reduction process. Through the addition of a fuzzy frequency measure, the reduction process endeavours to absorb anomalous data elements in the reduced rule base that is formed.

The data used within this primary stage is a labelled data set that has contextual similarities to the target task which is the focus of the classification process. The information gathered from this initial data set informs the rule and set construction.

The initial step is to divide each domain interval into fuzzy regions each containing the membership functions for the input or output variables. Taking a simple example containing two inputs (x_1,x_2) and one output (y), the process is to divide each of these domains by $2N + 1$ regions where N can be different for differing variables. In order to automate this step, the domain is divided equally based upon the minimum and maximum values of the interval and the defined number of regions. Fig. 2 shows the input and output domains divided into $2N + 1$ regions and labelled with linguistic values VS (Very Small), S (Small), M (Medium), L (Large) and VL (Very Large).

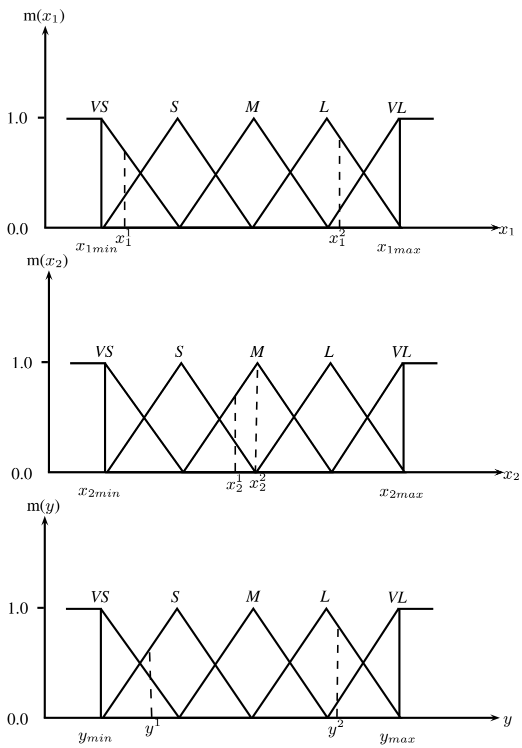


Fig. 2. Construction of Fuzzy Membership Functions

Within the application outlined within this paper, we applied triangular functions to the framework. The shape of the membership function can vary dependent on the application. Triangular membership functions were chosen for simplicity, generality and speed of execution. Differing functions are suitable for differing situations. The optimisation of membership functions through techniques such as Genetic Algorithms is a broad and well defined research area [9,1] which is beyond the scope of this paper.

In order to produce a defined set of rules from the numerical labelled data, each instance of the data has its overall maximum membership calculated. From this process each data instance creates an input-output rule. This can be demonstrated using a simple example using Fig. 2. Taking a value from each of the inputs, x_1 and x_2 and the output, y, the maximum degree of membership is taken from each domain. For input x_1 the membership values are 0.65 in VS and 0.35 in S with 0 in all other regions. The maximum is thus 0.65 in VS. Based upon this process we can produce the below linguistic rules.

$$
\begin{aligned}
(x_1^1, x_2^1, y^1) \Rightarrow \left[x_1^1 \text{ (0.65 in VS)}, x_2^1 \text{ (0.7 in M)}; y_1 \text{ (0.55 in S)}\right] \Rightarrow \text{ Rule 1} \\
\text{IF } x_1^1 \text{ is VS and } x_2^1 \text{ is M THEN } y_1 \text{ is S;} \\
(x_1^2, x_2^2, y^2) \Rightarrow \left[x_1^2 \text{ (0.75 in L)}, x_2^2 \text{ (1.0 in M)}; y_2 \text{ (0.75 in L)}\right] \Rightarrow \text{ Rule 2} \\
\text{IF } x_1^2 \text{ is L and } x_2^2 \text{ is M THEN } y_2 \text{ is L;}
\end{aligned}
$$

Each antecedent and consequent element of each rule that is constructed is mapped to the data value that was used to produce it. A two input - one output rule would have the construction: $(VL, c_1), (L, c_2), (S, c_3)$. The production of the fuzzy rules using the numerical data produces a rule base that is equal in size to that of the original dataset. As each individual data instance produces a single rule, this can become unmanageable in size. Additionally conflicts within the rule base can be constructed that result in an ineffective system. To reduce the rule base size and remove conflicts, each of the rules are assigned a weighted degree (d). The weight is based upon the maximum product of the individual inputs and outputs alongside a novel fuzzy representation of the frequency of the occurrence of each defined rule.

Initially rules that share the same antecedent values are combined into a single group, G_i. This group is a subset of the Full Rule Base (FRB). The second stage of the process is to form a universe of the rule frequency. Using the FRB, the minimum and maximum frequency values for each individual rule, based upon the antecedent components, is formed. Using these values, fuzzy sets are created following the rule extraction process outlined above. For each rule r, the membership value $\mu_A(z_r)$ is produced where z is the frequency occurrence of r within the FRB.

The final stage is to form the strength of the rule. This can be depicted as:

$$d^r = \prod_{m=1}^{s} \mu_{h_m^r}(x) \prod_{n=1}^{t} \mu_{k_n^r}(x) \mu_A(z_r) \tag{1}$$

The strength of the rule d^r is the combined membership of each antecedent and consequent value coupled with the membership value of the rule frequency, $\mu_A(z_r)$. Equation 1 shows the product of the input value x produced from the antecedent sets h and consequent sets k. This is combined with a membership value produced from the frequency input z. Using this value, a comparison is made eliminating all rules within G_i set except the rule with the highest overall value. By combining both a strength and frequency measure, a single anomalous rule instance with a high strength will have a reduced influence on the Reduced Rule Base (RRB). Rules that are frequently produced but have a very low strength equally are unable to heavily influence the overall rule base outcome.

Table 1. Frequency Based Rule Pruning

Time	$\mu(Time)$	Light	$\mu(Light)$	Temp	$\mu(Temp)$	Freq. of Rule	$\mu(Freq)$	WM	With Freq.
Very Low	0.30	Very Low	0.60	Very Low	0.80	1	0.20	0.14	0.03
Very Low	0.30	Very Low	0.60	Low	0.70	4	1.00	0.13	0.13
Very Low	0.30	Very Low	0.60	Low	0.70	4	1.00	0.13	0.13
Very Low	0.30	Very Low	0.60	Low	0.70	4	1.00	0.13	0.13
Very Low	0.30	Very Low	0.60	Low	0.70	4	1.00	0.13	0.13
Very Low	0.30	Very Low	0.60	Med	0.30	2	0.70	0.05	0.04
Very Low	0.30	Very Low	0.60	Med	0.30	2	0.70	0.05	0.04

Table 1 highlights an example of seven instances of data that each produce a single rule from environmental data (time, light and temperature sensors). Based upon the Wang-Mendel (WM) method, a rule would be produced based upon sets {*Very Low, Very Low, Very Low*} as this produces the highest overall product of membership values. However, this rule is based upon the lowest frequency of the occurrence implying a low number of instances of this rule type having occurred within the data set. Using the frequency measure, the lower valued rule {*Very Low, Very Low, Low*} is kept as the overall weighting is 0.13 as opposed to 0.03.

3.2 Transfer

The constructed fuzzy sets and the RRB form the main section of the transferable FIS. This can be seen in Fig. 1. These elements are used as the basis for the transferral of knowledge from the labelled dataset to the unlabelled, unknown data. The transfer process is encapsulated within the FIS through the use of the fuzzy sets and fuzzy rules that are bound within it. The source and target task datasets used in the FuzzyTL system can be contextually different, however, there is a need for them to be related. Environmental knowledge consisting of temperature, light and motion data from an office space can be transferred to a home consisting of a reduced sensor set. Variance in sensor type and accuracy can be incorporated into the system through preprocessing of the data. Variation in the intervals of the input and output domains produced through contextual change is accounted for within the system. The focus of this work is upon numerical data types, however there is scope to investigate the use of non-numerical data.

The transferable FIS forms the backbone from which the adaptation process assists in the optimising of the classification decision making process.

3.3 Adaptation

To produce an output, the first element of the process is the use of the unlabelled target task data to adapt the system. Two stages form the adaptation process as a whole:

1. Fuzzy Set Adaptation.
2. Fuzzy Rule Base Adaptation.

Fuzzy Set Adaptation. In order to absorb contextual differences in the source and target tasks, the FuzzyTL adopts a process of adapting the minimum and maximum values of the input universes. Taking each individual input instance of the dataset, the framework adjusts the minimum and maximum universe values according to any difference calculated between the transferable FIS and the new

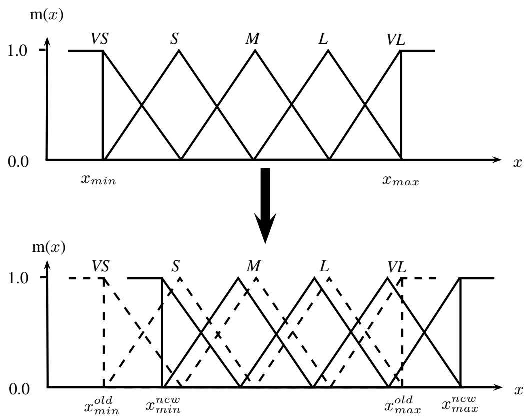


Fig. 3. Adaptation of Sets Based on New Minimum and Maximum Input Values

input values. The result is the adaptation of the sets that form the basis of the FIS. Each set can be adjusted both across the width of its footprint and in its position within the universe. Fig 3 shows the shifting of sets in a positive direction within the universe. All the sets are adjusted based on the new minimum and maximum values.

To remove the adverse effects of single anomalous reading within the system, a percentage change measurement is adopted to dictate the level of the minimum and maximum values used to influence the universe adaptation. The percentage change value is formed across a defined sliding window of input values. This value influences the rate of increment or decrement of the input universe. This can be depicted as:

$$p = 100 \times \left(\frac{f(x_n) - f(x_1)}{f(x_1)} \right) \tag{2}$$

$$y = \begin{cases} p > d & y = i \\ p < d & y = \big(i \times r\big)/100 \end{cases} \tag{3}$$

where in equation (2), p is the percentage change, x is the input variable within the sliding window n. In equation (3), y represents the universe adjusting value, d is a user defined threshold of the rate of change, r is a percentage value used for the reduction of y and i represents the input value.

Fuzzy Rule Adaptation. The adaptation of the RRB is based upon a two phased approach. In the initial phase, the FRB created by the Ad Hoc Data Driven Learning approach is used to search for a rule that produces a result when the required inputs are used. Each rule that is accrued through this search is added to the RRB. This draws on the over riding theme of transfer learning, drawing knowledge from the previously acquired data. The secondary step within the adaptation utilises the knowledge embedded within the labelled source task data to form new rules.

To achieve this goal, the framework captures the inputs from the target task and carries out a comparison with the known labelled data in order to approximate a rule based on the membership of the associated sets. This can be summarised within the Algorithm 1. Each target input value is compared to the source input value to locate the closest matching data item. The mapped antecedent "IF" component is captured and used to form the new rule. Using the captured data an additional comparison is formed between the source and target inputs using an n dimensional euclidean distance. Based upon the lowest distance value, the associated output consequent "THEN" is added to form the overall rule.

Algorithm 1. Rule Adaptation Process

Data: d is each data instance in target dataset TD
Data: e is each data instance in source dataset SD
Data: h is the antecedent component of the defined rule
Data: k is the consequent component of the defined rule
Data: p is the antecedent value of the rule
Data: q is the consequent value of the rule
Result: Adapted Rule Base α
input: a is each rule in the Full Rule Base β
input: b is each rule in the Reduced Rule Base γ
input: s is each input data variable of the each data instance
input: t is each output data variable of the each data instance
input: c minimum difference between each data item
input: l minimum euclidean difference between each data item

```
foreach d_s do
    foreach e_s do
        Find the smallest value within the source data.
        c ← min(δ(d_s, e_s));
        Produce the smallest combined euclidean distance between the
        source and task data.
        l ← min(sqrt(Σ_{i=1}^{n} (e_s − d_s)^2));
    end
end
foreach a in β do
    Output the antecedent set with the smallest source data value.
    h_a ← (β_1)^{-1}(c_s);
    p_s ⊕ (h_a, c_s);
    Output the consequent set with the smallest euclidean distance value.
    k_a ← (β_1)^{-1}(l_t);
    q_t ⊕ (k_a, l_t);
    Produce a new rule based on the antecedent and consequent sets.
    New Rule r = {p_s, q_t};
    if r ∉ γ then
        α = γ + r
    end
end
```

Through this process the adapted rule base is formed that is specific to the target task but based upon the knowledge within the source task.

3.4 Defuzzification

The final stage of the process within the framework is the production of a crisp output via a process of defuzzification. Many methods of defuzzification have been discussed and as such an in depth review is beyond the scope of this paper.

A centroid defuzzification strategy is used though there are a number of others that are applicable based on the context of the problem[18].

The centroid method can be expressed as:

$$z = \frac{\sum_{i=1}^{K} m_{o^i}^{i} \bar{y}^i}{\sum_{i=1}^{K} m_{o^i}^{i}}$$

where m_{o^i} denotes the degree of output using a product t-norm, $\bar{y}^i$ is the centre of the fuzzy region and K is number of fuzzy rules in the rule base.

4 Application

To evaluate the use of the Fuzzy Transfer Learning framework, a real world intelligent environment data set was chosen to demonstrate the applicability of the spatial and temporal contextual transfer process. The framework was constructed and run using C++ via Code:Blocks (Version 8.02) and compiled through GNU GCC on Ubuntu LTS Version 10.04.

4.1 Sensor Data and Construction

The data used is based upon information collected from 54 sensors deployed in the Intel Berkeley Research Laboratory between the 25th February and the 5th April, 2004. The network used XBow Micra2dot weatherboard based nodes to record environmental data. Four parameters were measured: time-stamped temperature (degrees Celsius), humidity ranging from 0-100%, light in Lux, and residual power of each sensor unit expressed in Volts. The data was collected using the TinyDB in-network query processing system built onto the TinyOS platform which recorded information every 31 seconds[13]. The layout of the nodes can be seen in Fig.4. A section of network was identified across the

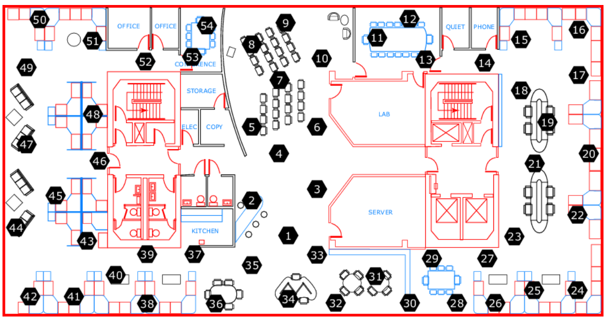


Fig. 4. Diagram of Intel Laboratory Showing Placement of Wireless Sensor Nodes from [13]

laboratory to examine the influence of variations in the spatial context and the transfer of this knowledge. Equally there was a desire to examine temporal contextual changes within the domains and so the influence on the FuzzyTL system as a result. To achieve both of these, the output of Sensors 7, 9, 12 , 24 , 34 , 42 and 51 where examined across five days from 28th February to 3rd March 2004. To form the source data for the model, light , time and temperature readings were taken from a single sensor (sensor 7) on a single day (28th February 2004). Using this data, target task datasets were created across the seven sensors within the seven day time span (omitting sensor 7 on the 28th February). This produced 48 separate contextual instances.

4.2 Experimentation Construction

For this set of experiments we segregated the domains of the sensor inputs, time (t) and light (l) and the output temperature (tmp) into equal regions with each section defined as $N = 2$ to maintain equality amongst them. To produce a comparative output, the source task data was compared to each sensor on each day. This produced 34 differing spatial and temporal contextual domains. In order to assess the performance of the fuzzy transferable framework, an average Normalised Root Mean Squared Error (NRMSE) of the predicted and actual temperature value was produced.

An inherent issue within the deployment of a Wireless Sensor Network (WSN) such as those used within Intelligent Environments comes with the collective aligning of individual sensor time stamps. Within the Intel data set, each sensor produced an individual timestamped dataset out of alignment with other network sensors. In order to compare the performance of the sensor readings over a temporally defined interval, a linear interpolation method was used. Using linear interpolation, each data set had a 1000 points defined at set temporal intervals spread across the task domain. Taking the overall points, a corresponding average NRMSE value was produced between the predictive values and the actual temperature value.

To analyse the difference in contexts between source and task dataset, an average NRMSE value of each data pair of the input variables was produced. Due to the linear interpolation of the time variable resulting in a zero difference, only the NRMSE value of the light value is displayed.

In order to assess the FuzzyTL framework a comparison is made with an FIS constructed using a standard ADDL process. The Wang-Mendel algorithm was employed to produce the FIS and to subsequently output values. The process to construct the fuzzy sets and fuzzy rules within this inference system can be found within [22].

4.3 Results

Within Fig. 5 a summary of the comparison between the ADDL and the FuzzyTL can be found. Overall the FuzzyTL out performs the ADDL system. In a high proportion of instances (24 out of 34) the margin is greater than 1.0. Due to

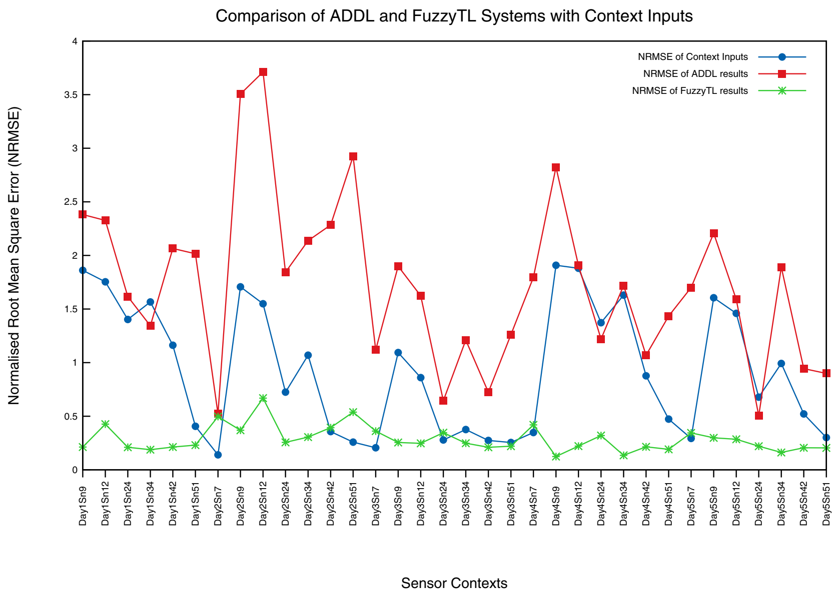


Fig. 5. A Comparison of the NRMSE Results of the ADDL and FuzzyTL System Against Each Contextual Situation

the variation in the system construction this is to be expected. It is evident that in selected instances the improvement of the FuzzyTL over the ADDL was only small. This can be seen if we compare the performance of the ADDL system with the FuzzyTL process on Day 2 with Sensor 7. The performance of the ADDL is only marginally lower than the FuzzyTL, 0.49642 to 0.5225. Due to the contextual difference of the input variable being significantly low in comparison to the other measured domains, it is believed that this may be a factor. Within Fig 5 we additionally show a comparison of the input contexts against the results produced by the ADDL. There appears from this data to be a correlation between these datasets. The FuzzyTL, by contrast, absorbs the difference in context and produces similar output values across all of the contexts despite the variations. This highlights the ability of the FuzzyTL framework to absorb the dynamic nature of the data produced by the application.

It can be seen from the graph that the Fuzzy Transfer Learning (FuzzyTL) framework was able to out perform the Ad-Hoc Data Driven Learning (ADDL) process across all five days and across the seven sensors that were tested. It is evident that the lack of adaptive ability within the ADDL can account for the failure to accommodate to the changing spatial and temporal contexts. By focussing on a single spatial-temporal instance we can see the ADDL systems inability to produce output for sections of the input values. Focussing on sensor 34 on day 1, Fig 6 demonstrates the failure at specific time segments. This is shown with the production of a 0 value. Between 07:31 and 16:25, the ADDL system

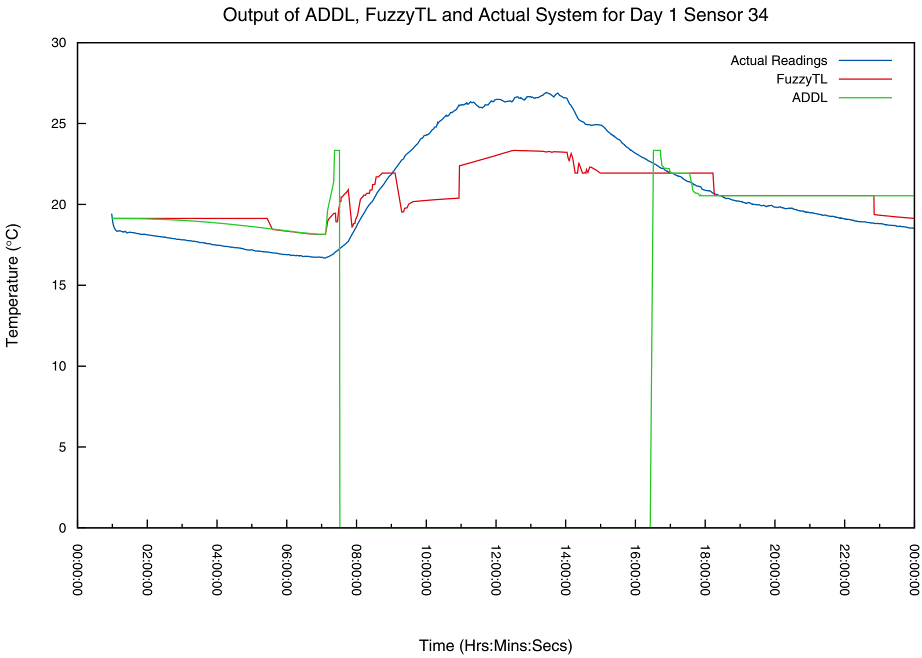


Fig. 6. A Comparison of the Ad-Hoc Dynamic Data Learning and the Fuzzy Transfer Learning systems Against the Actual Results Generated using a Normalised Root Mean Squared Error for Day 1 Sensor 34

fails to produce an output value. The flexible and online learning attributes of the FuzzyTL framework allows for the construction of an increased number of rules and adaptation of the set structure which assists in the production of values during this period. Across the 34 contexts, all rule bases had an increased quantity of rules when comparing the ADDL system to the FuzzyTL framework. Although not indicative of improved performance, it is evident that the process was able to absorb the changes in the data and adapt the system accordingly.

5 Conclusion

Within this paper we have discussed the use of a generic Fuzzy Transfer Learning system within the application of Intelligent Environments. We highlighted the benefits of using a Fuzzy Transfer Learning (FuzzyTL) process over conventional machine learning techniques, specifically focusing on its use within the prediction of sensor readings in Intelligent Environments (IE's). The proposed system was compared to a Ad Hoc Data Driven Learning system across data gathered from a real world sensor network. We were able to show that despite changing temporal and situational contexts, the FuzzyTL system was able to adapt to the changes.

The system was shown to surpass the ADDL system. Overall we highlighted the capacity of the system to utilise previously learned data to inform differing contextual situations.

We have isolated a number of areas within which we feel that the framework we have outlined can be extended and further improved. To understand the applications of FuzzyTL further we are investigating its implementation to other IE datasets. This will allow for a greater depth of understanding of the performance of the modelling structure. Currently we using the system to analyse an in-house sensor network dataset closely related to this work.

Focussing on the framework and the elements contained within, we are to investigate methods to improve the production of the fuzzy sets and fuzzy rules in respect to the source task so as to improve the overall performance of the system. The use of an optimisation technique is additionally being investigated to overcome issues encountered with the contextual differences across tasks. In combination with this approach, further work will be focussed on the online adaptation of the fuzzy transfer system with a view to greater exploration of the online pruning and tuning of the Fuzzy Inference System (FIS).

Acknowledgements. The authors would like to thank MIT Computer Science and Artificial Intelligence Laboratory for the provision of the data used in this work.

References

1. Arslan, A., Kaya, M.: Determination of fuzzy logic membership functions using genetic algorithms. Fuzzy Sets and Systems 118(2), 297–306 (2001)
2. Blanke, U., Schiele, B.: Remember and transfer what you have learned-recognizing composite activities based on activity spotting. In: 2010 International Symposium on Wearable Computers (ISWC), pp. 1–8. IEEE (2010)
3. Bokareva, T., Hu, W., Kanhere, S., Ristic, B., Gordon, N., Bessell, T., Rutten, M., Jha, S.: Wireless sensor networks for battlefield surveillance. In: Proceedings of Land Warfare Conference 2006, Citeseer (2006)
4. Casillas, J., Cordón, O., Herrera, F.: Improving the wang and mendels fuzzy rule learning method by inducing cooperation among rules. In: Proceedings of the 8th Information Processing and Management of Uncertainty in Knowledge-Based Systems Conference, pp. 1682–1688 (2000)
5. Chen, M.Y., Linkens, D.A.: Rule-base self-generation and simplification for data-driven fuzzy models. In: The 10th IEEE International Conference on Fuzzy Systems, vol. 1, pp. 424–427. IEEE (2001)
6. Chok, H., Gruenwald, L.: Spatio-temporal association rule mining framework for real-time sensor network applications. In: Proceeding of the 18th ACM Conference on Information and Knowledge Management, pp. 1761–1764. ACM (2009)
7. Deshpande, A., Guestrin, C., Madden, S.R., Hellerstein, J.M., Hong, W.: Model-driven data acquisition in sensor networks. In: Proceedings of the Thirtieth International Conference on Very Large Data Bases. VLDB Endowment, vol. 30, pp. 588–599 (2004)

8. Farhadi, A., Tabrizi, M.K.: Learning to Recognize Activities from the Wrong View Point. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 154–166. Springer, Heidelberg (2008)
9. Herrera, F.: Genetic fuzzy systems: taxonomy, current research trends and prospects. Evolutionary Intelligence 1(1), 27–46 (2008)
10. Jung, Y.J., Lee, Y.K., Lee, D.G., Ryu, K.H., Nittel, S.: Air Pollution Monitoring System based on Geosensor Network 3 (2008)
11. Kidd, C.D., Orr, R., Abowd, G.D., Atkeson, C.G., Essa, I.A., MacIntyre, B., Mynatt, E., Starner, T.E., Newstetter, W., et al.: The aware home: A living laboratory for ubiquitous computing research. LNCS, pp. 191–198 (1999)
12. Luhr, S., West, G., Venkatesh, S.: Recognition of emergent human behaviour in a smart home: A data mining approach. Pervasive and Mobile Computing 3(2), 95–116 (2007)
13. Madden, S.: Intel lab data (June 2004), http://db.csail.mit.edu/labdata/labdata.html (published on 2nd June 2004)
14. Pan, S.J., Kwok, J.T., Yang, Q.: Transfer learning via dimensionality reduction. In: Proceedings of the 23rd National Conference on Artificial Intelligence, pp. 677–682 (2008)
15. Popescu, M., Coupland, S.: A Fuzzy Logic System for Acoustic Fall Detection. In: Skubic, M. (ed.) Proc. 2008 AAAI Symposium on AI in Eldercare. AAAI, Washington DC (2008)
16. Raina, R., Battle, A., Lee, H., Packer, B., Ng, A.Y.: Self-taught learning: transfer learning from unlabeled data. In: Proceedings of the 24th International Conference on Machine Learning, pp. 759–766. ACM (2007)
17. Rashidi, P., Cook, D.J.: Multi home transfer learning for resident activity discovery and recognition. In: KDD Knowledge Discovery from Sensor Data, pp. 56–63 (2010)
18. Ross, T.J.: Fuzzy logic with engineering applications. Wiley Online Library (1997)
19. Simon, G., Maróti, M., Lédeczi, Á., Balogh, G., Kusy, B., Nádas, A., Pap, G., Sallai, J., Frampton, K.: Sensor network-based countersniper system. In: Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems, pp. 1–12. ACM, New York (2004)
20. Sixsmith, A., Johnson, N.: A smart sensor to detect the falls of the elderly. IEEE Pervasive Computing 3(2), 42–47 (2004)
21. Wang, L.-X.: The wm method completed: a flexible fuzzy system approach to data mining. IEEE Transactions on Fuzzy Systems 11(6), 768–782 (2003)
22. Wang, L.X., Mendel, J.M.: Generating fuzzy rules by learning from examples. IEEE Transactions on Systems, Man and Cybernetics 22(6), 1414–1427 (1992)
23. Wang, Q., Shin, W., Liu, X., Zeng, Z., Oh, C., AlShebli, B.K., Caccamo, M., Gunter, C.A., Gunter, E., Hou, J., et al.: I-Living: An open system architecture for assisted living, Citeseer (2006)
24. Wood, A., Stankovic, J., Virone, G., Selavo, L., He, Z., Cao, Q., Doan, T., Wu, Y., Fang, L., Stoleru, R.: Context-aware wireless sensor networks for assisted living and residential monitoring. IEEE Network 22(4), 26–33 (2008)
25. Zadeh, L.A.: Fuzzy logic. Computer 21(4), 83–93 (1988)

Gesture Profile for Web Services: An Event-Driven Architecture to Support Gestural Interfaces for Smart Environments

Radu-Daniel Vatavu[1], Cătălin-Marian Chera[2], and Wei-Tek Tsai[3]

[1] University Stefan cel Mare of Suceava, 720229 Suceava, Romania
[2] Computer Science Department, Politehnica University of Bucharest, Romania
[3] School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ 85287, USA & Department of Computer Science and Technology, Tsinghua University, Beijing, China

vatavu@eed.usv.ro, catalin.chera@cs.pub.ro, wtsai@asu.edu
http://www.eed.usv.ro/~vatavu

Abstract. Gestural interfaces have lately become extremely popular due to the introduction on the market of low-cost acquisition devices such as iPhone, Wii, and Kinect. Such devices allow practitioners to design, experiment, and evaluate novel interfaces and interactions for new smart environments. However, gesture recognition algorithms are currently the appanage of machine learning experts which sometimes leaves AmI practitioners dealing with complex pattern recognition techniques instead of focusing on prototyping ambient interactions. To address this problem, we propose GPWS (Gesture Profile for Web Services), a service-oriented architecture (SOA) designed to assist implementation of gestural interfaces. By providing gesture recognition as a web service, we leverage easy and fast adoption of gestural interfaces for various platforms and environments through simple service discovery and composition mechanisms. We discuss two GPWS designs based on SOA 1.0 and SOA 2.0 standards, analyze their performance, and demonstrate GPWS for a gesture-controlled smart home application.

Keywords: Gesture, gesture-based control, service-oriented computing, event-driven architecture, smart home, web services, gesture recognition, SOA, EDA.

1 Introduction

Moving into a world of intelligent ambient systems puts high demands on the interactive experience delivered to users [1,2,32]. However, traditional ways to interact with information systems have been restricted to keys and buttons, computer mice, and trackpads, while the remote control still represents the industry standard for interacting with household electrical appliances despite the potential of next-generation homes [5]. Such standard interfaces prove sometimes difficult to control (e.g., TV remotes are being perceived as confusing by most users [4]) or inappropriate for some tasks and contexts (e.g., the control of a remote screen in the kitchen [27]). As a suitable alternative, gestural interfaces are meant to deliver natural and intuitive interactions [26,37]. This angle has already been successfully exploited by the gaming industry which showed an increased interest for incorporating people' willingness to move and gesture into playing

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 161–176, 2012.

video games. The outcome has been a large palette of gesture acquisition devices with Nintendo Wii Remote[1] and Microsoft Kinect[2] being just a few popular examples which are being reused today for ambient intelligence applications [11,15,37,38].

As practitioners of ambient intelligence become more and more interested in implementing gestural interfaces into their own designs, high access to gesture recognizers is needed. One way to achieve this is to provide designers with detailed pseudocode for such recognizers [3,41] so that they can be adopted and implemented on various platforms. Another practice would be to reuse existing code, not necessarily in the form of libraries but rather as service provided over the web. The practice of service-oriented software engineering [10,14] has already shown the benefits of such software architectures in terms of platform independence, loosely coupled design, and alignment with life cycle support processes. In this context, we believe that a common framework hiding the complex details of gesture recognition algorithms and machine learning formalisms while exposing clear services would be extremely beneficial to practitioners. This way, gesture-based control could be more easily adopted by the service computing community just as any other service available on the web. Therefore, we discuss in this work web services that deliver gesture recognition by following the existing AmI practices of developing service-oriented software infrastructure [7,24,34,36].

This paper introduces GPWS (Gesture Profile for Web Services), a novel approach of presenting gesture recognition for control applications in a service-oriented event-driven manner. We highlight our main contributions:

1. We provide AmI practitioners with web services for gesture recognition in order to facilitate easy adoption and promote gesture-based interactions for smart environments.
2. GPWS is the first event-driven architecture for gesture recognition. This design choice is motivated by the fact that human gestures are naturally event-driven (they have clear start, execution, and ending timestamps) and that acquisition devices also deliver data in discrete events.
3. We motivate the need for an event-driven implementation (SOA 2.0) for gesture processing by discussing a comparison with a simpler GPWS design which uses simple request-response web services as per the standards of SOA 1.0. We show how each architecture brings benefits to practitioners in accordance to their application needs and requirements.

By introducing GPWS we hope to address the practical needs of researchers and practitioners interested in using gestures for their applications. We plan GPWS as an open-source framework with free services available to the community and refer the interested reader to `http://gpws.fcint.ro`. We demonstrate our architecture with a sample application designed to control household appliances.

[1] `http://nintendo.com/wii`
[2] `http://www.xbox.com/en-GB/kinect`

2 Related Work

A large amount of work exists in the pattern recognition and human-computer interaction communities with regards to capturing motion and gestures, implementing gesture recognizers, and designing gestural interfaces [25,28,29].

Common gesture capture technologies include video cameras [28], accelerometers embedded in mobile devices [31], game controllers [33], wrist watches [21], and worn equipment such as sensor gloves [13]. Also, the recent years have seen several low-cost acquisition devices being released by the gaming industry, such as Nintendo Wii Remote and Microsoft Kinect. Many researchers have reused the motion sensing capabilities of such devices for interactive applications and for exploring gesture-based interfaces. For example, Lee [17] has inventively used the Wii Remote for 3D head tracking and touch-based interactions. Bott et al. [6] used the remote for interactive art and music applications. Vatavu [38] described an ambient interactive system using Kinect in order to demonstrate the concept of nomadic gestures, while Panger [27] used Kinect for augmenting interactions in the kitchen. Also, researchers showed interest in the technical performance of such devices such as the pointing accuracy of the Wii Remote [22] or its performance for recognizing gestures [19,33]. Wii has also found applications in controlling home appliances. For example, Pan et al. [26] used the Nunchuck controller attached to the Wii Remote to implement GeeAir, a device for controlling household appliances using gestures. Vatavu [37] explored the Wii Remote functionalities for providing WIMP-like interactions (Windows, Icons, Menus, Pointer) for working with multiple TV screens for home entertainment systems.

Whilst paying careful attention to these works in order to select the most appropriate recognition technique for our implementation, we are more interested in software architectures designed for gesture-based control as well as in gesture taxonomies and ontologies informing the design of such architectures [8]. Following this perspective, previous research of direct interest to our work can be grouped into gesture ontologies and software architectures.

Ontologies represent a powerful tool for understanding concepts and relationships with important applications in informing software design and development [18,43]. Wang et al. [39] introduced and described an ontology for multi-touch gestures. The authors discussed atomic gestures and techniques for combining them using temporal, spatial, and logical relationships. Sonntag et al. [35] defined ontological representations for pointing gestures. van Seghbroeck et al. [34] described WS-Gesture which represents a framework for controlling devices based on DPWS (Devices Profile Web Services). DPWS relies on WS-* protocols initially designed for enterprise computing and then ported to the device networks that may have hardware constraints that an enterprise system does not have. Chera et al. [8] developed a gesture ontology for informing the design of service-oriented architectures for gesture controlled applications. The ontology groups gesture concepts and relationships into execution, implementation, and reflection levels. Although not explicitly developing ontologies, other researchers have classified interactive gestures by identifying common properties. For example, Wobbrock et al. [42] introduced a taxonomy of surface gestures by taking into consideration criteria such as form, nature, binding, and flow. Ruiz et al. [31] proposed a taxonomy for motion gestures in the context of mobile phones by considering gesture mappings (nature, context, and temporal) and physical characteristics (kinematic impulse, dimension, and complexity) as classification criteria.

The practitioners of ambient intelligence have already considered service-oriented architectures for the software infrastructure needs of smart environments [36]. For example, Mingkhwan et al. [24] proposed an architecture for interconnecting home appliances and their services; Chakraborty et al. [7] discussed service composition for mobile and pervasive computing environments; and Seghbroeck et al. [34] described the first protocol of using gestures in conjunction with web services. This work adds gesture recognition to the software infrastructure layers of ambient intelligence [36] by extending previous works [34] with the first event-driven service-oriented design for gesture recognition.

3 Gesture Representation and Recognition

Throughout this work we understand by gesture a continuous motion captured using some specific acquisition device. Irrespective of the acquisition technology, a gesture can be finally represented as a time-ordered series of points: $\{(x_i, y_i, z_i) | i = 1..n\}$, where n represents the sampling resolution. Figure 1 illustrates a "beat" gesture consisting in two rapid hand movements captured by a 3-axis accelerometer.

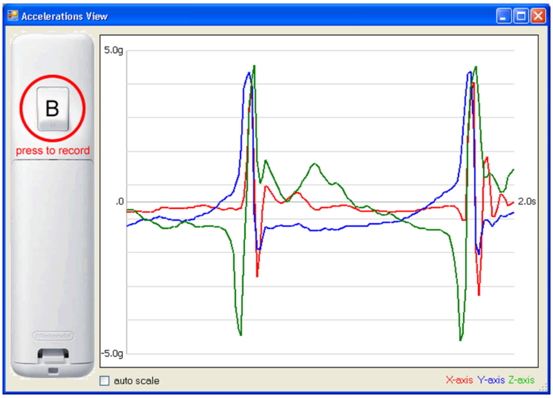


Fig. 1. Accelerated gesture composed of two beat-like movements captured by a 3-axis accelerometer (beat strokes can be easily identified by their high force peaks). Screen capture from our demo application (see the last section of the paper).

The pattern recognition and human-computer interaction communities have proposed many algorithms to date for recognizing gestures [3,9,19,28,29,40,41]. Out of these, this project employs the Dynamic Time Warping distance (DTW) [19] in the context of the Nearest-Neighbor supervised learning approach (NN). DTW computes the optimum alignment between two motion gestures by minimizing the total sum of point-to-point Euclidean distances. The NN approach classifies a candidate gesture by comparing it with every sample stored in the training set and returns the class of the closest sample. We selected the combination of DTW/NN to perform gesture recognition as it was found to work well in the presence of user variation in gesture execution even for long term testing [19]. As the gesture recognition technique is not the main focus of this paper and we are more interested in software architectures that deliver efficient web services to the community, we refer the interested reader to Liu et al. [19] for more details on the recognition process.

The gesture terminology introduced so far only deals with single stroke gestures. However, more complex gesture commands can be imagined by composing individual gesture executions. For example, drawing a "circle" followed by drawing digit "1" could stand as a command for increasing temperature with 1 degree for an air conditioner in a smart home scenario. We therefore define a gesture sequence command as a set of one or more gestures that have been associated with meaning: $gesture\ sequence = \{gesture_1, gesture_2, ...\}$.

4 GPWS: Gesture Profile for Web Services

The motivation behind our work is to make gesture recognition available as a service for researchers and practitioners interested in prototyping gesture-based control applications. Also, we are interested in providing such services with different levels of complexity to address applications with various requirements and needs. This section presents two such service-oriented architectures. The first one implements the minimum amount of services and functionality needed to fulfill the basic needs of a practitioner interested in gesture-based processing: gesture recognition and management of the gesture set. The solution is therefore characterized by low complexity and simple usage patterns in the standard of SOA 1.0 web services[3]. The second SOA 2.0 event-driven architecture[4] is more ellaborate, flexible, and therefore able to expose an increased level of functionality for subscribing clients. At the end of the paper, we discuss architecture performance and provide a comparison between the two designs.

4.1 GPWS 1.0: A Simple Architectural Design Using SOA 1.0 Response-Request Standards

This section presents our first implementation of a service-oriented GPWS using SOA 1.0 web services within the standard request-response protocol. By analyzing the practical requirements of a gesture-controlled application, we identified and implemented the following services:

1. **Recognition services**. Two services were implemented to recognize motion gestures (GESTURE-RECOGNITION-SERVICE) and sequences of gestures executed in order (GESTURE-SEQUENCE-RECOGNITION-SERVICE). The gesture service

[3] Oasis Reference Model for Service Oriented Architecture 1.0, https://www.oasis-open.org/committees/download.php/19679/soa-rm-cs.pdf

[4] Patricia Seybold Group, Event-Driven Architecture Overview, http://www.omg.org/soa/Uploaded%20Docs/EDA/bda2-2-06cc.pdf

implements the Nearest-Neighbor classification approach employing the DTW distance for motion trajectories, as described previously in the paper. The gesture sequence service implements a simple matching rule which compares the candidate sequence against a set of previously stored prototypes. For example, if gestures "circle" and "up" are performed consecutively by the same user in less than X seconds[5], this sequence of gestures is interpreted as a single command.

2. **Manager services for user-defined data**. Gestures and commands (i.e., functions controlled in the application) represent user-defined data. Two services were implemented for storing and retrieving such data from a database repository: GESTURE-MANAGER-SERVICE and GESTURE-SEQUENCE-MANAGER-SERVICE. They serve as a simple interface between the client application and the database allowing users to manage gesture-function mappings. The first service associates gesture motions to ID values representing class names such as "circle", "turn-on", "help", etc. The gesture sequence manager associates a list of gesture IDs to a command ID. For example, the command "increase the temperature of the air conditioner by 1 degree" can be stored as the set {"up", "digit-1"} with each member in the set representing the class ID of an individual gesture.

Figure 2 illustrates the basic components of this first GPWS architecture: client application, gesture acquisition device, web services, and database. During setup, clients use the manager services to upload gesture-function mappings as well as gesture samples for the training set. During runtime, the client application performs requests to the gesture recognition services that respond with class IDs. Gesture recognition details are therefore hidden away into the GPWS architecture which only exposes simple functions to client subscribers.

4.2 GPWS 2.0: An Elaborated Event-Driven Architectural Design Using SOA 2.0 Standards

Although the first design satisfies the basic needs of a client application implementing gesture recognition, it is limited in terms of flexibility and scaling (e.g., only the requesting client is notified of the classification result). However, complex applications with more demanding requirements such as multiple clients that need to be informed when a specific action has occurred, need more elaborated architectures. As SOA 1.0 services use the principle of synchronous request-response calls, they can't address such needs. SOA 2.0 goes further by allowing asynchronous calls. This way, event-driven architectures allow various components to monitor the environment, process events, and respond to changing conditions continuously. Multiple clients can be informed once a specific event occurred (e.g., the user performed the "turn-off air conditioner" gesture) in order to take specific decisions (e.g., the air conditioner client performs the command; the power consumption monitoring client updates its status and predictions; while a smart client monitoring user actions can verify whether this is typical behavior of the specific user). Also, in complex EDA [20], the traditional subscription ways in

[5] The number of seconds represents an application setting which depends on the length of the sequence, e.g. 10 seconds for small sequences of 2-3 gestures. As a thumb rule, we add 5 seconds for each additional gesture in the sequence.

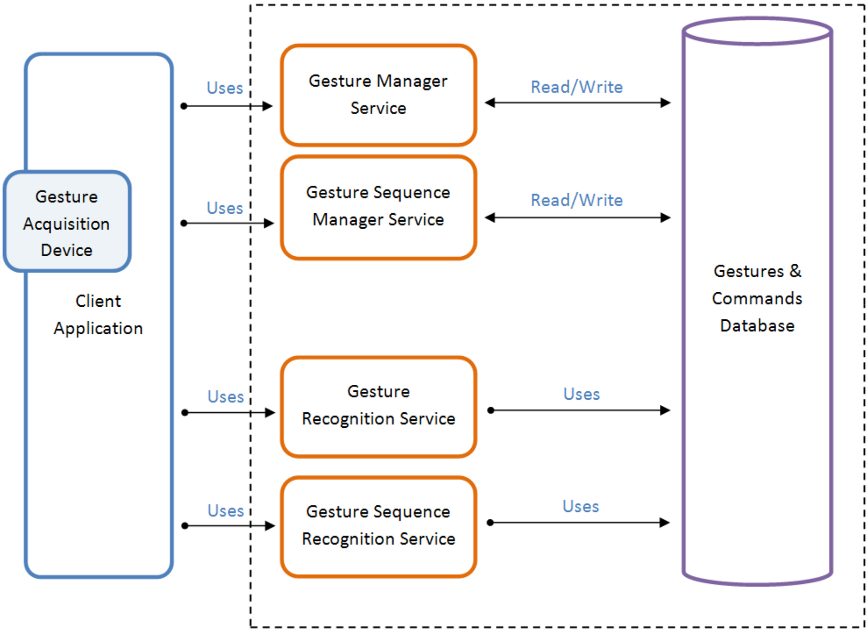


Fig. 2. 1st generation GPWS consisting in gesture managers and gesture recognition web services in the style of SOA 1.0 standards

which a filter is usually applied to a single event are replaced by more sophisticated mechanisms for which the decision is made using correlations across histories of multiple event streams.

We also note that gestures are inherently event-driven: they consist of preparation, stroke, and retraction phases [23]. Also, acquisition devices signal new incoming data at specific discrete intervals while recognizers process the acquired gestures and report recognition scores in a timely fashion. All these represent clear and distinct events from the application but also the users' point of view. Therefore, in order to model such mechanisms, we designed an event processing architecture which we refer to as the 2nd generation GPWS. The architecture has five main components (see Figure 3):

1. **Events** are generated and processed in correspondence to gesture acquisition, gesture recognition, and gesture sequence matching key points (specific events employed by the architecture are described in detail next in the paper).
2. **Event processing services** implement event generation, processing, logging, and publishing. For example, the EVENTS-DISPATCHER-SERVICE serves as the architecture key point for generating events while the EVENTS-PROCESSING-SERVICE implements all the logic required for processing, logging, and publishing events. The specific services of GPWS 2.0 are described next in the paper.
3. **Acquisition service**: a special-purpose GESTURE-ACQUISITION-SERVICE was implemented to serve as middleware between the client application and the GPWS 2.0

architecture. This assures the independence between the acquisition device and the functionality of GPWS. The acquisition service represents the entry point of the architecture for connecting clients.

4. **Recognition services**: the same recognition services from the first generation architecture were also kept in this one. However, this time they act as event handlers and are being called by the EVENTS-PROCESSING-SERVICE when needed.
5. **Manager services for user-defined data**: they are the same as in the first generation architecture (GESTURE-MANAGER-SERVICE and GESTURE-SEQUENCE-MANAGER-SERVICE) and are used to store and retrieve user-defined gestures and commands to/from the database.

GPWS 2.0 processes three types of events relating to gesture acquisition, recognition, and sequence matching. GESTURE-ACQUIRED-EVENT is generated once the gesture was successfully acquired from the user. The client application connects to the entry

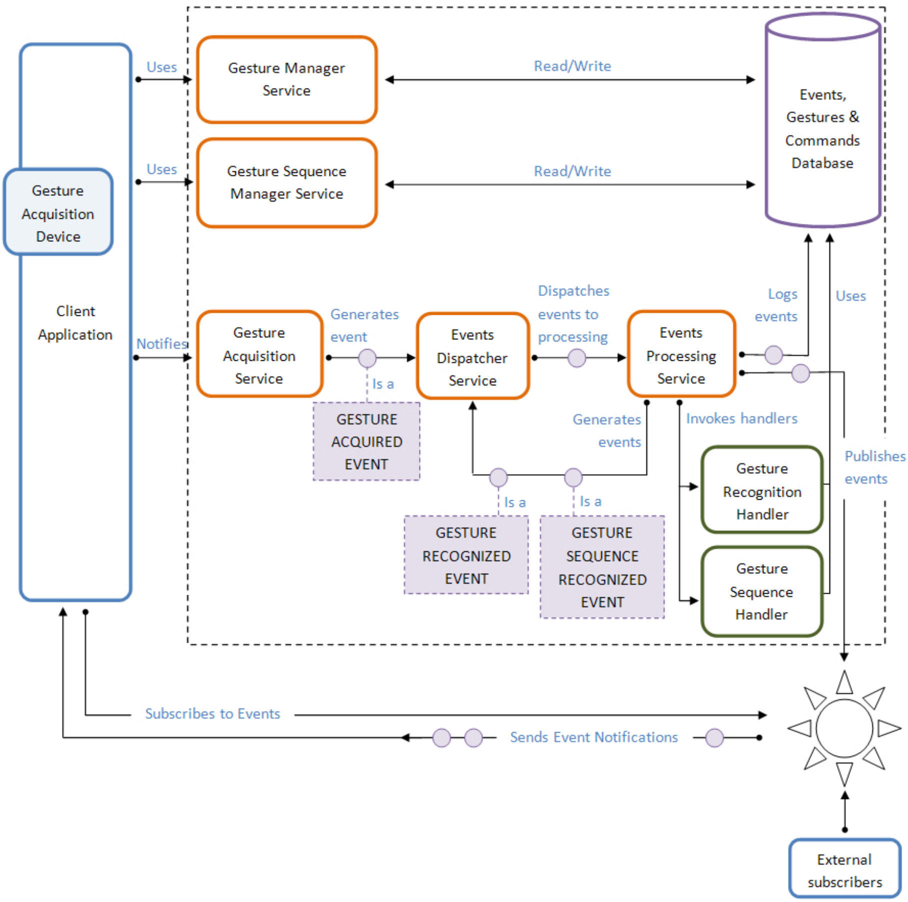


Fig. 3. Event-driven architectural design for the SOA 2.0 generation of GPWS

point of the GPWS architecture and performs a call to the GESTURE-ACQUISITION-SERVICE which will make a request to the EVENTS-DISPATCHER-SERVICE to generate a new acquisition event. The event has two arguments: client ID and the set of points acquired from the device. The GESTURE-RECOGNIZED-EVENT is generated once GESTURE-RECOGNITION-HANDLER has run the recognition algorithm. The event arguments (client ID and the class of the recognized gesture) are fed back to the EVENTS-PROCESSING-SERVICE. The gesture sequence event is produced once the GESTURE-SEQUENCE-RECOGNITION-HANDLER has run the sequence matching algorithm. This event also has two arguments: client ID and the class of the recognized command.

Most of the tasks in GPWS 2.0 are performed using the event processing service as per the following scenario. The client application acquires gesture data from a specific device. Once a gesture has been acquired, the client notifies the acquisition service that a new gesture is ready to be processed. The acquisition service represents the single entry point for the client when working with the GPWS 2.0 engine while it also separates the specific details of the acquisition device from gesture representation required by GPWS. The acquisition service uses the EVENTS-DISPATCHER-SERVICE to create a new GESTURE-ACQUIRED-EVENT having as arguments the client ID and the acquired motion. The event is passed to EVENTS-PROCESSING-SERVICE. The only function of the dispatcher service is to act as a central unified point in the architecture for creating events and for transmitting them towards processing. The event processing service implements the logic for handling incoming events:

- The event is logged in the database together with its arguments and timestamp.
- The appropriate handler is called: GESTURE-RECOGNITION-HANDLER will handle GESTURE-ACQUIRED-EVENT(s) and GESTURE-SEQUENCE-RECOGNITION-HANDLER will process GESTURE-RECOGNIZED-EVENT(s). These handlers implement gesture and sequence matching algorithms by using gesture samples and command definitions stored in the database via manager services.
- After calling and executing handlers, new events are generated as follows: GESTURE-RECOGNIZED-EVENT is generated after a call to the GESTURE-RECOGNITION-HANDLER and GESTURE-SEQUENCE-RECOGNIZED-EVENT may be generated after a call to the GESTURE-SEQUENCE-RECOGNITION-HANDLER. The processing service uses the dispatcher to generate such events (which by means of the dispatcher will finally reach the processing service later on).
- If the event should be published, the processing service will publish it to the relevant subscribers.

5 GPWS Performance Analysis and Demonstration

We report in this section GPWS performance in terms of recognition rate and response times and describe a demo application employing GPWS in the context of smart home control. Table 1 lists a summary comparison between the two GPWS architectures on implementation and performance criteria.

5.1 Recognition Rate

A simulation environment was designed for automatic testing in order to assess the recognition performance of the GPWS architectures. A custom designed application simulated users performing motion accelerated gestures. The simulation used data from a publicly available gesture set introduced by Hoffman et al. [12]. The set contains 25 types of motion gestures performed repeatedly by 17 participants for 20 times with a total number of 8,500 samples[6]. More details on the acquisition procedure, apparatus, and participants can be found in [12]. The testing application loaded gesture samples from the set which were sent for processing to GPWS, simulating thus user behavior.

As Nearest-Neighbor recognizers perform better as more gesture samples are available in the training set, we compute and report recognition performance for varying numbers of samples per gesture type. Figure 4 (left) shows the results obtained. Recognition accuracy reached 92.5% with just 4 samples per gesture type, rose up to 95.0% with 8 training samples, and reached 96.6% with 16 examples. A fixed sampling rate of n=32 points was used during testing. A Friedman test showed a significant effect of the number of training samples on recognition accuracy ($\chi^2(4) = 5634.966$, $p < .001$).

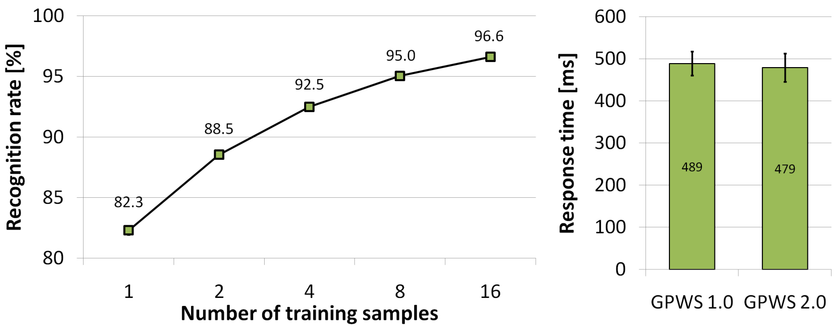


Fig. 4. GPWS architecture performance. Left: accuracy of the gesture recognition service (DTW and Nearest-Neighbor) vs. the number of training samples per gesture type. Right: response times for the two architecture designs (error bars show ± 1 SD).

5.2 Response Time

We also measured response times for the two different architectures. The simulation application measured and averaged response times for 500 consecutive gesture recognition requests. We hypothesized that the first generation GPWS would be faster due to direct request-response mechanisms. However, response times were close for both architectures: 488.6 ms (SD=28.5 ms) for the first generation GPWS and 479.2 ms (SD=33.6 ms) for the second event-driven implementation (Figure 4, right). A Wilcoxon signed-rank test showed that GPWS 2.0 was significantly faster than GPWS 1.0 ($Z = -6.107$,

[6] http://www.eecs.ucf.edu/isuelab/downloads.php

Table 1. Summary comparison between the two GPWS architectural designs

Criterion	GPWS 1.0	GPWS 2.0
FEATURES		
Web services standards	SOA 1.0	SOA 2.0
Communication	synchronous (request-response mechanisms)	asynchronous (event processing)
Supports 3rd party clients	no	yes
PERFORMANCE		
Gesture recognition accuracy[1]	95% with 8 samples per gesture type and 97% with 16 samples per gesture type	
Response time[2]	489 ms	479 ms
USAGE RECOMMENDATIONS		
Should be used when	simple gesture recognition is needed by a single client application	gesture events are important for 3rd party clients

[1] Measured on a set of 8,500 3D gesture samples [12], see the appropriate section for details.
[2] Measured on a 2.40 GHz Intel CoreDuo Quad CPU computer, see the appropriate section for details.

$p < .001$) with a small Cohen effect ($r < .2$). Execution times are acceptable for real-time response requirements. Times were measured on a 2.40 GHz Intel CoreDuo Quad CPU computer running Microsoft Windows with IIS hosting GPWS and MySQL as supporting data server.

5.3 Home Control Application

In order to demonstrate GPWS, we implemented a home control application. We chose such an application due to many works being interested in controlling household devices via gestures [16,30,37]. The testing environment was a laboratory room which contained four controllable devices: two light lamps and two air conditioners. While the lamps can only be switched on and off, the air conditioners exposed several functions: on/off; increase/decrease temperature by X degrees; select operation mode; and schedule settings. Users were able to define, edit, and publish gesture commands to the server and to associate commands to devices and functions. Figure 5 illustrates the operation of the application. Gestures were acquired with the Wii Remote which was selected for several reasons:

- The TV remote-like form factor makes the device look familiar to most users.
- Multiple options for capturing user input (a 3-axis accelerometer; an optical sensor with embedded software working in the infrared domain; 12 buttons such as +/-, home, 1/2, up/down, and left/right).

- Multiple options for providing user feedback in the form of audio through the embedded speaker, haptic through a vibromotor, and visual by using the 4 bright LEDs located at the bottom on the front side of the controller.
- Easy connectivity with PC systems via Bluetooth.
- Numerous users are buying the Wii console (over 95 million units sold before March 2012)[7]. Therefore, users are already familiar with how the Wii controller works and this acquired experience can be reused in our application.

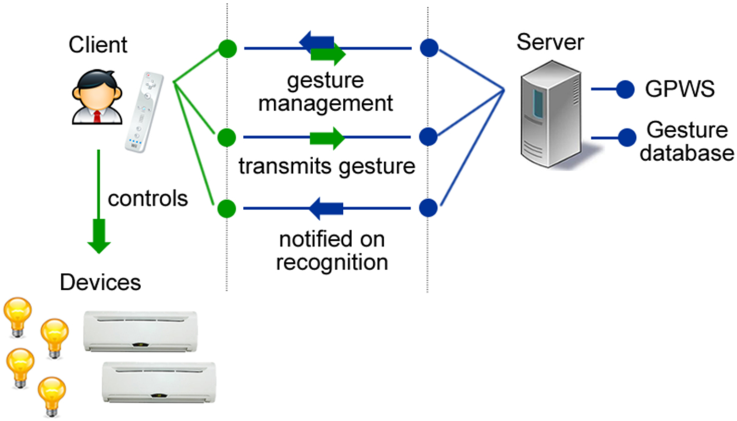


Fig. 5. Operation diagram for the smart home control application

As all devices exposed similar functions (on/off), we adopted the design choice of assigning a unique gesture for common functions and prefix it by an ID gesture identifying the device. For example, the air conditioner 1 (AC1) was identified by performing a quick beat-like movement while air conditioner 2 (AC2) with two quick movements (the two force peaks gesture illustrated in Figure 1). Lamp one (L1) was identified by a small circle (suggesting the shape of the bulb) and lamp two (L2) by two consecutive circles. Once the device identification gesture was performed, all following gestures were directed to that specific device. For example, an "S"-like shape was used to start and an "X" to stop the device. Valid commands in our application were {"Identify-AC1", "S"}, {"Identify-Lamp-2", "X"}, {"Identify-AC2", "arrow-down", "digit-1"}, etc. We limit our description here and note that the main goal of the smart home prototype was only to demonstrate the applicability of the GPWS architecture for gesture recognition. Therefore, we don't discuss here aspects related to gesture set design (i.e., gesture fit to function, learnability, and execution difficulty) for which special design techniques have been proposed in the literature [31,42]. We simply note the availability of the GPWS web services at `http://gpws.fcint.ro`. Also, in order to illustrate GPWS 2.0 ease of use, we present below a short section of C# code that employs the architecture to recognize a gesture candidate and receives the response as an event

[7] Nintendo, Consolidated Sales Transition by Region, `http://www.nintendo.co.jp/ir/library/historical_data/pdf/consolidated_sales_e1203.pdf`

(we assume that gesture samples have been submitted before by the client identified by `subscriberId`):

```
public class GPWS20Example : SubscriptionService.IGestureRecognizedCallback
{
    SubscriptionService.GestureRecognizedSubscriber subscriber = null;
    InstanceContext contextOnGestureRecognized = null;
    AcquisitionService.GestureAcquisitionClient clientAcquisition = null;
    string subscriberId = ‘‘1000’’;

    // Subscribe to GPWS and create an acquisition client
    public void Initialize()
    {
        contextOnGestureRecognized = new InstanceContext(null, this);
        subscriber = new SubscriptionService.GestureRecognizedSubscriber(
            contextOnGestureRecognized
        );
        subscriber.Subscribe();
        clientAcquisition = new AcquisitionService.GestureAcquisitionClient();
    }

    // Data comes from the acquisition device,
    // usually as one data point at a time but we simplify here for the clarity of this example
    public void OnGestureData(Gesture gesture)
    {
        // Send gesture to GPWS for processing
        clientAcquisition.GestureWasAcquired(subscriberId, gesture);
    }

    // Gesture was recognized and event received from GPWS
    public void OnGestureRecognized(string subscriberId, string gestureName)
    {
        // Gesture was recognized as gestureName. Perform action.
    }

    // Unsubscribe from GPWS and close clients
    public void Uninitialize()
    {
        subscriber.Unsubscribe();
        subscriber.Close();
        clientAcquisition.Close();
    }
}
```

6 Conclusion

We introduced in this paper Gesture Profile for Web Services, an event-driven architecture delivering gesture recognition services. We experimented an SOA 2.0 event-driven design after we first implemented an SOA 1.0 request-response architecture. The EDA design was motivated by the event-driven nature of human gesture production and gesture acquisition devices. Performance measurements showed real-time responsiveness and high recognition accuracy of GPWS. Future work will focus on adding further processing layers on top of GPWS. One example is complex events processing [20] for inferring advanced knowledge from simple events which will allow other service-oriented architectures [36] to automatically be notified when complex event patterns occur.

Acknowledgments. This work is supported partially by European Regional Development Fund and the Government of Romania under the grant no. 181 of 18.06.2010 and the European project Empowering Romanian Research on Intelligent Information Technologies (ERRIC), FP7-REGPOT-2010-1/264207.

References

1. Aarts, E., de Ruyter, B.: New research perspectives on Ambient Intelligence. J. Ambient Intell. Smart Environ. 1(1), 5–14 (2009)
2. Aarts, E., Grotenhuis, F.: Ambient Intelligence 2.0: Towards Synergetic Prosperity. J. Ambient Intell. Smart Environ. 3(1), 3–11 (2011)
3. Anthony, L., Wobbrock, J.O.: A lightweight multistroke recognizer for user interface prototypes. In: Proceedings of Graphics Interface 2010 (GI 2010), pp. 245–252. Canadian Information Processing Society, Toronto (2010)
4. Bernhaupt, R., Obrist, M., Weiss, A., Beck, E., Tscheligi, M.: Trends in the living room and beyond: results from ethnographic studies using creative and playful probing. Comput. Entertain. 6(1), Article 5 (2008)
5. Bonino, D., Corno, F.: What would you ask to your home if it were intelligent? Exploring user expectations about next-generation homes. J. Ambient Intell. Smart Environ. 3(2), 111–126 (2011)
6. Bott, J.N., Crowley, J.G., LaViola Jr., J.J.: Exploring 3D gestural interfaces for music creation in video games. In: Proceedings of the 4th International Conference on Foundations of Digital Games (FDG 2009), pp. 18–25. ACM, New York (2009)
7. Chakraborty, D., Joshi, A., Finin, T., Yesha, Y.: Service composition for mobile environments. Mob. Netw. Appl. 10(4), 435–451 (2005), doi:10.1145/1160162.1160168
8. Chera, C.M., Tsai, W.T., Vatavu, R.D.: Gesture Ontology for Informing Service-oriented Architectures. In: IEEE Multi-Conference on Systems and Control (2012)
9. Ghosh, D., Dube, T., Shivaprasad, A.: Script Recognition: A Review. IEEE Trans. Pattern Anal. Mach. Intell. 32(12), 2142–2161 (2010)
10. Gold, N., Knight, C., Mohan, A., Munro, M.: Understanding Service-Oriented Software. IEEE Softw. 21(2), 71–77 (2004), doi:10.1109/MS.2004.1270766
11. Harley, D., Fitzpatrick, G., Axelrod, L., White, G., McAllister, G.: Making the Wii at Home: Game Play by Older People in Sheltered Housing. In: Leitner, G., Hitz, M., Holzinger, A. (eds.) USAB 2010. LNCS, vol. 6389, pp. 156–176. Springer, Heidelberg (2010)
12. Hoffman, M., Varcholik, P., LaViola Jr., J.J.: Breaking the status quo: Improving 3D gesture recognition with spatially convenient input devices. In: Proceedings of the 2010 IEEE Virtual Reality Conference (VR 2010), pp. 59–66. IEEE Computer Society, Washington, DC (2010)
13. Ibarguren, A., Maurtua, I., Sierra, B.: Layered architecture for real time sign recognition: Hand gesture and movement. Eng. Appl. Artif. Intell. 23(7), 1216–1228 (2010)
14. Ibrhaim, M.H., Holley, K., Josuttis, N.M., Michelson, B., Thomas, D., de Vadoss, J.: The future of SOA: what worked, what didn't, and where is it going from here? In: Companion to the 22nd ACM SIGPLAN Conference on Object-Oriented Programming Systems and Applications Companion (OOPSLA 2007), pp. 1034–1038. ACM, New York (2007)

15. Jung, Y., Li, K.J., Janissa, N.S., Gladys, W.L.C., Lee, K.M.: Games for a better life: effects of playing Wii games on the well-being of seniors in a long-term care facility. In: Proceedings of the Sixth Australasian Conference on Interactive Entertainment (IE 2009). ACM, New York (2009)
16. Kühnel, C., Westermann, T., Hemmert, F., Kratz, S., Müller, A., Möller, S.: I'm home: Defining and evaluating a gesture set for smart-home control. International Journal of Human-Computer Studies 69(11), 693–704 (2011)
17. Lee, J.C.: Hacking the Nintendo Wii Remote. IEEE Pervasive Computing 7(3), 39–45 (2008)
18. Li, W., Lee, Y.-H., Tsai, W.-T., Xu, J., Son, Y.-S., Park, J.-H., Moon, K.-D.: Service-oriented smart home applications: composition, code generation, deployment, and execution. Serv. Oriented Comput. Appl. 6(1), 65–79 (2012), doi:10.1007/s11761-011-0086-7
19. Liu, J., Zhong, L., Wickramasuriya, J., Vasudevan, V.: uWave: Accelerometer-based personalized gesture recognition and its applications. Pervasive Mob. Comput. 5(6), 657–675 (2009)
20. Luckham, D.: The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems. Addison-Wesley (2002)
21. Maurer, U., Rowe, A., Smailagic, A., Siewiorek, D.P.: eWatch: A Wearable Sensor and Notification Platform. In: Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2006), pp. 142–145. IEEE Computer Society, Washington, DC (2006)
22. McArthur, V., Castellucci, S.J., MacKenzie, I.S.: An empirical comparison of "wiimote" gun attachments for pointing tasks. In: Proceedings of the 1st ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS 2009), pp. 203–208. ACM, New York (2009)
23. McNeill, D.: Hand and mind: What gestures reveal about thought. University of Chicago Press (1992)
24. Mingkhwan, A., Fergus, P., Abuelma'Atti, O., Merabti, M., Askwith, B., Hanneghan, M.B.: Dynamic service composition in home appliance networks. Multimedia Tools Appl. 29(3), 257–284 (2006), doi:10.1007/s11042-006-0018-2
25. Mitra, S., Acharya, T.: Gesture Recognition: A Survey. Trans. Sys. Man Cyber Part C 37(3), 311–324 (2007), doi:10.1109/TSMCC.2007.893280
26. Pan, G., Wu, J., Zhang, D., Wu, Z., Yang, Y., Li, S.: GeeAir: a universal multimodal remote control device for home appliances. Personal Ubiquitous Comput. 14(8), 723–735 (2010)
27. Panger, G.: Kinect in the kitchen: testing depth camera interactions in practical home environments. In: Proceedings of the 2012 ACM Annual Conference Extended Abstracts on Human Factors in Computing Systems Extended Abstracts (CHI EA 2012), pp. 1985–1990. ACM, New York (2012), doi:10.1145/2212776.2223740
28. Poppe, R.: Vision-based human motion analysis: An overview. Comput. Vis. Image Underst. 108(1-2), 4–18 (2007)
29. Poppe, R.: A survey on vision-based human action recognition. Image Vision Comput. 28(6), 976–990 (2010), doi:10.1016/j.imavis.2009.11.014
30. Mahfujur Rahman, A.S.M., Hossain, M.A., Parra, J., El Saddik, A.: Motion-path based gesture interaction with smart home services. In: Proceedings of the 17th ACM International Conference on Multimedia (MM 2009), pp. 761–764. ACM, New York (2009)
31. Ruiz, J., Li, Y., Lank, E.: User-defined motion gestures for mobile interaction. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (CHI 2011), pp. 197–206. ACM, New York (2011)
32. Sadri, F.: Ambient intelligence: A survey. ACM Comput. Surv. 43(4), Article 36, 66 (2011), doi:10.1145/1978802.1978815

33. Schlömer, T., Poppinga, B., Henze, N., Boll, S.: Gesture recognition with a Wii controller. In: Proceedings of the 2nd International Conference on Tangible and Embedded Interaction (TEI 2008), pp. 11–14. ACM, New York (2008)
34. van Seghbroeck, G., Verstichel, S., de Turck, F., Dhoedt, B.: WS-Gesture, a gesture-based state-aware control framework. In: Proceedings of the IEEE International Conference on Service-Oriented Computing and Applications, SOCA 2010 (2010)
35. Sonntag, D., Engel, R., Herzog, G., Pfalzgraf, A., Pfleger, N., Romanelli, M., Reithinger, N.: SmartWeb Handheld — Multimodal Interaction with Ontological Knowledge Bases and Semantic Web Services. In: Huang, T.S., Nijholt, A., Pantic, M., Pentland, A. (eds.) ICMI/IJCAI Workshops 2007. LNCS (LNAI), vol. 4451, pp. 272–295. Springer, Heidelberg (2007)
36. Urbieta, A., Barrutieta, G., Parra, J., Uribarren, A.: A survey of dynamic service composition approaches for ambient systems. In: Proceedings of the 2008 Ambi-Sys Workshop on Software Organisation and MonIToring of Ambient Systems (SOMITAS 2008), Brussels, Belgium (2008)
37. Vatavu, R.D.: Point & Click Mediated Interactions for Large Home Entertainment Displays. Multimedia Tools and Applications 59(1), 113–128 (2012)
38. Vatavu, R.D.: Nomadic Gestures: A Technique for Reusing Gesture Commands for Frequent Ambient Interactions. Journal of Ambient Intelligence and Smart Environments 4(2), 79–93 (2012)
39. Wang, D.-X., Xiong, Z.-H., Zhang, M.-J.: An application oriented and shape feature based multi-touch gesture description and recognition method. Multimedia Tools and Applications 58(3), 497–519 (2012)
40. Weinland, D., Ronfard, R., Boyer, E.: A survey of vision-based methods for action representation, segmentation and recognition. Comput. Vis. Image Underst. 115(2), 224–241 (2011)
41. Wobbrock, J.O., Wilson, A.D., Li, Y.: Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In: Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST 2007), pp. 159–168. ACM, New York (2007), doi:10.1145/1294211.1294238
42. Wobbrock, J.O., Morris, M.R., Wilson, A.D.: User-defined gestures for surface computing. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI 2009), pp. 1083–1092. ACM, New York (2009)
43. Xu, J., Lee, Y.-H., Tsai, W.-T., Li, W., Son, Y.-S., Park, J.-H., Moon, K.-D.: Ontology-Based Smart Home Solution and Service Composition. In: Proceedings of the 2009 International Conference on Embedded Software and Systems (ICESS 2009), pp. 297–304. IEEE Computer Society, Washington, DC (2009)

Using Markov Logic Network for On-Line Activity Recognition from Non-visual Home Automation Sensors

Pedro Chahuara[1], Anthony Fleury[2], François Portet[1], and Michel Vacher[1,⋆]

[1] Laboratoire d'Informatique de Grenoble
Grenoble 1/Grenoble INP/CNRS UMR 5217, F-38041 Grenoble, France
{pedro.chahuara,francois.portet,michel.vacher}@imag.fr
[2] Univ. Lille Nord de France, F-59000 Lille, France
EMDouai, IA, F-59508 Douai Cedex, France
anthony.fleury@mines-douai.fr

Résumé. This paper presents the application of Markov Logic Networks(MLN) for the the recognition of Activities of Daily Living (ADL) in a smart home. We describe a procedure that uses raw data from non visual and non wearable sensors in order to create a classification model leveraging logic formal representation and probabilistic inference. SVM and Naive Bayes methods were used as baselines to compare the performance of our implementation, as they have proved to be highly efficient in classification tasks. The evaluation was carried out on a real smart home where 21 participants performed ADLs. Results show not only the appreciable capacities of MLN as a classifier, but also its potential to be easily integrable into a formal knowledge representation framework.

Keywords: Activity Recognition, Markov Logic Network, Support Vector Machine, Smart Home, Ambient Assisted Living.

1 Introduction

In the Ambient Assisted Living domain there is a increased interest in automatic human activity recognition from sensors [1,2,3]. Recognition of human activity is recognised as one important variable for human behaviour monitoring but it is also extensively studied for the provision of context-aware services in smartphones and other smart objects [4].

In this paper, we focus on the recognition of activity within the home which is a private space in which multiple sensors, actuators and home automation equipments coexist. This research is linked to the Sweet-Home project which aims at developing a complete framework to enable voice command in smart homes. In this framework, the interpretation of the orders and the decisions to be made depend on the context in which the user is. This context is composed,

⋆ This work is part of the Sweet-Home project founded by the French National Research Agency (Agence Nationale de la Recherche / ANR-09-VERS-011).

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 177–192, 2012.

among other information, of the user's current activity which is essential for decision making. For instance, if the user utters "turn on the light", the action will be different if she is awaking in the middle of the night (in that case, the best action could be to provide low intensity light using the bedside lamp) or if she is dressing up in the morning (in that case, the best action could be to provide high intensity light using the ceiling lamp). Knowing whether the user is sleeping or dressing permits fine grain decision making and facilitate disambiguation of situations interpretation.

Most of the progresses made in the field of activity recognition come from the computer vision domain [5]. However, the installation of video cameras in the user's home is not only raising ethical questions [1], but is also rejected by some users [6]. Other approaches rely on information from RFID tags [7] and wearable devices [8]. In the first case, putting RFID tags on objects makes fastidious the maintenance of the smart home since any new object implies technical manipulation to attach the corresponding sensor. The case of wearable sensors is sometimes not applicable when inhabitants do not want (or forget) to wear sensors all the time. Moreover, cost and dissemination of assistive technology would be more efficient if it is built on standard home automation technology with minimal technical addition. This is why, the solution developed in the Sweet-Home project is to complete conventional sensors for home automation (infrared presence detectors, switches, etc.) by microphones to allow the user to control her environment through voice recognition.

This type of environment imposes constraints on the sensors and the technology used for recognition. Indeed, information provided by the sensor for activity recognition is indirect (no worn sensors for localisation), heterogeneous (from numerical to categorical), transient (no continuous recording), noisy, and non-visual (no camera). This application setting calls for new methods for activity recognition which can deal with the poverty and unreliability of the provided information, process streams of data, and whose models can be checked by human and linked to domain knowledge. To this end, we present a method based on Markov Logic Network (MLN) to recognise activities of daily living in a perceptive environment. MLN, a statistical relational method, makes it possible to build logical models that can deal with uncertainty. This method is detailed in Section 4. Before this, a state of the art in MLN based activity recognition is given in Section 2 and the Sweet-Home project is introduced in Section 3. The method was tested in an experiment in a real smart home involving more than 20 participants. The experiment and the results are described in Section 5. The paper ends with a discussion and gives a short outlook on further work.

2 State of the Art of Human Activity Recognition in the Home

Automatic recognition of human activity (eating, talking, watching TV, etc..) can be defined as the identification of a sequence of atomic actions (taking a cooking utensil, lying, etc.). This involves abstracting the raw signals into symbols (proposals) temporally labelled (e.g. : door slamming at 11 :32), signatures of specific situations of atomic events are detected through a process of hierarchical abstraction. For example, the movements detected in the bedroom can be part of the activity "get up" which itself can be part of a plan of the day (e.g ; : Sunday morning). Automatic recognition of activity is one of the most active and most ambitious research areas because of the large amount of noise in the data and the difficulty of modelling situations ; for the same person, an activity can take place in many ways.

Approaches for activity recognition can be divided mainly into three categories : statistical, probabilistic and logic. In the former category, machine learning methods have been applied to classify activities from information related to pervasive environments. For instance, Fleury *et al.* [9] proposed Support Vector Machines (SVM) to implement a classifier using data from sensors in a real pervasive environment. If statistical methods such as SVM or Neural Networks can lead to good performance, they lack a formal base to represent uncertainty and the obtained models are not easily interpretable.

As information in pervasive environments is uncertain in most cases, probabilistic approaches are suitable candidates to be applied for activity recognition. For instance, Dynamic Bayesian networks were used by Van Kasteren *et al.* [3] to recognise elders' activities. Considering the temporal nature of activities as a succession of actions or sub activities, literature presenting a modelling of activity by Hidden Markov Models (HMM) is vast. For instance, Duong *et al.* [10] extended a conventional HMM to model the duration of an activity and Naaem *et al.* [11] defined activities as a composition of tasks modelled by hierarchical HMMs. However, despite improved expressiveness, these models require a large amount of training data. These training data are costly to obtain and are often not generalisable to other settings than the one in which they were acquired. Moreover, it remains difficult to integrate *a priori* high-level knowledge in these probabilistic models.

The logical approach offers an ideal framework to model explicit knowledge. Ontologies have been used for this task [12] since they feature readability and formal definitions while the activity recognition can be performed by an ontology reasoner as a problem of satisfiability. Moreover, under the logic approach, logic rules facilitate the implementation of expert knowledge within a model [2]. For instance, Augusto *et al.* [13] used logical models to represent the temporal relations among events to recognise activities. The main drawback of this methods is the lack of systematic handling of uncertainty.

Recently, Statistical Relational Learning (SRL) [14], a sub domain of machine learning, has gained much attention as it integrates elements of logic and probabilistic models. Under the SRL schema, models are defined in a formal logical language that makes them reusable and easy to verify, that systematically take uncertainty into account, and that allows easily inclusion of *a priori* knowledge. SRL has recently attracted attention in the domain of human activity modelling and recognition. For instance, Logic HMM and relational Markov networks are both SRL methods that were considered for activity recognition [15,16]. In our work we proposed to use Markov Logic Networks (MLN). To the best of our knowledge the closed approach to our is the one of Trans *et al.* [17] who detected activities in video streams. Other MLN-based activity recognition methods were defined and tested under different settings from the SWEET-HOME project. For instance, [7] assumed RFID tags but this implies tagging every objects involved in the model and the practicability of the approach can be questioned. In our project, which is described in the next section input data imposes little constraint on the daily file of the user.

3 SWEET-HOME: An Audio-Based Smart Home System

The SWEET-HOME project (`sweet-home.imag.fr`) aims at designing a new smart home system based on audio technology focusing on three main aspects : to provide assistance via *natural man-machine interaction* (voice and tactile command), to ease *social inclusion* and to provide *security reassurance* by detecting situations of distress. If these aims are achieved, then the person will be able to pilot their environment at any time in the most natural way possible.

The input of the SWEET-HOME system is composed of the information from the home automation system transmitted via a local network and information from the microphones transmitted through radio frequency channels. Rather than building communication buses and purpose designed material from scratch, the project tries to make use of already standardised technologies and applications. As emphasised in [18] the interoperability of ubiquitous computing elements is a well known challenge to address. Thus, the use of home automation standards ensure compatibility between devices, ease the maintenance and orient the smart home design toward cheaper solutions. While the home automation system provides symbolic information, raw audio signals must be processed to extract information from speech and sound. The extracted information is analysed and either the system reacts to an order given by the user or the system acts pro-actively by modifying the environment without an order (e.g. turns off the light when nobody is in the room). ù Output of the system thus includes home automation orders but also interaction with the user when a vocal order has not been well understood for example, or in case of alert messages (e.g. turn off the gas, remind the person of an appointment).

The SWEET-HOME system will be piloted by an intelligent controller which will capture all streams of data, interpret them and execute the required actions. The diagram of this intelligent controller is depicted in Figure 1. All the knowledge of the controller is defined using two semantic layers : the *low-level* and the *high-level* ontologies. The former ontology is devoted to the representation of raw data and network information description while the high level ontology represents concepts being used at the reasoning level such as : Actions that can be performed in a home and the context in which a home can be (e.g., making coffee, being late). This separation between low and high levels makes possible a higher re-usability of the reasoning layer when the sensor network and the home must be adapted [19]. The estimation of the current context is carried

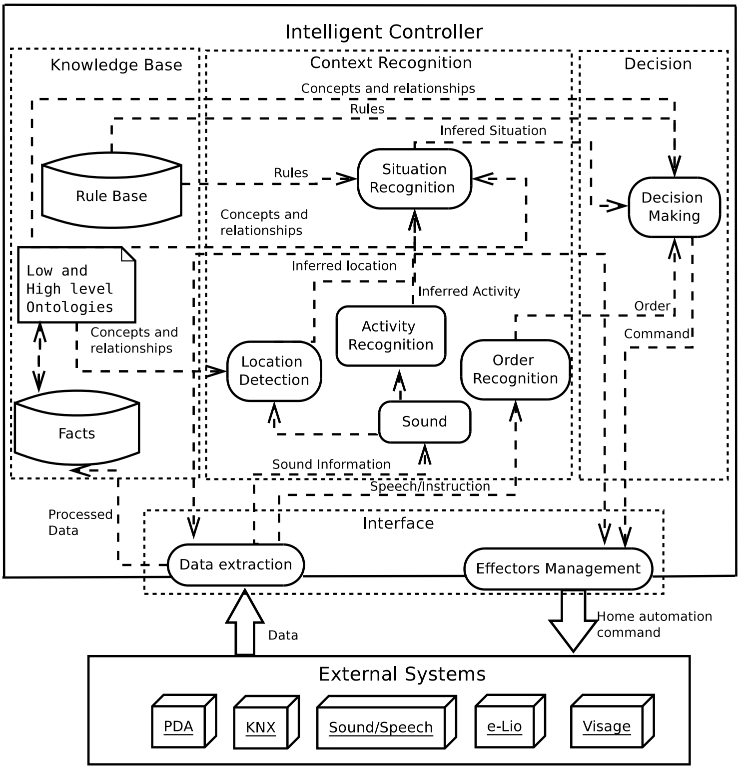


Fig. 1. The Intelligent Controller Diagram

out through the collaboration of several processors, each one being specialised in a certain context aspect, such as location detection or activity recognition. All processors share the knowledge specified in both ontologies and use the same repository of facts. Furthermore, the access to the knowledge base is executed under a service oriented approach that allows any processor being registered to be notified only about particular events and saving any inferred information to be available to other processors. This data and knowledge centred approach permits to ensure that all the processors are using the same data structure and that the meaning of each piece of information is clearly defined among all of them. In addition, the chosen architecture is more flexible than a classical pipeline of processors, making possible the easy insertion of new processors. Once the current context has been determined, the controller evaluates if an action must be taken. What supports the process of decision is a set of logic rules which are part of the knowledge base.

4 Method

As shown in the previous section, activities is an important element of the context. In this study, the activities under consideration are daily living tasks such as sleeping, dressing, eating, communicating, etc. [20] that a person performs during a temporal interval. Given that the data streams provide quite poor, indirect (no wearable sensors) and sporadic (no continuous measurement of some variables) information as they are only composed of classical home automation sensors and microphones data ; fine grain activity recognition, such as screwdriver usage, is impracticable. However, each instance of activity is composed of a set of events that generate observations from the set of home automation sensors. Our hypothesis is that these set of observations are signatures of the activities and that they can be described by statistics of predefined variables computed over temporal windows shorter that the minimal activity duration. Although activities captured in this manner might be gross, we argue that they can be sufficient to provide contextual information to the decision module.

The method to recognise activities from the streams of raw sensor data goes through different levels of abstraction as depicted Figure 2. The raw data are composed of symbolic timestamped values (e.g., infra-red sensors), state values (e.g., switches), time series (e.g., temperature) and signals (e.g., microphones). A pre-processing stage extracts higher-level information such as speech, sounds, location and agitation of the inhabitant. To represent the stream of data, all the raw and abstracted data are summarised as attribute vectors, each of which corresponding to a temporal windows of size W. So, at current time t, every segment of the data of interval $]t - W, t]$ seconds is represented by a vector $v(t)$ which is used as input to a classification model M which gives the main activity a of the person during $]t - W, t]$. All of the classification models were acquired using supervised machine learning techniques. This section summarises the pre-processing stage, and details the attributes and the classifier model.

4.1 Preprocessing

The raw data captured within the smart home (see bottom of Fig. 2) contains information that must be extracted for enhanced activity recognition. Three types of abstraction are considered : localisation of the inhabitant — which is of primary importance for activity recognition —, speech and sound recognition — which is important for activities of communication — and activity level — which specifies how active the person was during the temporal window.

Speech/Sound Detection. In this approach, sound events are detected in real-time by the AuditHIS system [21]. Briefly, the audio events are detected in real-time by an adaptive threshold algorithm based on a wavelet analysis and an SNR (Signal-to-Noise Ratio) estimation. The events are then classified into speech or everyday life sound by a Gaussian Mixture Model. The microphones being omnidirectional, a sound can be recorded simultaneously by multiple microphones ; AuditHIS identifies these simultaneous events.

Localisation. In smart homes, cheap localisation can be performed by using infra-red sensors detecting movement but these sensors can lack sensitivity. To improve localisation, our approach fuses information coming from different data

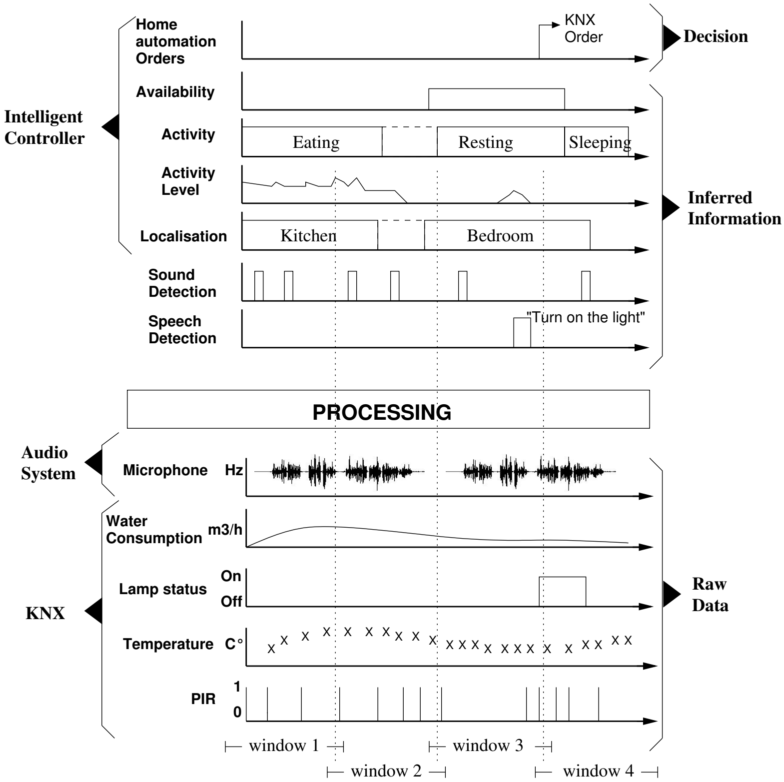


Fig. 2. Temporal windowing for computing the vectors from the sensors data

sources namely, infra-red sensors, door contacts and microphones. The data fusion model is composed of a two-level dynamic network [22] whose nodes represent the different location hypotheses and whose edges represent the strength (i.e., certainty) of the relation between nodes. This method has demonstrated a correct localisation rate from 63% to 84% using several uncertain sources.

Activity Level. In smart homes, the level of activity is a measure of the frequency of actions of the person in the house. This makes it possible to distinguish from activities that ask for many actions (e.g., cleaning the house) from quieter ones (e.g., sleeping). In our approach, level of activity is estimated not only from infra-red sensors but also from sound information as well as door contacts. These sources of data are fused using a linear model. Level of activity — which is a numerical measure— must not be confused with activity — which is the category of daily living activities being performed.

4.2 Generation of the Vectors of Attributes

The traces generated from human activities are difficult to generalise due to the high inter and intra-person variability of realisation of a task. This is why statistic attributes and inferred information were chosen to summarise the information present in each window. In total 69 attributes were extracted for each temporal window. They are summarised in Table 1. In addition to the set of attributes, the seven activities considered in the study are given at the end of the Table. It can be noticed that an unknown class is present. This class is composed of all windows during which none of the seven activities is being performed.

4.3 Markov Logic Network (MLN)

MLN [23] combines first-order logic and Markov Networks. A MLN is composed of a set of first-order formulae each one associated to a weight that expresses a degree of truth. This approach soften the assumption that a logic formula can only be true or false. A formula that does not contain variables is *ground* and is a *ground atom* if it consists of a single predicate. A set of ground atoms is a *possible world*. All possible worlds in a MLN are true with a certain probability which depends on the number of formulae they satisfy and the weights of these formulae. A MLN, however, can also have hard constraints by giving a infinite weight to some formulae, so that worlds violating these formulae have zero probability. Let's consider F a set of first-order logic formulae, $w_i \in \mathbf{R}$ the weight of the formula $f_i \in F$, and C a set of constants. During the inference process,

Table 1. Attributes used for activity recognition (all the attributes are numerical)

Attributes	Number	comments
PourcentageLoc_x	4	ratio of time spent at room x
PredominantLoc	1	most occupied room during the temporal window
LastRoomBeforeWindow	1	last room in which the person was just before the current window
TimeSinceInThisRoom	1	Time elapsed since the person entered the room
ActivationDeactivationWindow_y	6	number of state changes of the door contact of the window y
ActivationDeactivationDoor_w	4	number of state changes of the door contact of the door w
ActivationDeactivationLight_z	6	number of state changes of the light z
ActivationDeactivationCommDoor_f	5	number of state changes of the door contact of the furniture f
DetectionPIR_x	2	number of time the movement detector fired in room x
AmbientSensor	13	difference of value between temporal windows of : CO2, temperature, humidity, brightness, water and electricity
Power_LastUse	3	Id of the last used sockets or switches.
PercentageTime_Sound	1	ratio of time occupied by sounds during the temporal window
PercentageTime_Speech	1	ratio of time occupied by speech during the temporal window
sound_m	7	number of sound occurrences detected by microphone m
speech_m	7	number of speech occurrences detected by microphone m
PercentageAgitation_	6	number of events per windows for the category : room, doors, electricity, water, sounds and speech
TotalAgitation	1	sum of the PercentageAgitation_ in the temporal window
Class	1	eating, tidying up, eliminating, communicating, dressing up, sleeping, resting, unknown

every MLN predicated is grounded and Markov network $M_{F,C}$ is constructed where each random variable corresponds to a ground atom. The obtained Markov network allows to estimate the probability of a possible world $P(X = x)$ by the equation 1 :

$$P(X = x) = \frac{1}{Z} exp \left(\sum_{f_i \in F} w_i n_i(x) \right) \tag{1}$$

where $Z = \sum_{x' \in \chi} exp \left(\sum_{f_i \in F} w_i n_i(x') \right)$ is a normalisation factor, χ the set of possible worlds, and $n_i(x)$ is the number of true groundings of the i-th clause in the possible world x. Exact inference in MLN is intractable in most cases, so Markov Chain Monte Carlo methods are applied [23].

Learning an MLN consists of two independent tasks : structure learning and weight learning. Structure can be obtained by applying machine learning methods, such as Inductive Logic Programming, or rules written by human experts. Weight learning is an optimisation problem that requires learning data. The most applied algorithm in the literature is *Scaled Conjugate Gradient* [**?**].

The approach we implemented uses a list or rules modelling the capacity of each feature value to discriminate an activity. Formally, this rules have the following structure $feature_i(X, V_i) \rightarrow class(X, V_c)$ where the variables X, V_i, and V_c represent : the temporal window to be classified, the value of the i^{th} feature, and the value of the target class. Before creating the model, all numerical variables were discretised. In addition, the temporal relation between instances were modelled by means of the following rule $previous(X, V_c) \rightarrow class(X, V_c)$. The adequacy of including this rule comes from the fact that daily activities use to follow a certain pattern within the routine of an inhabitant. The weights for this model were learnt by *Scaled Conjugate Gradient*. Inference on test data is performed as though it was an on-line system. Thus, the classification of a temporal window not only considers the features describing it but also the results of the previous window classification.

5 Experiments and Results

The method was applied on data collected in a smart home during an experiment involving 21 persons. This section describes the pervasive environment in which the tests were performed (sec. 5.1), the data set (sec. 5.2) that was acquired and the results of the activity recognition (sec. 5.3).

5.1 Pervasive Environment

In the Sweet-Home project, the main pervasive environment considered is the Domus smart home depicted Figure 3. It was adapted to acquire realistic corpus and test the developed techniques. This smart home was set up by the Multicom team of the Laboratory of Informatics of Grenoble, partner of the project. It is a thirty square meters suite flat including a bathroom, a kitchen, a bedroom and a study, all equipped with sensors and effectors such as infra-red presence detectors, contact sensors, video cameras (used only for annotation purpose), etc. In

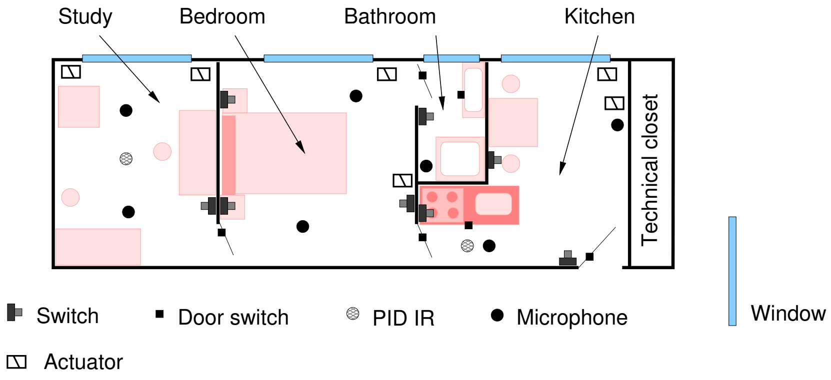


Fig. 3. The Domus Smart Home

addition, seven microphones were set in the ceiling. The technical architecture of Domus is based on the KNX bus system[1], a worldwide ISO standard (ISO/IEC 14543) for home and building control. Bus devices can either be sensors or actuators needed for the control of building equipments such as : lighting, shutters, security systems, energy management, heating, ventilation and air-conditioning systems, interfaces to service and building control systems, remote control, metering, audio video control. . .Besides KNX, several field buses coexist in Domus, such as UPnP (Universal Plug and Play) for the audio video distribution, X2D for the opening detection (doors, windows, and cupboards), RFID for the interaction with tangible objects (not used in the Sweet-Home project). More than 150 sensors, actuators and information providers are managed in the flat. A residential gateway architecture has been designed, supported by a virtual KNX layer seen as an OSGI service (Open Services Gateway Initiative). This layer guarantees the interoperability of the data coming from the different field buses and allows the communication between them and towards virtual applications, such as activity tracking. More than 60 bundles delivering more than 60 services are running. Thanks to this gateway, all home automation elements as well as multimedia elements can be controlled and parametrised remotely.

The sensors that were used in the experiments were the following :

- 7 radio microphones set into the ceiling (2 per room except for the bathroom). These microphones are useful to detect communication activities, recognise speech or detect abnormal sounds ;
- switches and door contact detectors connected to the KNX network which are useful for device usage (fridge, cupboard), informing about movement (e.g., door used between to rooms), knowing whether a window is open or a socket activated.
- Two infra-red movement detectors connected to the KNX network ;
- ambient sensors, such as temperature, humidity, CO_2, etc. ;

[1] www.knx.org

– electricity and water meters ;
– video cameras only used to perform annotation of the data.

5.2 Experimental Data

An experiment was conducted to acquire a representative data set of activities performed at home. The Domus flat was designed to seem as normal as a standard flat (e.g., no visible wire, no prominent sensor) so that any participant's activity instance would be as close as possible to an instance performed in their usual manner. 21 persons (including 7 women) participated to the experiment to record all sensors data in a daily living context. The average age of the participants was 38.5 ± 13 years (22-63, min-max). To make sure that the data acquired would be as close as possible to real daily living, the participants were asked to perform several daily living activities in the smart home. A visit, before the experiment, was organised to make sure that the participants will find all the items necessary to perform the activities. No instruction was given to any participant about how they should behave. All the home automation data was recorded at the virtual KNX layer apart from video data which was recorded separately and audio data which was recorded in real-time thanks to a dedicated PC embedding an 8-channel input audio card [24].

The experiment consisted in following a scenario of activities without condition on the time spent and the manner of achieving them (having a breakfast, simulate a shower, get some sleep, clean up the flat using the vacuum, etc.). During this first phase, participants uttered 40 predefined casual sentences on the phone (e.g., "Allo", "J'ai eu du mal à dormir") but were also free to utter any sentence they wanted (some did speak to themselves aloud). At the end, more than 26 hours of data were recorded. All the activities as well as other information such as location, position, etc. were annotated using the advene software [2].

The stream of data was analysed window by window to generate 2122 instances described by the 69 attributes detailed in Section 4.2. The temporal window size was chosen empirically to be 1 minute. Moreover, given that some activities might intersect with several windows, a 25% overlap was applied. This gives an on-line system that recognises activities every 45 seconds.

5.3 Results

The MLN approach was applied to the experimental data collected in the smart home. In order to assess improvement, we also applied a Naive Bayes (NB) classifier known to be simplistic but often efficient and a more elaborate approach based on Support Vector Machine (SVM) which showed very good results on a similar task [9]. These two classifiers NB and SVM provide the baseline results. The learning and testing strategy used was leave-one-out which consisted in learning models successively on 20 participants and testing on the remaining

[2] `liris.cnrs.fr/advene/`

Table 2. Overall Accuracy, precision and recall

Method	SVM		Naive Bayes		MLN	
Accuracy	59.64		66.1		85.25	
	Precision	Recall	Precision	Recall	Precision	Recall
Eating	64.8	71.0	75.1	79.8	90.4	91.9
Tiding Up	40.0	39.0	58.3	56.4	75.1	86.9
Hygiene	55.8	57.4	67.4	61.7	82.6	80.9
Communicating	83.7	71.9	40.9	47.4	100.0	82.5
Dressing Up	32.3	41.1	13.3	11.8	85.3	56.9
Sleeping	57.6	60.1	60.1	74.3	84.7	82.2
Resting	81.5	73.5	82.4	70.8	90.2	92.2
Unknown	10,2	8.2	19.7	17.9	63.6	25.0

one. The results on the 21 combinations are then averaged to give the overall performance.

Table 2 shows the precision and recall for the methods. The overall accuracy achieved with MLN, 85.3%, is significantly higher than the one obtained with SVM, 59.6% and Naive Bayes, 66.1% . The unknown class case, which corresponds to temporal windows that have not been labelled, exhibits the lowest performance due to the fact that none of the methods can characterise the class. These unlabelled windows can, indeed, take place in very different circumstances, mainly within the transition of two activities, what makes difficult to model them. When the Unknown class is excluded from the data set the overall performances significantly increase showing again the very good performance of MLN (acc. = 90.5%).

We believe that, in spite of its simplicity, our MLN model is well suited for activity modelling and classification because it covers exhaustively all the possible relations among sensors and activities. Even more, in this particular case it seems sound to assume that a sensor evidence about an activity being performed is independent of the other sensors information given the activity. Consequently, our model gives very acceptable results relying mostly in rules modelling independently the influence of a sensor value to recognise an activity. We show below some of the rules composing the MLN model having the highest weights after the process of weight learning :

$2.17 : PercentageLocationBedroom(win, High) => class(win, Sleeping)$
$2.14 : PercentageAgitationElectricity(win, Medium) => class(win, TidyingUp)$
$1.94 : SpeechStudy(win, High) => class(win, Communication)$
$1.85 : SoundBedroomMic1(win, Medium) => class(win, Dressing)$

This subset of rules characterises the implicit knowledge learnt by the model. Many of the rules having a high weight are concerned with location of the inhabitant, what is easily explained by the fact that most activities are performed in specific rooms. Indeed, several previous studies highlighted the importance of location for activity recognition. However, other information appropriately complement location, as in the above example a medium electricity consumption helps to detect the vacuum cleaner usage and consequently the tidying up activity.

Table 3 shows the confusion matrix for the recognition with MLN taking the unknown class into account. The matrix exhibits high values in the confusion of eating and tidying up since they are often performed in the same location (tidying up is also preformed in the bedroom) and complementary information are similar in some occasions : water consumption, sounds, opening of placard doors. Moreover, the confusion of these two activities represents the main difference in accuracy results among SVM and MLN. In general, confusion is higher when activities share a common location as : Tidying up/sleeping and Tidying up/resting. From these results we can conclude that MLN models better the complementary information to discriminate activities performed under similar settings. Actually, the set of learnt rules in our MLN model characterises any detail of a sensor value that can be useful to recognise an activity. Additionally, note that a weighted set of logic rules is a good model to represent knowledge. Being highly readable and clear, its analysis can help to understand core differences between sensor information related to activities in pervasive environments. It also can be noticed that the use of microphones, which is an essential feature of the Sweet-Home project, proved to be very important for activity recognition. Indeed, detection of speech is fundamental to recognise communication activity. This explains the high accuracy of this activity shown in the confusion matrix. Sound plays also an important role; for instance, the amount of sound in the bedroom can resolve the ambiguity between dressing up and sleeping, which are both performed in the same location.

An important finding drawn from these results is that SVM and Naive bayes models are more easily affected by the unbalanced number of instances. Thus, MLN performs better on classification of classes having few instances such as dressing up and hygiene. In real daily routines, some activities have a short duration and are carried out just a few times, then the capacity of MLN to deal with unbalanced training data seems appropriate to be applied in smart homes.

Table 3. Confusion Matrix for MLN results with Unknown class)

True class/Prediction	1	2	3	4	5	6	7	8
Eating(1)	546	47	0	0	0	1	0	0
Tiding Up(2)	23	373	3	0	1	18	4	7
Hygiene(3)	14	6	114	0	2	0	5	0
Communicating(4)	0	2	0	47	0	0	8	0
Dressing Up(5)	2	8	6	0	29	1	5	0
Sleeping(6)	17	9	11	0	0	221	11	0
Resting(7)	0	22	0	0	1	9	458	5
Unknown(8)	2	30	4	0	1	11	17	21

6 Discussion and Perspective

The findings of our study allow us to assert that MLN based models have features that make them more adapted than traditional classifiers for activity recognition. The main advantages of the presented MLN model are : the completeness to represent every detail of sensor values and activity relations ; the readability of the model to understand which sensors explain better the occurrence of an activity ; the possibility of including additional knowledge to the model through logic rules, as it was done with the temporal relations among instances ; and the ability to learn from unbalanced data sets.

Another great advantage of MLN rely on its formal logic nature. In Ambient Intelligent systems, the domain knowledge is often represented as logic formulae (e.g., Description Logic in the case of the OWL format). When possible, this permits translation from one representation to another to performs, for instance, consistency checking of MLN models or addition of relational knowledge as a priory knowledge in the MLN structure learning. In this perspective, the use of a formal domain knowledge description and logic-based recognition method could leads to a higher re-usability of the model learnt in one home to another home. Given the difficulty and cost of acquiring training data in the smart home domain this way seems promising to alleviate the need of large volumes of training data of purely statistical methods.

On the uncertainty modelling, the MLN being a probabilistic model, the outputs of the MLN inference can be exploited and fused with other uncertain evidence in the context-aware decision as it is the case in the intelligent controller described in Section 3. This ability might be highly interesting in other challenging application such as recognition of interleaving activities. In addition, we believe that some characteristics of MLN for activity recognition go beyond the mere classification. Thus, the analysis of set of logic rules composing the MLN and their weights can give us rich information to be exploited during the design of the Sweet-Home context aware system.

References

1. Tapia, E.M., Intille, S.S., Larson, K. : Activity recognition in the home using simple and ubiquitous sensors. Pervasive Computing 2, 158–175 (2004)
2. Storf, H., Becker, M., Riedl, M. : Rule-based activity recognition framework : Challenges, technique and learning. In : Pervasive Computing Technologies for Healthcare, London, UK, pp. 1–7 (2009)
3. van Kasteren, T., Krose, B. : Bayesian activity recognition in residence for elders. In : 3rd IET International Conference on Intelligent Environments, Ulm, Germany, pp. 209–212 (2007)
4. Coutaz, J., Crowley, J.L., Dobson, S., Garlan, D. : Context is key. Communications of the ACM 48(3), 49–53 (2005)
5. Aggarwal, J., Ryoo, M. : Human activity analysis : A review. ACM Comput. Surv. 43, 1–43 (2011)

6. Portet, F., Vacher, M., Golanski, C., Roux, C., Meillon, B. : Design and evaluation of a smart home voice interface for the elderly — acceptability and objection aspects. Personal and Ubiquitous Computing (to appear)
7. Helaoui, R., Niepert, M., Stuckenschmidt, H. : A Statistical-Relational Activity Recognition Framework for Ambient Assisted Living Systems. In : Augusto, J.C., Corchado, J.M., Novais, P., Analide, C. (eds.) ISAmI 2010. AISC, vol. 72, pp. 247–254. Springer, Heidelberg (2010)
8. Zappi, P., Lombriser, C., Stiefmeier, T., Farella, E., Roggen, D., Benini, L., Tröster, G. : Activity Recognition from On-Body Sensors : Accuracy-Power Trade-Off by Dynamic Sensor Selection. In : Verdone, R. (ed.) EWSN 2008. LNCS, vol. 4913, pp. 17–33. Springer, Heidelberg (2008)
9. Fleury, A., Vacher, M., Noury, N. : SVM-based multi-modal classification of activities of daily living in health smart homes : Sensors, algorithms and first experimental results. IEEE Transactions on Information Technology in Biomedicine 14(2), 274–283 (2010)
10. Duong, T., Phung, D., Bui, H., Venkatesh, S. : Efficient duration and hierarchical modeling for human activity recognition. Artificial Intelligence 173(7-8), 830–856 (2009)
11. Naeem, U., Bigham, J. : Activity recognition using a hierarchical framework. In : Pervasive Computing Technologies for Healthcare, pp. 24–27 (2008)
12. Chen, L., Nugent, C.D. : Ontology-based activity recognition in intelligent pervasive environments. IJWIS 5(4), 410–430 (2009)
13. Augusto, J.C., Nugent, C.D. : The use of temporal reasoning and management of complex events in smart homes. In : European Conference on Artificial Intelligence, pp. 778–782 (2004)
14. Getoor, L., Taskar, B. : Introduction to Statistical Relational Learning. The MIT Press (2007)
15. Natarajan, S., Bui, H.H., Tadepalli, P., Kersting, K., Wong, W. : Logical Hierarchical Hidden Markov Models for Modeling User Activities. In : Železný, F., Lavrač, N. (eds.) ILP 2008. LNCS (LNAI), vol. 5194, pp. 192–209. Springer, Heidelberg (2008)
16. Liao, L. : Location-based Activity Recognition. PhD thesis, University of Washington (2006)
17. Tran, S.D., Davis, L.S. : Event Modeling and Recognition Using Markov Logic Networks. In : Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 610–623. Springer, Heidelberg (2008)
18. Edwards, W., Grinter, R. : At Home with Ubiquitous Computing : Seven Challenges. In : Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 256–272. Springer, Heidelberg (2001)
19. Klein, M., Schmidt, A., Lauer, R. : Ontology-centred design of an ambient middleware for assisted living : The case of soprano. In : Kirste, T., König-Ries, B., Salomon, R. (eds.) Towards Ambient Intelligence : Methods for Cooperating Ensembles in Ubiquitous Environments (AIM-CU), 30th Annual German Conference on Artificial Intelligence (KI 2007), Osnabrück, September 10 (2007)
20. Katz, S. : Assessing self-maintenance : activities of daily living, mobility, and instrumental activities of daily living. Journal of the American Geriatrics Society 31(12), 721–727 (1983)

21. Vacher, M., Fleury, A., Portet, F., Serignat, J.F., Noury, N. : Complete Sound and Speech Recognition System for Health Smart Homes : Application to the Recognition of Activities of Daily Living, pp. 645–673. Intech Book (2010)
22. Chahuara, P., Portet, F., Vacher, M. : Fusion of audio and temporal multimodal data by spreading activation for dweller localisation in a smart home. In : STAMI, Space, Time and Ambient Intelligence, Barcelona, Spain, July 16-22, pp. 17–22 (2011)
23. Richardson, M., Domingos, P. : Markov logic networks. Machine Learning 62(1-2), 107–136 (2006)
24. Vacher, M., Portet, F., Fleury, A., Noury, N. : Development of audio sensing technology for ambient assisted living : Applications and challenges. International Journal of E-Health and Medical Communications 2(1), 35–54 (2011)

Multi-Classifier Adaptive Training: Specialising an Activity Recognition Classifier Using Semi-supervised Learning

Božidara Cvetković, Boštjan Kaluža, Mitja Luštrek, and Matjaž Gams

Jožef Stefan Institute,
Department of Intelligent Systems,
Jamova ceta 39, 1000 Ljubljana, Slovenia
{boza.cvetkovic,bostjan.kaluza,mitja.lustrek,matjaz.gams}@ijs.si
http://ijs.dis.si

Abstract. When an activity recognition classifier is deployed to be used with a particular user, its performance can often be improved by adapting it to that user. To improve the classifier, we propose a novel semi-supervised Multi-Classifier Adaptive Training algorithm (MCAT) that uses four classifiers. First, the General classifier is trained on the labelled data available before deployment. Second, the Specific classifier is trained on a limited amount of labelled data specific to the new user in the current environment. Third, a domain-independent meta-classifier decides whether to classify a new instance with the General or Specific classifier. Fourth, another meta-classifier decides whether to include the new instance in the training set for the General classifier. The General classifier is periodically retrained, gradually adapting to the new user in the new environment where it is deployed. The results show that our new algorithm outperforms competing approaches and increases the accuracy of the initial activity recognition classifier by 12.66 percentage points on average.

Keywords: semi-supervised learning, adaptation to the user, MCAT, activity recognition.

1 Introduction

Activity recognition applications are often faced with the problem that a classifier trained in a controlled environment can demonstrate a high accuracy when tested in such environment, but a drastically lower one when deployed in a real-life situation. This issue could be resolved if one were to label enough data specific for the user in the real-life situation where the classifier is deployed. However, since this is often unpractical, semi-supervised learning can be employed to label the data of real users automatically. In semi-supervised learning, one or multiple classifiers usually classify each instance, some mechanism selects the final class based on their outputs, and if the confidence in this class is sufficient, the class is used to label the instance and add it to the training set of the classifiers. This

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 193–207, 2012.

approach raises three challenges that need to be addressed: (i) how to design the multiple classifiers and what data to use for training each of them; (ii) how to choose the final class; and (iii) how to decide whether an instance will be added to the training set of the classifiers.

This paper presents a novel algorithm for adaptation of an activity recognition classifier to a user called Multi-Classifier Adaptive Training (MCAT). It addresses all the above-mentioned challenges. The algorithm is based on the following key contributions: (i) it introduces two classifiers for labelling, a General one, which is trained on the general activity recognition domain knowledge, and a Specific one, which is trained on a limited number of labelled instances specific for the current end-user; (ii) the selection of the final class is handled by a meta-classifier, which uses the specific knowledge to improve the general knowledge of the activity recognition domain; and (iii) the decision about which instance to include in the training set is tackled with another meta-classifier, which weighs the decision of the first meta-classifier. The two meta-classifiers allow us to combine the general knowledge of the activity recognition domain with the specific knowledge of the end-user in the new environment in which the classifiers are deployed. The final contribution of the paper is the training procedure, which uses all the available data to properly train each of the four classifiers.

The algorithm was implemented and evaluated on an activity recognition domain based on Ultra-Wideband (UWB) localisation technology. The people had four wearable sensors attached to their clothes (chest, waist, left ankle, right ankle). The general activity recognition domain knowledge was induced from the data of people performing activities wearing the sensors in the laboratory. The specific data was obtained from an individual person to whom the system is adapted. The proposed approach is compared to two non-adaptive approaches and to three adaptive approaches: the initial version of the proposed approach [1], the self-learning algorithm [2] and majority vote algorithm [3,4].

The results show that the MCAT method successfully increases the accuracy of the initial activity recognition classifier and significantly outperforms all three compared methods. The highest absolute increase in accuracy, when using the MCAT method, is by 20.58 percentage points.

The rest of the paper is structured as follows. Section 2 contains the related work on the semi-supervised learning approaches and their use in adaptation of the activity recognition. The motivating domain is introduced in Section 3. The MCAT method is explained in the Section 4 and Section 5 contains the experiment and the results. Section 6 concludes the paper.

2 Related Work

Semi-supervised learning is a technique in machine learning that uses both labelled and unlabelled data. It is gaining popularity because the technology makes it increasingly easy to generate large datasets, whereas labelling still requires human effort, which is very expensive. The main idea of the semi-supervised approach is to use either i) supervised learning to label unlabelled data or ii) to utilise additional labelled data with unsupervised learning.

A similar approach to semi-supervised learning is active learning. This approach uses supervised learning for the initial classification. However, when the classifier is less confident in labelling the human annotator is required [5,6]. Since human interaction is undesirable the Active learning is inappropriate.

There are four ways of categorising semi-supervised learning approaches [2]: (i) single-classifier and multiple-classifier; (ii) multi-view and single-view; (iii) inductive and transductive; and (iv) classifier- and database-based approaches. The single classifier methods use only one classifier for classification task, where multiple classifiers use two or more classifiers. The key characteristic of a multi-view method is to utilise multiple feature-independent classifiers in one classification problem. Single-view methods use classifiers with the same feature vector but differentiate in the algorithm used for learning. Inductive methods are those that first produce labels for unlabelled data and secondly a classifier. Transductive methods only produce labels and don't generate a new classifier. Classifier-based approaches start from one or more initial classifiers and enhances them iteratively. The database-based approaches discover an inherited geometry in the data, and exploits it to find a good classifier. In this paper we will focus on a single-view, inductive and classifier-based semi-supervised learning.

The most common method that uses a single classifier is called self-training [2]. After an unlabelled instance is classified, the classifier returns a confidence in its own prediction, namely the class probability. If the class-probability threshold is reached, the instance is added to its training set and the classifier is retrained. The self-training method has been successfully applied to activity recognition by Bicocchi et al. [7]. The initial activity recognition classifier was trained on the acceleration data and afterwards used to label the data from a video camera. This self-training method can be used only if the initial classifier achieves high accuracy, since the errors in confident predictions can decrease the classifier's accuracy. The self-training has also been successfully applied on several other domains such as handwriting word recognition [8], natural language processing [9], protein-coding gene recognition [10].

Co-training [11] is a multi-view method with two independent classifiers. To achieve independence, the attributes are split into two feature subspaces, one for each classifier. The classifier that surpasses a confidence threshold for a given instance can classify the instance. The instance is afterwards added to the training set of the classifier that did not surpass the confidence threshold. A major problem of this algorithm is that the feature space of the data cannot always be divided orthogonally. If the attributes are split at random it is possible that classifiers do not satisfy the requirement of self-sufficiency.

The modified multi-view Co-training algorithm called En-Co-training [12] was used in the domain of activity recognition. The method uses information from 40 sensors, 20 sensors on each leg to identify the posture. The multi-view approach was changed into single-view by using all the data for training three classifiers with the same feature vector and a different learning algorithm, which is similar to previously mentioned democratic Co-learning. The final decision on the classification is done by majority voting among three classifiers and the classified instance is added into the training set for all classifiers.

Democratic co-learning [13] is a single-view technique with multiple classifiers. All the classifiers have the same set of attributes and are trained on the same labelled data but with different algorithms. When an unlabelled instance enters the system, all the classifiers return their class prediction. The final prediction is based on the weighed majority vote among the classifiers. If the voting results return sufficient confidence, the instance is added into the training set of all the classifiers.

The MCAT method uses two classifiers; both are trained with the same algorithm but on different data. We use a third, meta classifier, to make the final prediction on the class. The decision whether to put an instance into the training set or not is solved by employing another meta classifier and not slightly arbitrarily like in the case of the Democratic co-learning. In contrast to Co-training, our two domain classifiers have the same attribute set, thus the problem of dividing the sets is not present.

3 Motivating Domain

In this paper the MCAT method is applied to activity recognition, a very common domain in ambient intelligence. Activity recognition classifier is usually trained on the data retrieved from the people performing activities in a controlled environment such as a research laboratory. The classifier trained in this fashion can perform with high accuracy when tested in the same environment, but it is likely that it will perform poorly when deployed in a new environment with a new person, since each person tends to have specific manner of performing the activities. We have faced this problem during the development of the Confidence system [14][15].

The Confidence system is a real-time system developed for constant monitoring of human activity and detection of abnormal events such as falls, or unusual behaviour that can be a result of a developing disease. It is based on a multi-agent architecture where each module, task or activity is designed as an agent providing a service. The multi-agent architecture is shown in Figure 1. It consists of seven agent groups: (i) sensor agents that serve raw localisation data to the next group of agents, (ii) refining agents that filter out the noise, (iii) reconstruction agents that determine the location and activity of the user and are the main trigger for other agents, (iv) interpretation agents that try to interpret the state of the user and provide the emergency state information, (v) prevention agents that observe the user and detect possible deviations in the behaviour, (vi) cognitive agents that monitor the cognitive state of the user, and (vii) communication agents dedicated to user interaction in case of emergency. For detailed description of the multi-agent architecture of the Confidence system the reader is referred to [16].

The input to the system or the sensor agents are the coordinates of the Ubisense location tags [17] attached to the user's waist, chest and both ankles. Since the Ubisense system is noisy, we use three filters, implemented in the refinement agents group, to reduce it. First the data is processed by Median filter, secondly the data is processed with Anatomic filter, which applies anatomic constraints imposed by the human body and the last filter is the Kalman filter. For more details on the filters the reader is refereed to [18]. To get a representation for all the tag positions in time, the snapshots are created with frequency of 4 Hz. Each snapshot is augmented with positions and number of attributes, which are used for activity recognition and other purposes. Primarily the activity recognition was developed to enhance the fall detection, therefore the accuracy of the activity recognition is crucial. The reader is referred to [19] for details about the used attributes.

The activity recognition agent is included in the reconstruction agents group and is able to recognise nine atomic activities: standing, sitting, lying, falling, on all fours, sitting on the ground, standing up and sitting down. The high accuracy of the activity recognition is an essential information to be passed to the other agents for further effective reasoning about the abnormal events. Adapting the activity recognition agent to each user helps achieve that. Our method that does exactly that is described in the next section.

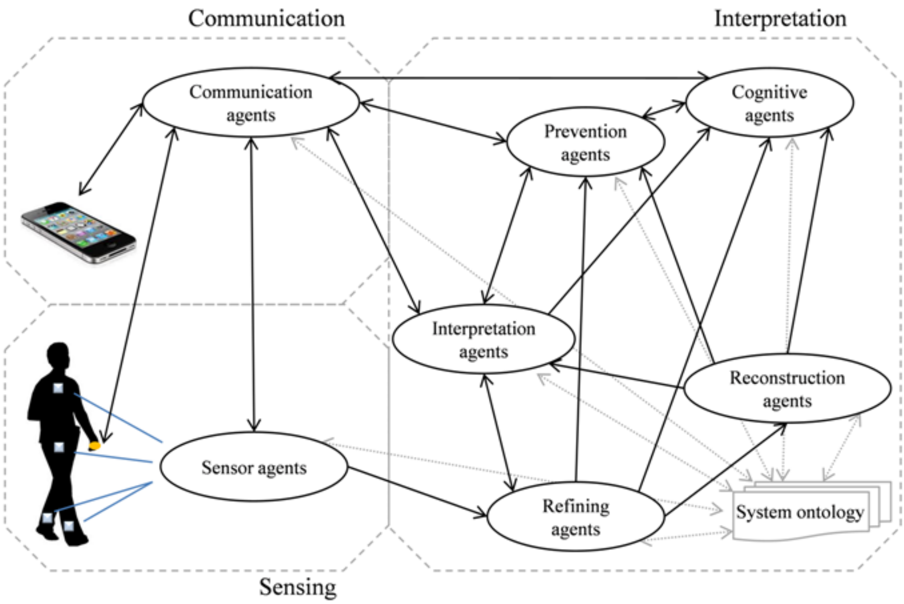


Fig. 1. The multi-agent system architecture of the Confidence system

4 The Multi-Classifier Adaptive Training Method (MCAT)

In this section we propose the MCAT method that improves the classification accuracy of an initial activity recognition classifier utilising unlabelled data and auxiliary classifiers. Before deploying the classifier in a real-world environment with the new user, it is usually trained in a controlled environment on a limited amount of data. In addition to this general approach, a small amount of labelled data from the new real-world environment and a new user is obtained. The MCAT method uses the data from the new environment to adapt the initial classifier to the specifics of the environment while using it.

Consider the following example: The initial classifier is trained on activities performed by several people. When using this classifier on a new person whose physical characteristics are different, and who was not used for training, the recognition accuracy can be low, since each person has specific movement signature. The MCAT method utilises a few activities performed by the new person to learn his/her specifics and thus improve upon the classification of the initial classifier.

4.1 The Algorithm

The proposed MCAT algorithm is shown in Figure 2. The General classifier is the initial classifier trained on the general domain knowledge in the controlled environment. This would be the only classifier deployed in a typical application that does not use the proposed method. To improve the General classifier, we propose a set of three auxiliary classifiers: (i) the Specific classifier, which is trained on a small subset of manually labelled data; (ii) the Meta-A classifier, which decides which classifier (Specific or General) will classify an instance; and (iii) the Meta-B classifier, which decides whether the instance should be included in the training set of the General classifier.

The *General classifier* is trained on a general set of labelled data available for the activity recognition domain, i.e., a controlled environment. The attributes are domain-specific and the machine-learning algorithm is chosen based on its performance on the domain.

The *Specific classifier* is trained on a limited amount of labelled data specific for the new environment in which the classifiers are deployed (new person). Note that this limited dataset does not necessary contain all the classes that are present in the dataset of the General classifier. This may happen due to an unbalanced distribution of class labels, for example, in the activity-recognition domain quick and short movements such as falls are rare. The classes known to the Specific classifier are denoted as basic classes. The attributes and the machine-learning algorithm should preferably be the same as those used in the construction of the General classifier.

After both classifiers return their classes, the *Meta-A classifier* is activated. The decision problem of the Meta-A classifier is to select one of the classifiers to classify a new instance. The Meta-A classifier can be trained with an arbitrary machine-learning algorithm. The attributes for the Meta-A classifier should describe the outputs of the General and Specific classifier as completely as possible, while remaining domain-independent. If we add domain attributes to the meta-attributes, the decision of the Meta-A classifier will also be based on the specifics of the training data available prior to the deployment of the classifiers, which may be different from the specifics of the situation in which the classifiers are

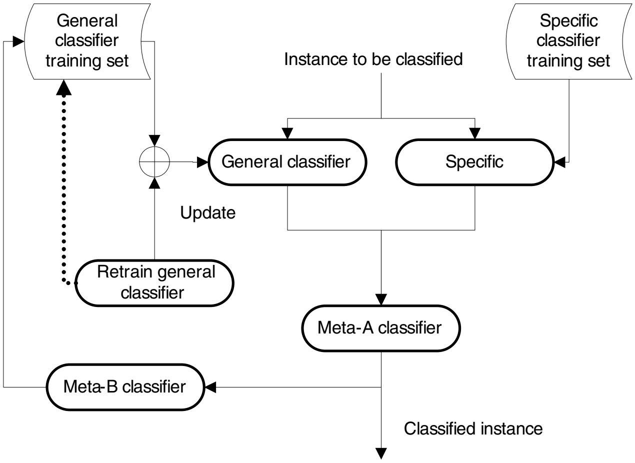


Fig. 2. The work flow of the algorithm proceeds as follows. The method contains two activity recognition classifiers, the General and Specific, and two additional meta-classifiers. The Meta-A classifier decides on the final class of the instance and the Meta-B classifier decides whether the instance is to be included in the training set of the General classifier.

deployed. It was experimentally shown in previous work [1] that domain attributes do not contribute to higher accuracy of the classifier, on the contrary, they can result in over-fitting to the specifics of the training data.

Although the attributes in two Meta-A classifiers, deployed in two different systems, need not be exactly the same, we can propose the following set of attributes considering the previous research: the class predicted by the generic classifier (C_G), the class predicted by the Specific classifier (C_S), the probability assigned by the generic classifier to C_G, the probability assigned by the Specific classifier to C_S, is the C_G one of the basic classes, are the classes C_G and C_S equal, the probability assigned by the generic classifier to C_S, the probability assigned by the Specific classifier to C_G. For more information on the tested sets of attributes and algorithms the reader is referred to [1].

The *Meta-B classifier* is used to solve the problem of whether an instance should be included in the General classifier's training set. The output of the Meta-B classifier should answer the question whether the current instance contributes to a higher accuracy of the General classifier. This question is not trivial and there are several approaches that specifically address it [20]. We use a heuristic instead, which answers the question: "Did the Meta-A classifier select the correct class for the current instance?". The heuristic performs well and is computationally inexpensive, so we left the investigation of more complex approaches for future work. The attributes used in the Meta-B classifier are the same as those used in the Meta-A classifier, with one addition: the confidence of Meta-A classifier in its prediction. The Meta-B classifier can be trained with an arbitrary machine-learning algorithm.

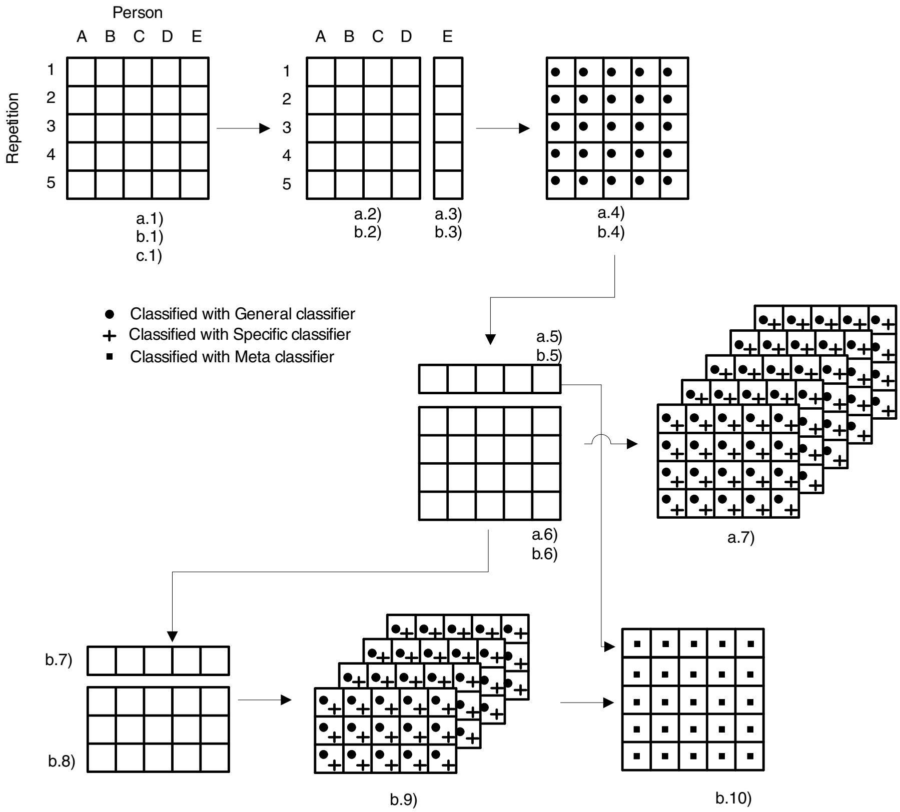


Fig. 3. Training of the classifiers. The figure shows three steps (a, b and c), each resulting in the training set for an individual classifier. *Step a* results in the training set for the Meta-A, *step b* results in the training set for Meta-B and *step c* in the training set for the General classifier. Each step is composed of sub- steps marked with numbers.

4.2 Training Procedure

Training of the four classifiers requires the data to be separated into non-overlapping datasets. This means that the data used to train the Meta-A classifier must not be used to train the General or Specific classifier, and the data used to train the Meta-B classifier must not be used to train these two or the Meta-A classifier. Since we are usually provided with a limited amount of data, our efforts must be focused on maximally utilising all the data.

We propose a training procedure that divides the data into several sets for training both meta-classifiers and the General classifier. The procedure consists of three steps as shown in Figure 3. The steps are marked with a, b and c, each consisting of several sub-steps. The first step trains the Meta-A classifier, marked with sub-steps a. It requires data classified by both domain classifiers (the General and Specific). In the second step the Meta-B classifier is trained, marked with sub-steps b. It requires data classified by the Meta-A classifier and data classified by both domain classifiers. The third step includes training of the Generic classifier, marked with step c. It requires only the originally labelled data.

Our procedure is general, but for the purpose of this paper we will make use of the activity-recognition domain. Specifically the publicly available dataset "Localization Data For Person Activity" from the UCI Machine Learning Repository [21]. The dataset consists of recordings of five people performing predefined scenarios five times. Each person had four Ubisense [17] tags attached to the body (neck, waist, left ankle and right ankle) each returning its current position. The goal is to assign one of eleven activities to each time frame. The dataset can be divided by the person five times and by the scenario repetition five times (sub-steps a.1,b.1,c.1).

The first step is building a training set for the Meta-A classifier (sub-steps a). Four persons are selected for training the General classifier (sub-step a.2), which is used to classify the fifth person (sub-step a.3). This is repeated five times for each person and the result is the complete dataset classified with the General classifier (sub-step a.4). The data classified with the General classifier is represented with dots in Figure 3. The data classified with the General classifier is split by repetition five times. Since the specific classifier should be trained on a small amount of data, one repetition for each person is used for building the Specific classifier (sub step a.5) and the remaining four repetitions are classified with the Specific classifier (sub step a.6). The data classified with the Specific classifier is represented with crosses in Figure 3. In total, we have five Specific classifiers (one per person), each of which was used to classify four repetitions. This gives us the data in the sub-step a.7, which represents the training set for the Meta-A classifier.

The second step is building the training set for the Meta-B classifier (sub-steps b). The sub-steps from b.1 to b.4 are executed identically as sub-steps from a.1 to a.4 as explained earlier. The data classified with the General classifier (dotted data in sub-step b.4) is divided by the repetition. One repetition for each person is kept aside for the Meta-B training (sub step b.5). The data that is left contains five people times four repetitions (sub-step b.6). It is used for the temporary Meta-A training. One repetition for each person is used to train the Specific classifier (sub-step b.7). The three repetitions that are left (sub-step b.8) are classified with the General and Specific classifier (sub-step b.9) represented with dots and crosses. The result is used to train the temporary Meta-A classifier. The data that was kept aside for the Meta-B classifier is now retrieved and classified with the temporary Meta-A classifier. The results are stored into the training set for the Meta-B classifier (sub step b.10). After all the repetitions are executed we get the training set (represented with squares) for training the final Meta-B classifier.

The last step is building the dataset for training the General classifier, which is trained on the complete dataset as shown in step c.1.

5 Experimental Evaluation

The experimental evaluation is focused on the activity recognition discussed in the previous sections. The main reason why the general classifier will likely not perform well in a real-life environment on a particular person is that each person has its own specifics such as the height and the characteristics of movement. The general classifier, which is trained on several people, can thus recognise the activities of a "general person". Obtaining enough training data for a particular end user is difficult, so our method is well suited for solving this problem.

5.1 Experimental Setup

The experimental data was collected using the real-time Confidence system described in Section 3. We have collected two different sets of data, one for the classifier training, namely *the training dataset* and one for testing the semi-supervised adaptation to the users, namely the *test dataset*.

The first dataset, *the training dataset* was contributed to the UCI Machine Learning Repository [21]. These data consists of the recordings of five people performing a scenario five times. The scenario is a recording of a person performing a continuous sequence of activities that represents typical daily activities. The data was divided into segments as discussed in Section 4.2 and used for training the classifiers.

The second, the *test dataset* consists of the recordings of ten people, again performing typical daily activities. All the people are different than in the *training dataset*. The scenario was repeated by each person five times, which gives us 2.7 hours of data per person on average, and 30.2 hours altogether. The scenario was designed to reflect the distribution of the activities during one day of an average person. The scenario contains eight activities: standing, lying, sitting, going down, standing up, falling, on all fours and sitting on the ground. The scenario performed by the people in the *test dataset* contains additional repetition of the sitting on the ground activity with all the respective transitions from sitting and afterwards to walking compared to the *training dataset*. Both datasets were primarily recorded for the purpose of fall detection and the difference between them is not significant for our problem.

The *training dataset* contains more labels that the test data, so we had to merge the transition activities. The activities "lying down" and "sitting down" were merged into the activity "going down". The activities "standing up from lying", "standing up from sitting" and "standing up from sitting on the ground" were merged into the activity "standing up". The activity "walking" was merged into "standing". The distributions of the activities per merged class for both datasets are shown in Table 1.

To create datasets needed for the proposed method we divided the *test dataset* as follows: *Basic activity dataset* - ten people performing basic activities (exactly 30 seconds per activity), which are lying, standing and sitting and *Unlabelled test dataset* - ten people performing scenario five times (125 minutes per person on average). The *Basic activity dataset* is used to train the Specific classifier per

Table 1. Activities and their distributions per dataset

Activities	Distributions (%)	
	Training dataset	Test dataset
Standing	19.56	29.55
Sitting	16.67	15.53
Lying	33.14	31.58
Sitting on the ground	7.20	16.09
On all fours	3.05	0.59
Falling	1.81	0.89
Going down	4.80	1.26
Standing up	13.75	4.50

person. The *Unlabelled test dataset* per person is the data used for the semi-supervised adaptation, the core of MCAT.

The experiment was done in two steps: first, the classifier training; and second, the use of MCAT as a semi-supervised adaptation of the initial General classifier.

The classifier training was executed as presented in Section 4.2. The Meta-A classifier was trained using the Support Vector Machines algorithm [22] and the Meta-B classifier using the C4.5 machine learning algorithm [23]. Note that the classifiers could be trained using other machine learning algorithms, but the chosen two algorithms proved satisfying in our experiments. Both were used with default parameters as implemented in the Weka suite [24]. The General classifier was trained using the Random Forest algorithm [25]. The accuracy of the classifier is 86% when evaluated using leave-one-person out approach, which implies that one can expect the same accuracy when the classifier is deployed in the new environment; however, this is not the case as we see in the Table 2 in the column labelled with G (initial general classifier). This algorithm was selected according to the results from previous research [26].

For each test person the Specific classifier was trained using the Random Forest algorithm on the *Basic activity dataset*. The *Unlabelled dataset* of that person was processed with the MCAT method. The General classifier was re-trained after each repetition of the test scenario (five repetitions per person), to take advantage of the instances from the unlabelled dataset that were included in its training dataset.

The data for each person was processed by our method two times. In the first run the instances selected for the inclusion in the training set of the General classifier were weighed with the weight 2. In the second run the weight was 1. The reason for increasing the weight in the first run is to accelerate the adaptation. The experiments showed that in case the weight in the first run is set to the default value 1, the number of non-basic class instances, which are sitting on the ground, falling, standing up, going down, on all fours, selected for the inclusion in the second run is lower. The weight value in the second run is decreased to avoid the elimination of the general knowledge from the General classifier. If an instance already existed in the training set and was selected for the inclusion again, it was discarded.

The MCAT method was compared to five competing methods: two transductive approaches that only label the instance and do not perform the adaptation; and three inductive semi-supervised learning approaches. The transductive approaches are: *the baseline approach*, which first merges the training data of the Generic and the Specific classifier into one training set and then trains a new General classifier (G'); and the *MCAT without Meta-B*(without semi-supervised learning) approach that builds both the Generic and Specific classifier and uses the Meta-A to decide which one to trust.

The inductive semi-supervised learning approaches used for comparison are *self-training*, *majority vote* and *threshold-based MCAT*. The self-training method uses the General classifier for classification. The instances in the *Unlabelled dataset* with the 100% classification confidence are added to its training set. The majority vote uses three classifiers trained on the training set with different machine learning algorithms. We used the algorithms that achieved the highest accuracy in the General classifier evaluation using the leave-one-person out approach. These were the Support Vector Machines, Random Forest and C4.5. The instances in the *Unlabelled dataset* with 100% classification confidence were included in the training set of all classifiers. The threshold-based MCAT uses threshold rule instead of the Meta-B classifier, which selects the instance to be included in the training set of the Generic classifier.

5.2 Results

The classification accuracies of the methods are shown in Table 2. The left side of the table shows the accuracy of the initial General classifier G and the Specific classifier S. The right side of the table shows the gain/loss in accuracy of the methods compared to the initial General classifier. The compared methods are: (i) transductive approaches: baseline method (G&S merged); MCAT without Meta-B (Meta-A) and (ii) semi-supervised learning approaches: threshold-based MCAT (previous version PV); Self-training (ST) and Majority vote (MV) algorithm. The last column shows the results of the MCAT method.

The results for the General classifier show a decrease in accuracy when used in the different environment than the controlled where the accuracy was 86%. The average accuracy on ten people is 70.03% (Table 2, column *Initial G*). The results of the Specific classifier (Table 2, column *Specific S*) show that it achieves a higher accuracy than the initial General classifier overall and in several individual cases, even though it is able to predict only three basic classes (lying, standing, sitting).

The results on the baseline training set are the worst compared to the General classifier accuracy for four of the people. This is because some instances from the Specific and General classifier are similar but differ in the label. The results of using the Meta-A classifier to select the final class show that this method outperforms the General classifier by 6.14%. The higher accuracy is gained because of the knowledge of basic classes, which represent 76.67% of the dataset on average and by 16.81 percentage points in the best case (Person I). This increase in accuracy reveals that using a classifier for class selection in one of the reasons for the success of the proposed approach, while the other reason is the semi-supervised adaptation. The average accuracy of the General classifier after adaptation using the previous version is 78.59% and the average gain in accuracy is 8.56 percentage points. The results of the Self-training show that in a few cases, where the General classifier has low initial accuracy, the method performs poorly and further weakens the classifier. The average accuracy after adaptation is 71.91%. The results of the Majority vote method show that introducing extra classifiers contributes to gain in accuracy upon the initial classifier. The average accuracy after adaptation using the majority vote method is 74.29%.

The results for the MCAT method show the average gain in accuracy of 12.66 percentage points and the average accuracy of the classifier 82.70%. In the best case (Person C) it achieves an increase in accuracy of 20.58 percentage points and in the worst case (Person B) 7.38 percentage points. The highest absolute accuracy is achieved for Person J, 85.32%, and Person A, 84.46%, where both the General and the Specific classifier returned a similar initial accuracy. The MCAT method outperformed the self-training method by 10.79 percentage points, the majority vote by 8.41 percentage points and the previous version by 4.11 percentage points.

Table 2. Classifier accuracies and comparison of the MCAT method to the (i) transductive approaches: baseline (G and S merged G&S), MCAT without adaptation (Meta-A) and (ii) semi-supervised approaches: threshold-based MCAT (previous version PV), Self-training (ST) and Majority vote (MV) method

	Classifier Accuracy(%)		Method Comparison (gain/loss in pp against G)					
Person	Initial G	Specific S	G&S	Meta-A	PV	ST	MV	MCAT
A	75.28	76.82	-10.57	+3.62	+3.82	+2.36	+0.38	+9.18
B	76.28	60.06	+1.17	+0.58	+0.70	+0.28	+0.64	+7.38
C	62.87	70.85	+8.52	+12.82	+13.46	-1.18	+3.07	+20.58
D	69.55	76.17	-0.21	+8.99	+9.77	+2.23	+2.10	+12.57
E	68.13	74.23	+0.20	+5.78	+9.76	-1.81	+6.73	+16.22
F	73.57	68.18	-4.42	+2.62	+8.08	+6.07	+3.67	+8.86
G	65.42	67.72	5.97	+6.99	+9.76	-0.21	+8.57	+12.60
H	73.45	67.46	-8.02	+0.01	+4.51	+4.31	+0.41	+10.53
I	62.09	68.08	+7.75	+16.81	+16.98	-0.18	+11.03	+17.08
J	73.67	74.17	+1.18	+3.15	+8.73	+6.87	+5.98	+11.65
Average	70.03	70.37	+0.16	+6.14	+8.56	+1.87	+4.26	+12.66

6 Conclusion

This paper is focused on the problem of enhancing an initial activity recognition classifier using unlabelled data. The main contributions of this paper are: (i) a novel method for specialising the activity recognition classifier by semi-supervised learning MCAT; and (ii) a procedure for training multiple classifiers from a limited amount of labelled data that fully utilises the data. The MCAT method for the semi-supervised learning uses auxiliary classifiers to improve the accuracy of the initial General classifier. The method utilises a classifier trained on a small amount of labelled data specific for the current person, in addition to the general-knowledge classifier. The two additional meta-classifiers decide whether to trust the General or the Specific classifier on a given instance, and whether the instance should be added into the training set of the General classifier or not.

The MCAT method was compared to two transductive approaches and three inductive semi-supervised approaches. The MCAT method significantly outperforms both the baseline approach by 12.51 percentage points and the MCAT without Meta-B by 6.53 percentage points on average. The MCAT method also significantly outperforms inductive approaches: the self-training by 10.79 percentage, the majority vote by 8.41 percentage points, and the threshold-based MCAT by 4.11 percentage points on average. On average, the initial classifier is improved by 12.66 percentage points.

This method can significantly contribute to further development of the ambient intelligence applications, since many of them rely on recognition of human activity. Accurate recognition of atomic activities can also contribute to more reliable recognition of the complex activities.

Acknowledgments. The research leading to these results was partly supported by the ARRS under the Research Programme P2–0209, partly by the European Community's Framework Programme FP7/2007-2013 under grant agreement N 214986.

References

1. Cvetković, B., Kaluža, B., Luštrek, M., Gams, M.: Semi-supervised Learning for Adaptation of Human Activity Recognition Classifier to the User. In: Workshop on Space, Time and Ambient Intelligence, IJCAI 2011, Barcelona, pp. 24–29 (2011)
2. Zhu, X.: Semi-Supervised Learning Literature Survey. CS Technical Report. University of Wisconsin-Madison (2005)
3. Kuncheva, L.I., Whitaker, C.J., Duin, R.P.W.: Limits on the majority vote accuracy in classifier fusion. Pattern Analysis and Applications 8, 22–31 (2003)
4. Longstaff, B., Reddy, S., Estrin, D.: Improving activity classification for health applications on mobile devices using active and semi-supervised learning. In: 4th International Conference on Pervasive Computing Technologies for Healthcare, pp. 1–7. IEEE Press (2010)
5. Dasgupta, S.: Two faces of active learning. Theoretical Computer Science 412, 1767–1781 (2011)
6. Settles, B.: Active Learning Literature Survey. CS Technical Report. University of Wisconsin–Madison (2009)
7. Bicocchi, N., Mamei, M., Prati, A., Cucchiara, R., Zambonelli, F.: Pervasive Self-Learning with Multi-modal Distributed Sensors. In: 2nd IEEE International Conference on Self-Adaptive and Self-Organizing Systems, pp. 61–66. IEEE Press, Washington (2008)

8. Frinken, V., Bunke, H.: Self-training Strategies for Handwriting Word Recognition. In: Perner, P. (ed.) ICDM 2009. LNCS, vol. 5633, pp. 291–300. Springer, Heidelberg (2009)
9. Guzmán-Cabrera, R., Montes-y-Gómez, M., Rosso, P., Villaseñor-Pineda, L.: A Web-Based Self-training Approach for Authorship Attribution. In: Nordström, B., Ranta, A. (eds.) GoTAL 2008. LNCS (LNAI), vol. 5221, pp. 160–168. Springer, Heidelberg (2008)
10. Feng-Biao, G., Chun-Ting, Z.: ZCURVEV: a new self-training system for recognizing protein-coding genes in viral and phage genomes. BMC Bioinformatics 7, 9 (2006)
11. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: 11th Annual Conference on Computational Learning Theory, pp. 92–100. Morgan Kaufmann Press (1998)
12. Guan, D., Yuan, W., Lee, Y.K., Gavrilov, A., Lee, S.: Activity Recognition Based on Semi-supervised Learning. In: IEEE International Conference on Embedded and Real-Time Computing Systems and Applications, pp. 469–475. IEEE Press, Washington (2007)
13. Zhou, Y., Goldman, S.: Democratic Co-Learning. In:16th IEEE International Conference on Tools with Artificial Intelligence, pp. 594–602. IEEE Press, Washington (2004)
14. Project Confidence, http://www.confidence-eu.org/
15. Dovgan, E., Luštrek, M., Pogorelc, B., Gradišek, A., Burger, H., Gams, M.: Intelligent elderly-care prototype for fall and disease detection from sensor data. Zdravniski Vestnik - Slovenian Medical Journal 80(11), 824–831 (2011)
16. Kaluža, B., Mirchevska, V., Dovgan, E., Luštrek, M., Gams, M.: An Agent-Based Approach to Care in Independent Living. In: de Ruyter, B., Wichert, R., Keyson, D.V., Markopoulos, P., Streitz, N., Divitini, M., Georgantas, N., Mana Gomez, A. (eds.) AmI 2010. LNCS, vol. 6439, pp. 177–186. Springer, Heidelberg (2010)
17. Ubisense, http://www.ubisense.net
18. Kaluža, B., Dovgan, E.: Glajenje trajektorij gibanja cloveškega telesa zajetih z radijsko tehnologijo. In: Proceedings of the 12th International Multiconference Information Society, IS 2009, Ljubljana, Slovenia, vol. A, pp. 97–100 (2009)
19. Luštrek, M., Kaluža, B.: Fall detection and activity recognition with machine learning. Informatica 33, 197–204 (2009)
20. Yanhua, C., Manjeet, R., Ming, D., Jing, H.: Non-negative matrix factorization for semi-supervised data clustering. Knowledge and Information Systems 17, 355–379 (2008)
21. UC Irvine Machine Learning Repository, http://archive.ics.uci.edu/ml/
22. Platt, J.C.: Fast training of support vector machines using sequential minimal optimization. MIT Press, Cambridge (1999)
23. Ross, Q.J.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco (1992)
24. Hall, M., Frank, E., Holems, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA Data Mining Software: An Update. SIGKDD Explorations 11, 10–18 (2009)
25. Breiman, L.: Random Forests. Machine Learning 45, 5–32 (2001)
26. Gjoreski, H.: Adaptive human activity/posture recognition and fall detection using wearable sensors. Master thesis, Jožef Stefan International Postgraduate School (2011)

Sound Environment Analysis in Smart Home

Mohamed A. Sehili[1,3], Benjamin Lecouteux[2], Michel Vacher[2], François Portet[2], Dan Istrate[1], Bernadette Dorizzi[3], and Jérôme Boudy[3]

[1] ESIGETEL, 1 Rue du Port de Valvins, 77210 Avon - France
{mohamed.sehili,dan.istrate}@esigetel.fr
[2] Laboratoire d'Informatique de Grenoble
Grenoble 1/Grenoble INP/CNRS UMR 5217, F-38041 Grenoble - France
{benjamin.lecouteux,francois.portet,michel.vacher}@imag.fr
[3] Telecom SudParis, 9 Rue Charles Fourier, 91000 Évry - France
{bernadette.dorizzi,jerome.boudy}@it-sudparis.eu

Abstract. This study aims at providing audio-based interaction technology that lets the users have full control over their home environment, at detecting distress situations and at easing the social inclusion of the elderly and frail population. The paper presents the sound and speech analysis system evaluated thanks to a corpus of data acquired in a real smart home environment. The 4 steps of analysis are signal detection, speech/sound discrimination, sound classification and speech recognition. The results are presented for each step and globally. The very first experiments show promising results be it for the modules evaluated independently or for the whole system.

Keywords: Smart Home, Sound Analysis, Sound Detection, Sound Recognition, Speech Distant Recognition.

1 Introduction

1.1 General Aspects

Demographic change and aging in developed countries imply challenges for the society to continue to improve the well being of its elderly and frail inhabitants. Since the dramatic evolution of Information and Communication Technologies (ICT), one way to achieve this aim is to promote the development of smart homes. In the health domain, a *health smart home* is a habitation equipped with a set of sensors, actuators, automated devices to provide ambient intelligence for daily living task support, early detection of distress situations, remote monitoring and promotion of safety and well-being [1]. The smart home domain is greatly influenced by the *Ubiquitous Computing* domain. As introduced by Weiser [2], ubiquitous computing refers to the computing technology which disappears into the background, which becomes so seamlessly integrated into our environment that we do use it naturally without noticing it. Among all the interaction and sensing technologies used in smart homes (e.g., infra-red sensors, contact doors, video cameras, RFID tags, etc.), audio processing technology has

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 208–223, 2012.

a great potential to become one of the major interaction modalities. Audio technology has not only reached a stage of maturity but has also many properties that fit the Weiser's vision. It is physically intangible and depending on the number and type of the sensors (omnidirectional microphones) that are used, it does not force the user to be physically at a particular place in order to operate. Moreover, it can provide interaction using natural language so that the user does not have to learn complex computing procedures or jargon. It can also capture sounds of everyday life which makes it even more easy to use and can be used to communicate with the user using synthetic or pre-recorded voice. More generally, voice interfaces can be much more adapted to disabled people and the aging population who have difficulties in moving or seeing than tactile interfaces (e.g., remote control) which require physical and visual interaction. Moreover, audio processing is particularly suited to distress situations. A person, who cannot move after a fall but being conscious, still has the possibility to call for assistance while a remote control may be unreachable. Despite all this, audio technology is still underexploited. Part of this can be attributed to the complexity of setting up this technology in a real environment and to important challenges that still need to be overcome [3].

1.2 Related Work

Compared to other modalities (e.g., video cameras, RFID tags), audio technology has received little attention [4,5,6,7,8,9,10]. To the best of our knowledge, the main trends in audio technology in smart home are related to dialog systems (notably for virtual assistant/robot) and emergency and the main applications are augmented human machine interaction (e.g., voice command, conversation) and security (mainly fall detection and distress situation[s] recognition).

Regarding security in the home, audio technology can play a major role in smart homes by helping a person in danger to call for help from anywhere without having to use a touch interface that can be out of reach [4,10]. Another application is the fall detection using a the signal of a wearable microphone which is often fused with other modalities (e.g., accelerometer) [6,5]. In [5] a patient awareness system is proposed to detect body fall and vocal stress in speech expression through the analysis of acoustic and motion data (microphones and accelerometers). However, the person is constrained to wear these sensors all the time. To address this constraint, the dialog system developed by [8] was proposed to replace traditional emergency systems that requires too much change in lifestyle of the elders. However, the prototype had a limited vocabulary (yes/no dialog), was not tested with aged users and there is no mention about how the noise was taken into account. In [9], a communicative avatar was designed to interact with a person in a smart office. In this research, enhanced speech recognition is performed using beam-forming and a geometric area of recording. But this promising research is still to be tested in a multiroom and multisource realistic home.

Most of the speech related research or industrial projects in AAL are actually highly focused on dialog to build communicative agent (e.g., see the EU funded

Companions or CompanionAble projects or the Semvox system [1]). These systems are often composed of Automatic Speech Recognition, Natural Language Understanding, Dialog management, and Speech Synthesis parts supplying the user the ability to communicate with the system in an interactive fashion. However, it is generally the dialog module (management, modeling, architecture, personalization, etc.) that is the main focus of these projects (e.g., see Companions, OwlSpeak or Jaspis). Moreover, this setting is different from the Sweet-Home one as the user must be close to the avatar to speak (i.e., not a distant speech setting). Indeed, in Sweet-Home, the aim is to enable speech interaction from anywhere in multiroom-home. Furthermore, few research projects have considered using daily living sound in their systems though it can be useful information [11,4]. In this perspective, the project addresses the important issues of distant-speech and sound source identification and the outcomes of this research is of high importance to improve the robustness of the systems mentioned above.

1.3 The Sweet-Home Project

Sweet-Home is a project aiming at designing a new smart home system based on audio technology focusing on three main aspects : to provide assistance via *natural man-machine interaction* (voice and tactile command), to ease *social inclusion* and to provide *security reassurance* by detecting situations of distress. If these aims are achieved, then the person will be able to pilot, from anywhere in the house, their environment at any time in the most natural possible way. The targeted users are elderly people who are frail but still autonomous. There are two reasons for this choice. Firstly, a home automation system is costly and would be much more profitable if it is used in a life-long way rather than only when the need for assistance appears. Secondly, in the case of a loss of autonomy, the person would continue to use their own system with some adaptations needed by the new situation (e.g., wheelchair) rather than having to cope simultaneously with their loss of autonomy and a new way of life imposed by the smart home. Qualitative user evaluation studies showed that such speech technology is well accepted by the elderly people [12].

2 Proposed Sound Analysis System

The proposed sound processing system (Figure 1) is composed of the following 4 stages which will be described in the next subsections :

1. the **Sound Detection and Extraction** stage, detecting sound events (daily sounds or speech) from input streams ;
2. the **Sound/Speech Discrimination** stage, discriminating speech from other sounds to extract voice commands ;
3. the **Sound Classification** stage, recognizing daily living sounds ; and
4. the **Speech Recognition** stage, applying speech recognition to events classified as speech.

[1] http://www.semvox.de

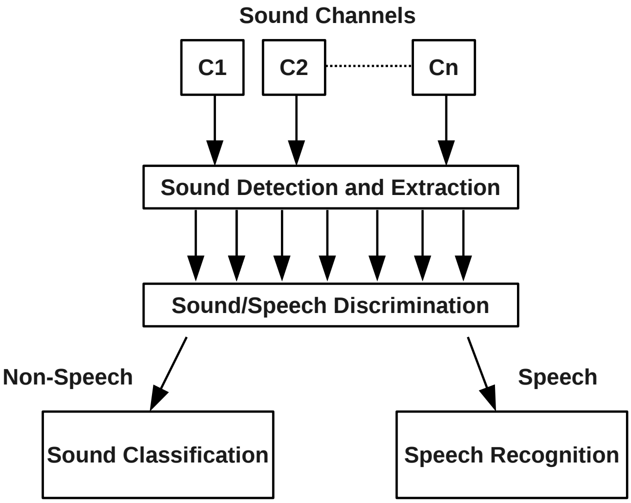


Fig. 1. Sound Analysis System in the SWEET-HOME Project

2.1 Sound Detection and Extraction

To detect intervals of sound occurrence an algorithm based on Discrete Wavelet Transform (DWT) is applied. The algorithm computes the energy of the three high frequencies coefficients of the DWT and an auto-adaptive threshold is estimated from the current Signal to Noise Ratio (SNR). In this work the SNR is computed based on the hypothesis that the noise present within the sound event interval is similar to the noise in the frames directly preceding the current event. The estimation of the SNR is useful both for adapting the threshold and for rejecting events too noisy to be treated. For more details, the reader is referred to [13].

2.2 Sound/Speech Discrimination and Sound Classification

Once intervals of sound occurrences are detected, the most important task is to recognize their category. In everyday life, there is a large number of different sounds and modeling all of them is intractable and not necessary in our case. Indeed, given that the primary aim for the SWEET-HOME project is to enable distant voice command and distress situation[s] detection, speech is the most important sound class. However, other kinds of sound are of interest such as crying and screaming which might reveal that the person needs urgent help. Moreover, some environmental sounds such as glass breaking and water flowing for a long while could also indicate a situation that requires external intervention.

These facts motivated us to build a hierarchical sound recognition system. First, speech is differentiated from all other sounds using a speech-against-sound model rather than including speech as one of the classes of interest. The two-class classification strategy is generally more reliable than numerous-class classification schemes. Then a multi-class sound classification is performed in order to recognize non-speech sounds.

We use the same method for sound/speech discrimination and sound classification. It is a combination of two well-known methods, GMM (Gaussian Mixture Models) and SVM (Support Vector Machines), and it belongs to the so called sequence discriminant kernels [14] [15]. Sequence discriminant kernels for SVM were successfully used for speaker recognition and verification and they became a standard tool [16]. Their main advantage is their ability to deal with sequences of observations of variable length. A sequence of vectors is classified as a whole without performing frame-level classification as with GMM. This makes them quite suitable for sound classification as sound duration may greatly vary. The kernel used in this work is called the GMM Supervector Linear Kernel (GSL) [17] [18]. The following subsections explain this kernel in details.

Support Vector Machines. Support Vector Machines [19] [20] is a discriminative method often used for classification problems, especially when dealing with non-linear data spaces. The non-linear classification is achieved thanks to kernel functions :

$$f(\mathbf{x}) = \sum_{i=1}^{N_{sv}} \alpha_i y_i K(\mathbf{x}, \mathbf{x_i}) + b \tag{1}$$

where $\mathbf{x_i}$ are the support vectors chosen from training data via an optimization process and $y_i \in \{-1, +1\}$ are respectively their associated labels. $K(.,.)$ is the kernel function and must fulfill some conditions so that : $K(\mathbf{x}, \mathbf{y}) = \mathbf{\Phi}(\mathbf{x})^t \mathbf{\Phi}(\mathbf{y})$ where $\mathbf{\Phi}$ is a mapping from the input space to a possible infinite-dimensional space.

Using SVM at frame-level for sound classification is very time consuming both for training and recognition processes and leads to low performances [21]. Thus, sequence discrimination kernels are used to overcome these problems. In [18] the following general definition of a sequence kernel is proposed :

$$K(X, Y) = \mathbf{\Phi}(X)^t \mathbf{R^{-1}} \mathbf{\Phi}(Y) = \left(\mathbf{R}^{-\frac{1}{2}} \mathbf{m}^X\right)^t \left(\mathbf{R}^{-\frac{1}{2}} \mathbf{m}^Y\right) \tag{2}$$

where $\mathbf{\Phi}(X)$ is the high-dimensional vector resulting from the mapping of sequence X and $\mathbf{R^{-1}}$ is a diagonal normalization matrix. In [18] a comparison between two sequence kernels for speaker verification was made. The GSL kernel showed better performances than the Generalized Linear Discriminant Sequence Kernel (GLDS) and thus we retained it for our sound classification system.

GMM Supervector Linear Kernel. The GSL scheme is depicted in Figure 2. To compute the kernel K as in equation (2), we define as $\mathbf{\Phi}_{\mathrm{GSL}}(X) = \mathbf{m}^X$ the supervector composed of the stacked means from the respective adapted Universal Background Model (UBM) components. For each sequence of vectors X extracted from one sound event, a GMM UBM of diagonal covariance matrices is adapted via a MAP (Maximum *a posteriori*) procedure [22]. The normalization matrix $\mathbf{R^{-1}}$ is defined using the weights and the covariances of the UBM model. For more details about the GSL procedure, the reader is referred to [18].

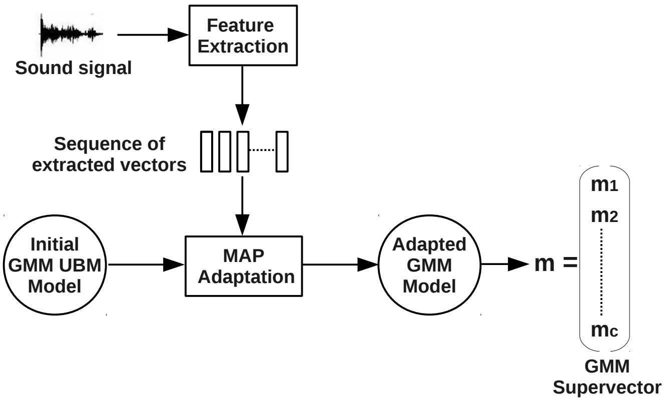


Fig. 2. GMM Supervector mapping process

2.3 Speech Recognition

Automatic Speech Recognition systems (ASR) have reached good performances with close talking microphones (e.g. head-set), but the performance decreases significantly as soon as the microphone is moved away from the mouth of the speaker (e.g., when the microphone is set in the ceiling). This deterioration is due to a broad variety of effects including reverberation and presence of undetermined background noise such as TV, radio and devices. All these problems should be taken into account in the home context and have become hot topics in the speech processing community [23].

In the Sweet-Home project, only vocal orders, large speech or some distress sentences need to be detected. Term detection has been extensively studied in the last decades in the two different contexts of spoken term detection : large speech databases and keyword spotting in continuous speech streams. The first topic recently faced a growing interest, stemming from the critical need of content-based structuring of audio-visual collections. Performances reported in the literature are quite good in clean conditions, especially with broadcast news data. However, performances of state-of-the-art approach are unknown in noisy situation such as the one considered in Sweet-Home. This section summarizes experiments that were run to test to which extend standard and research ASR systems can be used in this context.

The LIA (Laboratoire d'Informatique d'Avignon) speech recognition tool-kit Speeral [24] was chosen as unique ASR system. Speeral relies on an A$\star$ decoder with Hiden Markov Models (HMM) based context-dependent acoustic models and trigram language models. HMMs use three-state left-right models and state tying is achieved by using decision trees. Acoustic vectors are composed of 12 PLP (Perceptual Linear Predictive) coefficients, the energy, and the first and second order derivatives of these 13 parameters. In the study, the acoustic models were trained on about 80 hours of annotated speech. Given the targeted application of Sweet-Home, the computation time should not be a breach of real-time use. Thus, the 1xRT Speeral configuration was used. In this case, the time required by the system to decode one hour of speech signal is real-time (noted 1xRT). The 1xRT system uses a strict pruning scheme. Furthermore, acoustic models were adapted for each of the 21 speakers by using the Maximum Likelihood Linear Regression (MLLR) and the annotated Training Phase of the corpus. MLLR adaptation is a good compromise while only a small amount of annotated data is available. For the decoding, a 3-gram language model (LM) with a 10K lexicon was used. It results from the interpolation of a generic LM (weight 10%) and a specialized LM (weight 90%). The generic LM was estimated on about 1000M of words from the French newspapers Le Monde and Gigaword. The specialized LM was estimated from the sentences (about 200 words) that the 21 participants had to utter during the experiment.

The Speeral choice was made based on experiments we undertook with several state-of-the-art ASR systems and on the fact that the Driven Decoding Algorithm (DDA) is only implemented in Speeral.

2.4 Driven Decoding Algorithm

DDA aims at aligning and correcting auxiliary transcripts by using a speech recognition engine [25,26]. This algorithm improves system performance dramatically by taking advantage of the availability of the auxiliary transcripts. DDA acts at each new generated assumption of the ASR system. The current ASR assumption is aligned with the auxiliary transcript (from a previous decoding pass). Then a matching score α is computed and integrated with the language model [26].

We propose to use a variant of the DDA where the output of the first microphone is used to drive the output of the second one. This approach presents two main benefits : the second ASR system speed is boosted by the approximated transcript and DDA merges truly and easily the information from the two streams while voting strategies (such as ROVER) do not merge ASR systems outputs ; the applied strategy is dynamic and used, for each utterance to decode, the best SNR channel for the first pass and the second best channel for the last pass. A nice feature of DDA is that it is not impacted by an asynchronous signal segmentation since it works at the word level.

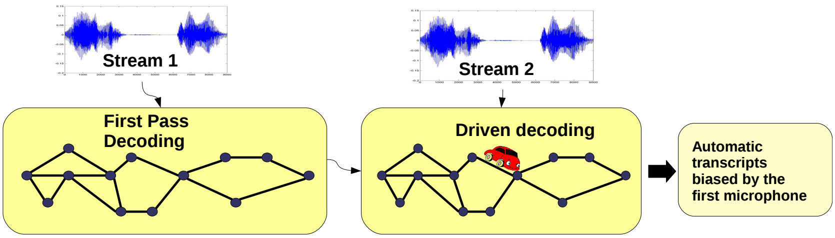


Fig. 3. DDA used with two streams : the first stream drives the second stream

The proposed approach benefits from the multiple microphones of the smart home and from *a priori* knowledge about the sentences being uttered. This approach is based on the DDA which drives an audio stream being decoded by the results of the decoding on another one [10]. The first stream (channel with the best Signal to Noise Ratio) is used to drive the second stream and to improve the decoding performances by taking into account 2 simultaneous channels (Figure 3). An important aspect to mention is that the purpose of the experiment is to assess the impact of the automatic segmentation. Unlike our previous experiments [10] we do not use grammar in order to bias strongly the ASR system.

3 Multimodal Data Corpus

To provide data to test and train the different processing stages of the Sweet-Home system, experiments were run in the Domus smart home that was designed by the Multicom team of the laboratory of Informatics of Grenoble to observe users' activities interacting with the ambient intelligence of the environment. Figure 5 shows the details of the flat. It is a thirty square meters suite flat including a bathroom, a kitchen, a bedroom and a study, all equipped with sensors and effectors so that it is possible to act on the sensory ambiance, depending on the context and the user's habits. The flat is fully usable and can accommodate a dweller for several days. The technical architecture of Domus is based on the KNX system (KoNneX), a worldwide ISO standard (ISO/IEC 14543) for home and building control. A residential gateway architecture has been designed, supported by a virtual KNX layer seen as an OSGI service (Open Services Gateway Initiative) to guarantee the interoperability of the data coming and to allow the communication with virtual applications, such as activity tracking.

The following sensors were used for multimodal data acquisition :

- 7 radio microphones set into the ceiling that can be recorded in real-time thanks to a dedicated PC embedding an 8-channel input audio card ;
- 3 contact sensors on the furniture doors (cupboards in the kitchen, fridge and bathroom cabinet) ;
- 4 contact sensors on the 4 indoor doors ;
- 4 contact sensors on the 6 windows (open/close) ;
- 2 Presence Infrared Detectors (PID) set on the ceiling.

The multimodal corpus was acquired with 21 persons (including 7 women) acting in the Domus smart home. To make sure that the data acquired would be as close as possible to real daily living data, the participants were asked to perform several daily living activities in the smart home. The average age of the participants was 38.5 ± 13 years (22-63, min-max). The experiment consisted in following a scenario of activities without condition on the time spent and the manner of achieving them (e.g., having a breakfast, simulate a shower, get some sleep, clean up the flat using the vacuum, etc.). Figure 4 shows participants performing activities in the different rooms of the smart home. A visit, before the experiment, was organized to make sure that the participants will find all the

Fig. 4. Images captured by the DOMUS video cameras during the experiment

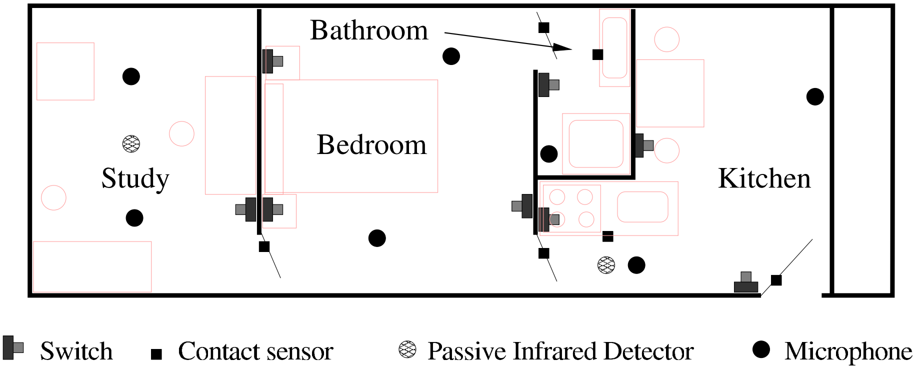


Fig. 5. Layout of the DOMUS Smart Home and position of the sensors

items necessary to perform the activities. During the experiment, event traces from the home automation network, audio and video sensors were captured. Video data were only captured for manual marking up and are not intended to be used for automatic processing. In total, more than 26 hours of data have been acquired.

For the experiment described in this article, we used only the streaming records of the 7 microphones (the remaining data are used for others research work). These records contain the daily living sounds resulting from routine activities during the performance of the scenario as well as two telephone conversations (20 short sentences each in French : "Allô oui", "C'est moi", "J'ai mal à la tête" . . .).

The human annotation of the audio corpus was a very fastidious task given the high number of sound events generated by the participants. To speed up the process, a detection algorithm was applied to the seven channels to detect intervals of sounds of interest. Then, for human annotation purpose, a unique signal resulting of the combination of the seven channels using a weighted sum with coefficients varying with the signals energy was created. Moreover, sound intervals were fused by making the union of overlapping intervals of the seven channels. This signal, the merged intervals, and the videos were then used by the authors to annotate the sound events. The resulting annotation file contains the start, the end and the class of the sound or speech events.

Table 1. Detection sensitivity for different values of the overlapping criteria τ

τ (%)	20	50	80	100
Sensitivity (%)	96.1	93.4	90.3	88.6

4 Experimental Results

To assess the performance of the method, audio records from five participants, S01, S03, S07, S09 and S13, were extracted from the multimodal corpus for Sound/Speech discrimination. They last respectively 88, 50, 72, 63 and 60 minutes. This small amount of data is due to delay in the annotation process. Furthermore, for sound recognition, S13 was not used due to incomplete annotation. For each subject, the 7 channels were used for detection and classification. In this section, we present the results for the four stages of the system namely : Sound Detection, Sound/Speech Discrimination, Sound Classification and ASR.

4.1 Sound Detection and Extraction

Evaluating the performance in detecting temporal intervals of event is known to be a hard task as several strategies can be employed (discrepancy in start/end occurrence time, duration difference, intersection, etc.). In our case, we used the temporal intersection between the humanly annotated reference intervals and the automatically detected ones. In this strategy, a reference sound event is correctly detected if there is at least τ% overlap between a detected interval and the reference interval. An evaluation of the detection results was made on each channel and a reference interval was considered detected if at least one detection was correct on one of the 7 channels. For τ we tested values between 20% and 100% in order to measure the decrease of sensitivity (or recall). The average results for four person[s] are presented in Table 1. The decrease of sensitivity between $\tau = 20\%$ and $\tau = 50\%$ is less than 3%. In reality if the detected segment contains half of the useful signal, it is sufficient for the sound classification system. This is why the 50% value for τ was chosen.

The detection algorithm was applied to each of the 7 channels in order to be able to make information fusion in the recognition stages. The evaluations were performed with 4 different records of the corpus. The results of the detection ratio are presented in Table 2. The detection sensitivity is very stable across the participants. Moreover, the best detection is also stable over the channel. Indeed, channel 1 and channel 2 gave the best detection sensitivity and exhibited the highest SNR. This is due to the scenario of the corpus which led the participant to be mostly close to these microphones when performing noisy activities.

4.2 Sound/Speech Discrimination

For Sound/Speech discrimination, an UBM model of 1024 components was created using 16 MFCC (Mel Frequency Cepstral Coefficients [27]) feature vectors

Table 2. Detection Sensitivity for four participants with $\tau = 50\%$

Participant	S01	S03	S07	S09
Detection Sensitivity (%)	93.2	93.1	93.0	94.4
Average Sensitivity (%)	93.4			

extracted from 16ms signal frames with an overlap of 8ms. The UBM model was learned from speech data from three participants different from those used for the evaluation. The utterances made by the five participants mentioned above were used to adapt the UBM model and generate the supervectors for the SVM stage.

Sound/Speech discrimination was performed on each of the 7 channels. Channels 6 and channel 7 gave the best results, as the subjects were close to them whilst talking. Table 3 shows Sound/Speech discrimination results. [The] number of False Negatives (missed detection or classification of speech) and True Positives (correct classification of speech) are given for detection and Sound/Speech discrimination. The missed utterances were either caused by the detection step (utterance not detected), or by the discrimination step (utterance detected but not recognized as speech). It should be noted that these results do not include false positive recognitions (non-speech sounds recognized as speech, which were actually rare), nor do they take into account speech from radio or improvised phrases that are not be used for speech recognition.

Table 3. Sound/Speech discrimination performances

Participant	# of Utt.	Channel	False Neg. Det.	False Neg. Reco.	True Positive
S01	44	C6	0	4	40
		C7	1	2	41
S03	41	C6	0	2	39
		C7	1	7	33
S07	45	C6	4	5	36
		C7	6	3	36
S09	40	C6	0	0	40
		C7	0	0	40
S13	42	C6	0	4	38
		C7	1	0	41

4.3 Sound Classification

Among 30 annotated sound classes, 18 classes were retained for our experiment (Brushing Teeth, Coughing, Hands Clapping, Door Clapping, Door Opening, Electric Door Lock, Window Shutters, Curtains, Window Shutters + Curtains, Vacuum Cleaner, Phone Ring, Music Radio, Speech Radio, Speech + Music Radio, Paper, Keys, Kitchenware and Water). Examples of classes not used include typing on a keyboard, eating biscuits, manipulating a plastic bag etc.

Although it would have been better to use sounds related to abnormal situations such as human screams or glass breaking, the participants were not asked to perform these kinds of activity. Indeed, the corpus acquisition was performed mainly to test the daily living usage rather than distress situations which are very difficult to produce. Many sounds were either considered as noise or very hard to recognize even by human ears and were annotated as *unknown*.

Given that detection and classification were separately applied to each channel, this led us to a problem of synchronization. Indeed, temporal intervals from a same event recorded on several channels may not be recognized as the same sound class. Our multi-channel aggregation strategy was the following : if an automatic detection does not cover at least 50% of an annotation, then its recognition result will not be taken into account for the classification. To take the final recognition decision using several channels, the sound event with the best SNR is compared to the annotation to compute the classification score.

Table 4 shows the results obtained for each participant. Sounds of Interest (SOI) are the acoustic events in the annotated recordings that belong to the set of the 18 sound classes mentioned above. The measure used to evaluate the performance was the ratio of the number of well recognized detections to the number of detected Sounds of Interest. Results are presented for the detection (column "TP D") and the classification (column "TP C", that is the ratio of well classified SOI taking into account only the number of well detected SOI) as well as for both (column "TP D+C", the ratio of well classified SOI on the total number of SOI). In order to evaluate how far the method is from the optimal performance, the Oracle values are given. The Oracle performances were computed by considering that a SOI is well recognized if at least one of the channels is correct regardless of its SNR. Table 5 shows the average performance per channel for the four subjects. In this experiment, we are only interested in true positives. In other words, for the time being, the system does not include any rejection for unknown sounds. This will be implemented in future work via the use of thresholds, or the creation of one large class for unknown sounds.

Table 4. Sound classification performance using the multi-channel fusion strategy

Subject	# SOI Occur.	TP D(%)	TP C(%)	TP D+C(%)	Oracle TP D+C(%)
S01	230	83.9	69.4	58.2	58.2
S03	175	80.6	65.2	52.6	52.6
S07	245	82.4	68.8	56.7	56.7
S09	268	91.8	74.8	68.7	68.7

Table 5. Average performance per channel for all subjects

Channel	C1	C2	C3	C4	C5	C6	C7	Fusion of all Channels
Avg. TP D+C(%)	31.3	33.3	14.9	24.1	21.8	13.3	9.6	59.1

4.4 Automatic Speech Recognition

The ASR stage was assessed using the classical Word Error Rate (WER) metric : $\frac{insertion+deletion+substitution}{Number of reference words}$ (WER is above 100 if there is more word insertion than reference words). The errors of automatic segmentation (WER ranges between 35.4% to 140%) results in two types of degradation : 1) The insertion of false positive speech detection ; and 2) The deletion of speech that was missed by the detection/discrimination.

In our experiments, the type of degradation differs for each speaker. In order to show the impact of insertions we present the "Global WER" and the "speaker WER". The first one takes into account all the insertions computed out of the reference segmentation while the second one compare the WER only on well detected segments. In all the experiments, the insertions generate a lot of errors and there is constantly a degradation of the "speaker WER" (non detected speech). Nevertheless DDA allows one to improve the WER by taking advantage of the combination of two segmented streams.

Table 6 shows experiments for the ASR system by using the automatic segmentation. We present three baselines ("ref", "ref-DDA" and "mix"). The baselines "ref" and "ref-DDA" are focused on the distant speech recognition issue. On the two "ref" baselines the ASR system is launched on the reference segmentation. The purpose of these baselines is to highlight the encountered difficulties of the ASR system in distant conditions :

- The "ref" baseline is a classical ASR system running on the best SNR channel.
- The "ref-DDA" baseline uses DDA in order to combine the two best SNR channels : this method allows one to improve the ASR robustness.

By using the reference segmentation, the WER ranges between 17.5% and 36.3%. This error rate is mostly due to the distance and the noisy environment. However these results are quite usable for the detection of distress or home automation orders [10] : our previous work using a grammar (with DDA) in order to constrain the ASR system allows to use these high WER rates. Moreover DDA improves by 5% relative of the WER.

In a second time we present the results on the whole system : speech detection, speech segmentation and speech transcription :

- "mix" is a classical ASR system running on the automatic segmentation (and the mix of all channels).
- "DDA" corresponds to the ASR system used in real conditions : DDA acts on the automatic segmentation by combining the two best SNR channels.

The most important WER to consider is related to the speaker : the decision system will be able to filter out false detections. In the presented experiments this WER ranges between 23% and 54%. Our previous work has shown that despite imperfect recognition the system is still able to detect a high number of original utterances that were actually home automation orders and distress sentences. This is due to the restricted language model coupled with the DDA

Table 6. Speech recognition results for each speaker

	Correct Words (%)	Global WER (%)	speaker WER (%)
S01-ref	84.4	17.5	17.5
S01-ref-DDA	84.0	16.0	16.0
S01-mix	77.4	36.8	24.1
S01-DDA	78.7	35.4	23.1
S03-ref	87.3	20.1	20.1
S03-ref-DDA	86.0	20.9	20.9
S03-mix	32.3	105.8	75.7
S03-DDA	56.0	74.1	51.9
S07-ref	69.6	36.3	36.3
S07-ref-DDA	69.7	36.2	36.2
S07-mix	47.9	81.8	56.5
S07-DDA	50.2	81.8	54.0
S09-ref	85.1	15.4	15.4
S09-ref-DDA	85.3	15.3	15.3
S09-mix	69.1	57.4	32.4
S09-DDA	69.1	54.3	31.9
S13-ref	75.3	27.4	27.4
S13-ref-DDA	76.4	26.3	26.3
S13-mix	44.7	140.5	57.4
S13-DDA	47.9	120.0	54.5

strategy which permits to retrieve correct sentences that are of lowest likelihood than the first hypothesis among the set of hypotheses [28].

5 Discussion and Conclusion

In this paper we propose a complete system for sound analysis in smart home. The system is designed hierarchically and can deal with multiple channels. This multi-layer design makes it easy to test the performance of each module separately. The sound detection module detects most sounds events in the house thanks to the use of 7 channels. Sound/Speech discrimination gave good result for the two channels with highest SNR as the subject does not move around the house whilst talking. Sound classification was more challenging because of the greater number of sound classes and the fact that the "best" channel(s) varies all the the time and some kind of sound events can occur anywhere in the house. Multichannel fusion strategy based on SNR gave very encouraging results (Table 4 and 5), taking into account that the sound detection's performance also affects that of sound classification. However, the current system does not include any rejection. We intend to implement this feature in future work. Furthermore, as many daily sounds vary considerably in terms of representation, it would be very desirable to be able to use simpler features for sound classification, at least for some sounds. This will also be investigated in further work.

As for the ASR system, the proposed approach based on DDA lead to moderate robustness. The impact of automatic segmentation for the ASR system highlights new challenges for integration within a smart home ; each stage spreading its errors. Despite the occasionally low performance, the ASR system offers the possibility to be exploited in industrial context. Our future work will focus on the cooperation with the decision-making module.

References

1. Chan, M., Campo, E., Estève, D., Fourniols, J.Y. : Smart homes — current features and future perspectives. Maturitas 64(2), 90–97 (2009)
2. Weiser, M. : The computer for the 21st century. Scientific American 265(3), 66–75 (1991)
3. Vacher, M., Portet, F., Fleury, A., Noury, N. : Development of audio sensing technology for ambient assisted living : Applications and challenges. International Journal of E-Health and Medical Communications 2(1), 35–54 (2011)
4. Istrate, D., Vacher, M., Serignat, J.F. : Embedded implementation of distress situation identification through sound analysis. The Journal on Information Technology in Healthcare 6, 204–211 (2008)
5. Charalampos, D., Maglogiannis, I. : Enabling human status awareness in assistive environments based on advanced sound and motion data classification. In : Proceedings of the 1st International Conference on Pervasive Technologies Related to Assistive Environments, pp. 1 :1–1 :8 (2008)
6. Popescu, M., Li, Y., Skubic, M., Rantz, M. : An acoustic fall detector system that uses sound height information to reduce the false alarm rate. In : Proc. 30th Annual Int. Conference of the IEEE-EMBS 2008, August 20-25, pp. 4628–4631 (2008)
7. Badii, A., Boudy, J. : CompanionAble - integrated cognitive assistive & domotic companion robotic systems for ability & security. In : 1st Congres of the Société Française des Technologies pour l'Autonomie et de Gérontechnologie (SFTAG 2009), Troyes, pp. 18–20 (2009)
8. Hamill, M., Young, V., Boger, J., Mihailidis, A. : Development of an automated speech recognition interface for personal emergency response systems. Journal of Neuro Engineering and Rehabilitation 6 (2009)
9. Filho, G., Moir, T.J. : From science fiction to science fact : a smart-house interface using speech technology and a photo-realistic avatar. International Journal of Computer Applications in Technology 39(8), 32–39 (2010)
10. Lecouteux, B., Vacher, M., Portet, F. : Distant Speech Recognition in a Smart Home : Comparison of Several Multisource ASRs in Realistic Conditions. In : Interspeech 2011, Florence, Italy, p. 4 (August 2011)
11. Chen, J., Kam, A.H., Zhang, J., Liu, N., Shue, L. : Bathroom Activity Monitoring Based on Sound. In : Gellersen, H.-W., Want, R., Schmidt, A. (eds.) PERVASIVE 2005. LNCS, vol. 3468, pp. 47–61. Springer, Heidelberg (2005)
12. Portet, F., Vacher, M., Golanski, C., Roux, C., Meillon, B. : Design and evaluation of a smart home voice interface for the elderly – acceptability and objection aspects. Personal and Ubiquitous Computing (in press)
13. Rougui, J., Istrate, D., Souidene, W. : Audio sound event identification for distress situations and context awareness. In : Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2009, Minneapolis, USA, pp. 3501–3504 (2009)

14. Jaakkola, T., Haussler, D. : Exploiting generative models in discriminative classifiers. In : Advances in Neural Information Processing Systems, vol. 11, pp. 487–493. MIT Press (1998)
15. Temko, A., Monte, E., Nadeu, C. : Comparison of sequence discriminant support vector machines for acoustic event classification. In : Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (2005)
16. Wan, V., Renals, S. : Speaker verification using sequence discriminant support vector machines. IEEE Transactions on Speech and Audio Processing, 203–210 (2005)
17. Campbell, W.M., Sturim, D.E., Reynolds, D.A., Solomonoff, A. : SVM based speaker verification using a gmm supervector kernel and nap variability compensation. In : Proceedings of ICASSP 2006, pp. 97–100 (2006)
18. Fauve, B., Matrouf, D., Scheffer, N., Bonastre, J.F. : State-of-the-art performance in text-independent speaker verification through open-source software. IEEE Transactions on Audio, Speech, and Language Processing 15, 1960–1968 (2007)
19. Burges, C.J.C. : A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Discov., 121–167 (1998)
20. Schölkopf, B., Smola, A.J. : Learning with Kernels. MIT Press (2002)
21. Sehili, M.A., Istrate, D., Boudy, J. : Primary investigations of sound recognition for a domotic application using support vector. Automation, Computers, Electronics and Mechatronics, vol. 7(34(2)), pp. 61–65. Annals of the University of Craiova (2010)
22. Reynolds, D.A., Quatieri, T.F., Dunn, R.B. : Speaker verification using adapted gaussian mixture models. In : Digital Signal Processing 2000 (2000)
23. Wölfel, M., McDonough, J. : Distant Speech Recognition, p. 573. John Wiley and Sons (2009)
24. Linarès, G., Nocéra, P., Massonié, D., Matrouf, D. : The LIA Speech Recognition System : From 10xRT to 1xRT. In : Matoušek, V., Mautner, P. (eds.) TSD 2007. LNCS (LNAI), vol. 4629, pp. 302–308. Springer, Heidelberg (2007)
25. Lecouteux, B., Linarès, G., Estève, Y., Gravier, G. : Generalized driven decoding for speech recognition system combination. In : Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008, pp. 1549–1552 (2008)
26. Lecouteux, B., Linarès, G., Bonastre, J., Nocéra, P. : Imperfect transcript driven speech recognition. In : InterSpeech 2006, pp. 1626–1629 (2006)
27. Logan, B. : Mel frequency cepstral coefficients for music modeling. In : Proceedings of International Symposium on Music Information Retrieval (2000)
28. Vacher, M., Lecouteux, B., Portet, F. : Recognition of Voice Commands by Multisource ASR and Noise Cancellation in a Smart Home Environment. In : EUSIPCO, Bucarest, Romania, pp. 1663–1667 (August 2012)

Contextual Wizard of Oz

A Framework Combining Contextual Rapid Prototyping and the Wizard of Oz Method

Doris Zachhuber[1], Thomas Grill[1], Ondrej Polacek[1], and Manfred Tscheligi[1]

ICT&S Center, University of Salzburg
Salzburg, Austria
{firstname.secondname}@sbg.ac.at

Abstract. Exploring user interaction in specific contexts is often based on simulated environments and semi-functional prototypes of interactive systems. In this paper, we address a combination of context simulation with the Wizard of Oz (WOz) technique, where a hidden human "wizard" simulates missing functionalities and system intelligence. The goal of our work is to provide a software framework for fast prototyping and concurrent evaluation through user studies during iterative interaction design processes. Contextual interaction research is particularly challenging in high-dynamic interaction contexts like ambient environments and includes the simulation of various context parameters to elaborate interaction designs in the target context. For this purpose, we have developed a prototyping framework that allows the setup and handling of different contextual situations during user studies. The framework and the proposed WOz protocol, which is used for integrating the WOz technique, are highly flexible, modular and adjustable at runtime. This allows their application in a big variety of study contexts in in-situ and in-vitro settings. A detailed description of the framework's requirements and architecture as well as a user study, where we successfully applied our framework, are presented. First results have been collected through interviews with evaluators and developers who used the framework to develop the particular study setups. The identified improvements and potentials experienced during the usage of the framework have been analyzed and provide valuable findings for further iterations.

1 Introduction

In the past years Human-Computer Interaction (HCI) has moved beyond the desktop accelerating the trend towards ambient intelligent environments. Novel forms of interaction in various contexts have been explored and user studies have shown that interaction design benefits from explicit studies in the target context. Results support theories and models motivating context-oriented thinking, such as the situated action theory suggesting that people's behavior in particular situations is determined by the context [15].

Interfaces in ambient intelligent environments need to be designed according to the needs and behavior of people in very specific contextual situations. This re-

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 224–239, 2012.

quires an understanding of the particular context and its influencing components on the user and user interaction. Thus, contextual interaction research comprises analysis, evaluation and comparison ranging from user behavior in specific contexts to context-aware interfaces that adapt themselves in real-time to optimize user support and interaction. Contextual situations have multiple aspects and change rapidly. In consequence, it is highly challenging and time-consuming to consider all possible situations and corresponding reactions. The goal of our work is to make interaction design for specific contexts easier by enabling fast prototyping of flexible user study setups. We have developed a software framework combining sensor/actuator- and software-based context simulation enabling situated interaction, and the Wizard of Oz (WOz) technique to cover unexpected or rare situations and simulate system behavior in a context which is not fully implemented yet.

The WOz technique is a rapid-prototyping method where a hidden human "wizard" simulates a system's intelligence [9]. According to Dow et al. [4] it is particularly useful in exploring user interfaces for pervasive, ubiquitous, or mixed-reality systems that combine complex sensing and intelligent control logic. The WOz method helps designers to explore and evaluate designs and identify user preferences before investigating a considerable development time for a fully functional prototype. Especially in complex ubiquitous environments the system can implement basic functionality, while a mobile WOz interface (i) enables the selection and triggering of simulated system behavior and (ii) supports the cognitively expensive task of a wizard. [12]

We have developed a service-oriented context simulation framework that enhances rapid prototyping approaches by combining them with the WOz technique. Because of its configuration flexibility and extensibility the framework is of high value especially for contextual situations where real-world deployment is not feasible or at least difficult due to safety reasons, privacy concerns or limited technological possibilities in the study environment, e.g. in the factory or automotive domain. Simulating specific task contexts thus refers to a number of parameters combined during a simulation in order to appropriately replicate specific contextual situations. The framework also allows to measure, identify, and analyze relevant parameters in detail in a controllable and replicable setting.

2 Related Work

WOz is an experimental evaluation technique. The user is observed when operating an apparently fully functional system whose missing features are supplemented by a hidden "wizard" [12]. This approach results in a higher degree of freedom regarding user as well as system behavior, independent from available hardware and algorithms. While in early years WOz studies focused on the evaluation of futuristic but not yet implementable interface designs, later experiments concentrate on human behavior instead. HCI researchers try to go beyond technological constraints and towards a user-centered design for intuitive interfaces.

The WOz technique has its origins in natural language interface (NLI) [3,13] and intelligent agent design [9]. In the work by Lee et al. [7] computer vision based hand-tracking is combined with simulated speech input by the wizard to gain insights about user behavior and preferences when interacting with an augmented reality application with multimodal input. According to Salber and Coutaz[12] another possibility to deal with concurrent input modalities is a two-wizard configuration to lower the cognitive load for each wizard. However, this requires a higher number of people for conducting studies and their synchronization. Instead, in our work we aim for reduced task complexity for the wizard by computer support such as recognizing, analyzing and abstracting user behavior and contextual situations, and enabling to keep track of current states and available options. One of advantages of our computer-supported WOz approach is the possibility to easily combine it with the development process of contextual interactive applications itself. Other advantage is the fact that the quality of study results is shifted from the reliability of technology to the consistency of the wizard only as described by [12]. Liu and Khooshabeh [8] state that especially at the scale of applications in ambient intelligent environments, increased system fidelity and automation allow to run user studies more efficiently, and can improve the validity of study results.

In order to keep the WOz technique effective, designers have to bridge the gap between the wizard's role and actual system implementation. According to Dow et al. [4] this is one reason why WOz is used rarely. A WOz interface has to support fast and easy integration into the prototype programming environment and be reusable to enable rapid prototyping. Software extensions such as DART (Designer's Augmented Reality Toolkit), a WOz plug-in for Adobe Director, or automatic WOz interface generation are possible solutions and provide methods for programmers to integrate Wizard of Oz functionality in their prototype [4]. A WOz tool, that has been used to simulate two mobile multimodal scenarios, is described in work by Ardito et al. [1].

Context simulation can be covered by different hardware actuators and/or software applications. Examples for such are e.g. Ubiwise, which is a rapid prototyping toolkit for UbiComp from [2]. It allows users to explore a 3D environment model in first-person view and interact using mouse and keyboard. RePlay from [10] contributes to research on design tools by enabling capturing and an adjustable playback of contextual data such as location sensor traces. The Sketchify tool from [11] enhances freehand user interface sketching with interactive hard- and software to increase user experience.

In our framework we support and extend existing context simulation and WOz approaches. Our framework supports in-situ and in-vitro evaluation environments by defining and implementing a method that allows to control contextual data during a study. Evaluators are supported already during the setup and prototyping phase of a study with tools like the study editor that allows to prototype and configure contextual scenarios. We support evaluators during the study with study controls like setting participants id's, logging functionality, and a mobile wizard tool that allows to control and monitor contextual and prototype states. Developers are provided with a distributed, service oriented framework that allows to combine existing functionality and add additional functionality by simply loading developed services to the framework and wiring them with functionality already available. If a system is also meant to react on explicit/implicit user interaction, the system has to be aware of the user's context in the sense of e.g. a user's location or gestures. This bridge to the context information is typically implemented by *context frameworks* using various sensors, advanced tracking systems or computer vision approaches plus classification and reasoning algorithms.

Based on an analysis about applying the WOz technique we could identify a high potential for applying WOz tools in the area of contextual interaction research especially in ambient intelligent environments. The key criteria in order to successfully use the technique is a close integration of the tools supporting WOz into a context simulation framework that is enabling concurrent prototype development. In comparison to existing work we address this in our framework which focuses on a tight integration of both aspects with the overall goal to support the developer and usability engineer with a framework to design contextual interfaces.

3 Contextual Interaction Framework (CIF)

For studying contextual interaction design we developed a framework that allows to combine contextual interaction in a simulated context by applying the WOz technique. By context simulation we refer to simulating specific situations by setting and changing different context parameters intentionally. Examples are the control of light intensities to simulate day and night, some speakers' output for ambient noise, or projectors visualizing surrounding people. In general, all human senses can be addressed, including not only the visual and auditive, but also e.g. the tactile and olfactory sense.

3.1 Requirement Analysis

The requirements for the CIF were identified in requirements engineering sessions that included nine researchers responsible for defining and conducting research projects that include contextual simulation. Their background was partly in computer science, psychology, communication science, and HCI. The main goal of the CIF is based on providing the engineer with support for fast prototyping for user studies, while developing actual interactive applications concurrently, i.e. without much additional implementation effort. This requires a modular programming approach such as service-oriented computing, which is is well-known in context-aware application development in ambient intelligent environments. In smart homes or vehicles, for instance, context recognition and reasoning are hot topics for device control, e.g. for energy saving or safety reasons. Our framework focuses on the control of hard- and software components for context simulation. It also supports the integration of user behavior recognition and analysis. Based on the requirements engineering sessions we elaborated the following requirements that such a system needs to fulfill.

- *Extensibility.* Easy implementation and integration of new sensors, actuators and software components such as context reasoners and formal modeling (using e.g. ontologies, OWL from – see [5]) or third party applications simulating context (e.g. a game engine) has to be ensured in order to enable the reuse of developed prototypes (at least parts of it) during the whole iterative design process.
- *Configuration Flexibility.* Available framework components must be selectable and configurable fast and easy. In the best case configuration changes should be possible even during runtime and via a graphical user interface.
- *Communication Protocol Flexibility.* The communication between different hard- and software components and the WOz interface has to be flexible. It has to be able to cover a high through-put of commands and data to allow contextual simulations in real-time.
- *Performance and Reliability.* The performance of all framework components should allow an approximation to a real-time behavior of the system to ensure that the simulation including actions invoked by the wizard appear to be "real" features of the system. Failures during a study or data ambiguity for multi-modal interaction scenarios have to be handled appropriately.
- *Contextual Information Abstraction.* During WOz studies the task complexity should be minimized for the wizard, e.g. by providing high-level context interpretations or encapsulating elementary actions into higher abstractions such as commands that change multiple context parameters at once.
- *Data Analysis Support.* For evaluation purposes the framework has to ensure that all relevant data and actions can be logged. Playback and annotation possibilities simplify the development of implications.

3.2 Framework Architecture

The CIF as depicted in Figure 1 is based on a service-oriented architecture (SOA). The framework consists of components to support modular programming, scenario configuration functionalities that allow to setup complex study setup, context simulation possibilities, as well as a WOz module to allow controlling the configured setup.

Modular Programming. The core communication and data exchange between all framework components has to be fast and flexible. Implementing own data protocols for every new component must be avoided for fast prototyping. Our work is based on the *Open Services Gateway Initiative (OSGi)*[1], a dynamic specification that defines an architecture for modular application development. OSGi container implementations allow to divide an application into multiple modules (called 'bundles') and manage cross-dependencies between them. Modular programing enables the reuse of single bundles for different setups, thus supporting

[1] `http://www.osgi.org`

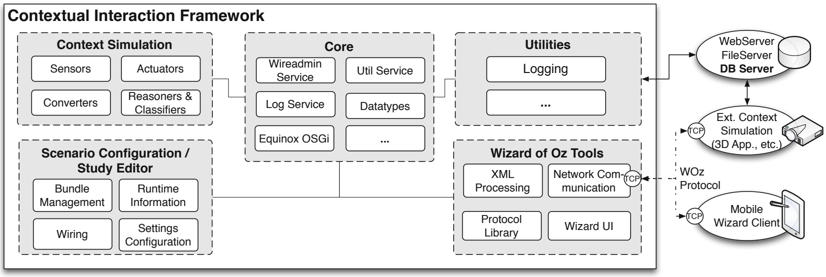


Fig. 1. CIF Architecture: The *OSGi core framework* provides the services to administer bundles and a data flow. *Context Simulation* refers to a set of plugins implementing the interface to contextual information. *The Study Editor* allows to setup, configure, and test contextual setups. *The contextual WOz tools* provide mechanisms to control configured settings during a study.

fast prototyping. Modules can be defined to provide and/or consume services, which are managed by a service registry.

Compared to web services that require a transport layer, OSGi services use direct method invocations, which makes them faster. Another advantage is that there is no need to define a protocol for intra-framework communication via services, instead instances of classes can be exchanged directly. In our framework we offer a bundle holding specified classes for recurrent types such as position or acceleration data used by all other bundles. If parts of the framework cannot be written in Java or already exist as remote stand-alone applications, communication to such software components can be implemented using given tools and interfaces in the CIF, e.g. for the Java Native Interface (JNI) for native applications or basic network communication which is implemented through a TCP Connector bundle. In the CIF all bundles implement the following structure to enable easy navigation and reuse of all modules. The main bundle types are as follows.

Sensors. Sensor bundles are service producers. We distinguish between physical, network, virtual, playback and synthetic sensors, based on the sensor abstractions described in [6]. While physical sensors are devices that measure and gather real-world quantities, virtual ones do not origin from any device, but from an application (e.g. a calendar) or online source (e.g. weather information) instead. Network sensors are basically physical sensors with the difference that our framework does not gather the data from a device directly, but over a network connection. By synthetic sensors we refer to data not coming from any real-world device or application, but an algorithm that produces data randomly (e.g. position data of simulated people surrounding the user). Finally, playback sensors replay sensor data recorded from one of the other four sensor types in advance and thus, enabling the replay on specific contextual situations and events.

Actuators. Actuator bundles are service consumers which control hardware devices or any application specific features. Similar to the sensor bundles, we distinguish between physical, network and software actuators. The latter refers to a software application which allows to control some of its features, e.g. the lighting in a virtual 3D environment, which runs in a separate game engine.

Converters. Converter bundles consume data to process it and provide the results as a new service. Therefore, these bundles often apply different algorithms for classification, feature extraction and interpretation, data fusion and many more. Again, we differ between physical, network and software converters.

Core. The core bundles of the interaction framework contain the basic functionality of the framework. It contains e.g. OSGi services, wiring services that allow to define the data flow, logging services, as well as a set of generic data types being exchanged over wires.

Tools. Tool bundles implement the Study Editor providing a user interface to configure the setup of the study, the bundle management allowing to load, start, and stop bundles, as well as logging functionalities useful to capture and store data during a study. The tools further consist of additional GUI bundles providing functionality useful for setting up and testing a study configuration.

Wizard. The wizard bundles handle the WOz protocol and provide the server for client connections such as the mobile wizard client.

Scenario Configuration

The scenario configuration provides a way that allows to setup a study based on the previously described modular approach. Our framework is based on OSGi, which provides predefined services through Java interfaces. One of these services is the *OSGi Wire Admin Service* which implements the concept of interchanging data between producer and consumer services over a configured wire instead of letting the bundles find services they want to work with by themselves.

In our framework we implemented a graphical user interface, i.e. the *Study Editor* that provides a graphical programming approach for setting up the study based on the OSGi Wire Admin Service. An essential part of the Study Editor are wires that allow to manipulate application elements graphically rather than by specifying them textually. The wires implemented in our framework are lines used to configure and visualize the data flow between bundles. This is done by simply drawing a line between output and input pins (representing services offered/consumed by bundles) from producer, converter and/or consumer bundles (left part of Figure 2). The appropriate service registration in the framework is done automatically in the background. The Study Editor provides functionality to find, start and stop modules, which allows a dynamic reconfiguration of the system. Configurations can be saved and reloaded for future sessions.

Context Simulation and Recognition. Scenarios define which sensors, actuators and converter bundles are used and connected. Sensors are used to capture user behavior and other context parameters which are provided as services (producers). Actuators are used to control hard- or software for context simulation and require input (consumers). Converters consume and produce data. Usually,

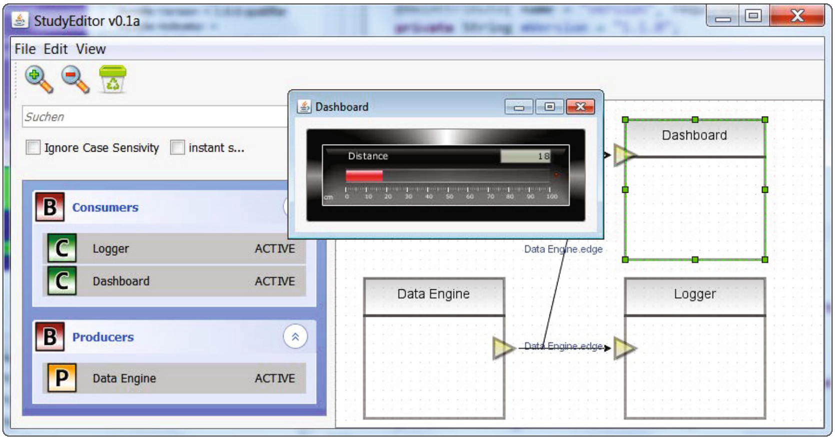


Fig. 2. CIF Study Editor & Bundle Management showing data from a producer linked to a dashboard and a logging module. The study editor provides an interface to interactively setup, define and conduct studies by allowing to combine different tools that are tailored for contextual study needs.

the consumed data is analyzed or supplemented in-between, e.g. a converter may derive high-level context data from low-level sensor input such as the classification of hand gestures from acceleration data.

Wizard of Oz Tools

The WOz tool bundles provide the functionality to integrate the WOz technique into the framework. This includes the ability to receive information from sensors and converters and send data to actuators via consumer and producer services in the framework. A client-server architecture was used while the server component is implemented in the CIF and provides the possibility to connect external clients that allow the monitoring of contextual situations as well as controlling contextual objects configured within the CIF study editor. The Mobile Wizard Client (see Figure 3) implements such features and represents a generic, flexible, and configurable tool that is based on study configurations defined within the CIF. This avoids the necessity reimplementing specific wizard functionalities required by different study setups. The communication between the different components is based on the developed wizard protocol that provides a generic way of enabling communication between the wizard tools but also to external software that can be integrated. By combining both, the wizard server and the mobile wizard tool, the WOz tools can be used for context simulation in multiple ways: control actuators in the framework and/or send commands to external applications such as a 2D/3D visualization of a study environment.

(a) Selection of configured studies

(b) Scenario control provides access to wizardable objects

Fig. 3. Mobile Wizard Client: A mobile application that allows to control the configured study. Study support is provided by the possibility to dynamically adapt the interface for different situations and task scenarios.

4 Framework Evaluation

To show the applicability and validity of the contextual WOz method, we developed the "Operator Guide" study using the CIF. Figure 4 depicts the wiring diagram of the study setup developed using the CIF and the study editor.

The goals behind this were two-fold. We wanted to show the usefulness of the CIF's Wizard of Oz functionality for the purpose of measuring contextual parameters and simulating contextual situations. In order to do this, we developed the "OperatorGuide" study and conducted it to evaluate a workflow support system in a semi-conductor factory. The factory context was simulated in a laboratory. The main task of participants was to move boxes of semiconductors from one machine to another to simulate the workflow pipeline, while we measured the overall throughput in several conditions. Total of 24 participants took part in the evaluation. One session lasted approximately 1.5 hour. The concrete results of the "Operator Guide" study are described in detail in [14].

During the usage of the framework different roles of people participating in the lifecycle of a user study have evolved.

Developers. With developers we refer to people with programming skills that have the task to setup a study. This includes the development of prototypes. Developers are intended to use the CIF in two different ways. They use and reuse existing functionality by setting up a configuration using the wiring metaphor implemented in the framework. They also develop bundles with new functionality to complete the functionality required for a specific study setup. In the case of a WOz study developers also may take the role of a wizard.

Evaluators. By evaluators we refer to people that design and conduct user studies in order to analyze people's behavior or to evaluate a prototype of an interactive

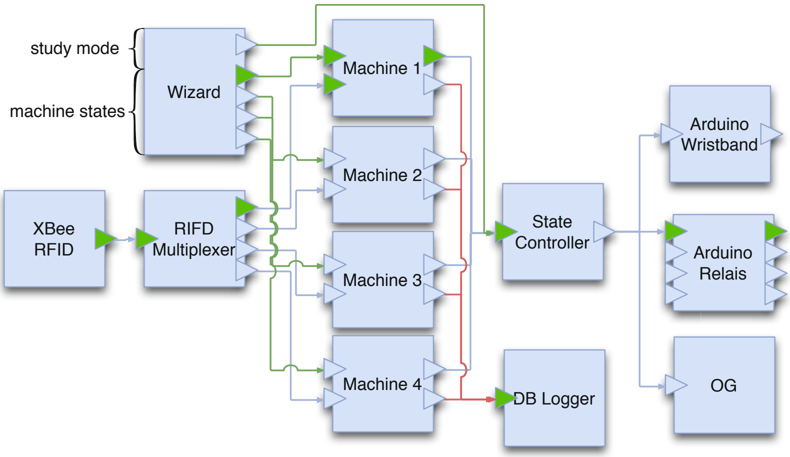


Fig. 4. Wiring diagram of the "Operator Guide" study: The developed setup contains multiple machines that control relays, a vibra-wristband, and the content of a ambient display (OG–Operator Guide). The machines receive input from sensors (RFID) as well as from wizard tools that allow to control the states of the machines.

system. In the case of a WOz study evaluators also may take the role of a wizard.

Based on this categorization we conducted structured interviews with people of both groups. We interviewed 5 developers and 5 evaluators with an experience of each at least 2-3 years in developing and conducting user studies. Both groups had varying experience regarding our focus on contextual user studies. The interviewees differed from the participants of the requirements engineering sessions.

The interviews have been structured into three main parts reflecting different research questions of interest.

The research questions that have been the basis for the interviews are derived from the requirements elaborated for the CIF.

– Does the modular service-oriented system design of the CIF decrease the time of prototype development and does it ease the reuse of previously developed system components?
– Is the graphical programming used in the CIF easy to learn and use and flexible enough to set up the hard- and software components for a study scenario with a focus on simulating different contextual situations?
– Is the WOz methodology an appropriate and convenient way to simulate contextual situations in real time depending on the current user behavior?

The different research questions have been addressed differently in the interviews with respect to the particular user group. For the interviews with the developers we focused on implementation and usage issues regarding the setup of the studies.

The interviews with the evaluators aimed to investigate on research questions targeting the WOz technique.

4.1 Structure of the Interviews

The interviews conducted followed a predefined procedure with the goal to get comparable results. The procedure was adopted for the particular user group and described in the following.

1. *Welcome.* The goal of the interview was explained and a consent form has been given.
2. *Demographic questionnaire.* A demographic questionnaire adopted for the particular user group has been designed and was handed to them.
3. *Hands on Example.* The interviewees were given an example of a study. The example differed between the developers and evaluators. The example was based on the OperatorGuide study setup already described. The developers had the task to use our framework to setup the study according to the concept handed to them. The evaluators were told to use the CIF to load an existing configuration and complete such by adding the wizard functionality. Further they had to test and use the WOz functionality during a test run of the OperatorGuide study where they had to control specific contextual settings like e.g. states of some machines, alarm modes, switching of lights, etc. using the mobile wizard tool.
4. *Structured Interviews.* Guidelines for the structured interviews have been developed based on the previously defined research questions. While the interview guidelines for the developers covered all research questions, the interview guidelines for the evaluators focused on the applicability and use aspects of the contextual WOz method.
5. *Closing.*

4.2 Analysis and Discussion of the Interviews with the Developers

The results regarding the interviews with the developers provide an in-depth insight about the applicability, utility, and usefulness of the framework to setup user studies focusing on the simulation and measurement of contextual information. The interviews have been transcribed, analyzed and clustered. The results are presented in the following.

The Modular Service-Oriented Approach of the System Design. All of the developers agreed that the modular concept of the CIF makes sense and is useful for setting up a study. One of the developers also said that the usefulness of the concept was already clear to him after the presentation. Three of five developers identified the need of clear guidelines covering the procedure of how to setup a study as well as guidelines how to extend it.

Regarding the question whether they have the feeling that the approach may save time when developing a study setup all of them agreed with the restriction that a rather complete library of bundles representing the sensors and actuators already exists. All of the developers agree that it is easy to familiarize oneself with the framework and that they are willing to use the framework if they have to develop a setup of a contextual interaction study. In case they have already invested the time to learn the framework they would stick to the framework and reuse it in further projects if it has shown to work and fit their specific problems regarding the usage of contextual information in prototypes. All developers agreed on the perspective that the framework could grow with every project and implementation of additional bundles to be more complete. Regarding the appropriateness and time-saving there were doubts if it is more appropriate to use other existing frameworks providing functionality to attach sensors. Such frameworks are e.g. *Processing*[2] or *VVVV*[3].

Other issues arising was that a developer or a technician would be needed anyway for setting up the system. The interviewees stated that the current state of the interaction framework does not enable an evaluator to setup a system alone. Also the Study Editor has to be extended. Currently bundle descriptions are missing and a search and sorting functionality for bundles would be "nice" as it is quite hard to find bundles in the Study Editor. Another idea was to include representative images for the different bundles in order to make them easily distinguishable. Developers state that the possibility to include, load, and start bundles during run-time is great and provides a potential not only for prototyping but also debugging the prototypes. All of the developers think that the flexibility of the framework is high enough because of modular structure. Also it is based on OSGi which is currently an industrial standard and shows to be continued through the consortium. One of the developers also mentioned that he is currently missing bundles representing logical operations that allow to compose functionality in the form of a state-machine within the wiring metaphor. It was stated that this is no problem and easily extensible through the modular approach of the framework.

Graphical Programming Approach. Regarding the usage of the wiring metaphor for the graphical programming approach all developers agree that it is an appropriate way. One developer stated "Yes, it is the only right way to do it". Another developer said that for him it does not make any difference if he writes a config file directly but he also agreed that for the most of the developers and evaluators using the framework it makes sense. One user didn't know the approach before and thus did not dare to evaluate its appropriateness. Regarding the estimated learning effort for extending the functionality of the framework under the condition that they know to program with Java and the Eclipse environment and the CIF has been set up, they estimate the effort to make a small "Hello World"-like bundle and learn how to implement within the framework in short time. Estimations range from 1 hour to 2 days. For simulating contextual situations all developers but one stated that it makes sense for them as they can immediately try out what happens and use the framework to debug the setup

[2] Processing – http://www.processing.org
[3] http://vvvv.org

of a contextual study. One developer said that for him it does not make any difference whether to configure contextual situation in a text file or to use the graphical configuration approach.

The Contextual WOz Methodology. When questioning the developers about the appropriateness of the WOz technique for setting up and developing contextual studies all stated that they think it is appropriate. Although two of five said that they are not experts and thus it is hard to answer this question. This reflects the fact that the contextual WOz methodology is mainly used by evaluators and developed and set up by the developers of a contextual study. All of them liked the way the wizard tools are implemented. The client server architecture makes sense and is extensible with other interfaces that may connect to the server. Also they like that the mobile application is configured through the server and needs no adoption on the client side in order to address standard behavior as implemented in the wizard client. Also the usage of a generic protocol was highly favored. One interviewee even thought about making the particular protocol even broader for supporting any communication done within the whole framework. Also the mobile application approach was favored although there were doubts if it is necessary and useful for a standard contextual WOz laboratory study where the wizard sits at an observation place. Anyway, the possibility to enhance the wizards functionality within a mobile context provides high potential for using the technique also in other non laboratory scenarios.

Developers see a potential in the modularization approach. Also they like the CIF and the graphical programming approach for early debugging of study setups using the CIF Study Editor. A high potential is seen in being able to focus on specific functionality during study setups and using existing functionalities to integrate sensors and actuators addressing contextual information. The learning effort required is not feared and estimated quite low. One potential problem of the framework could be that no matter how much functionality is available in the framework, a developer would always be needed. Thus it will not be an all-in-one solution but and extensible framework growing with future studies.

4.3 Analysis and Discussion of the Interviews with the Evaluators

The evaluators have been interviewed especially with regard to the WOz functionality, while the system structure and the modular approach are not targeted in the results.

All of the five evaluators agreed that the wizard of Oz methodology is appropriate for simulating contextual situations. One of these five added that he does not rely that much on conducting user studies but more on putting the evaluation in the test users hand resulting in the instrumentation of prototypes. Further it was stated by one interviewee that the most important thing is that the wizard software is functional and working without errors. He already experienced problems with non-working wizard software which destroyed parts of the study. One of the evaluators doubted if it might work out of the laboratory context as he didn't have experience with it. All the evaluators agree that having a mobile device combining the whole wizard functionality is comfortable during the study and enhances the flexibility of the evaluator during the study. This allows the evaluator to focus more on the user which results in better results. Anyway, four of five evaluators state that they would use the WOz technique only for bigger studies. If it is more effort to setup the study than the advantage they gain they would abandon the framework immediately. Also there were doubts if the whole framework really could be used without a developer. One interviewee stated "If something is missing we need developers anyway". Two of five evaluators saw a potential in setting up the studies and having a mean to try out early different study conditions.

Basically all of the evaluators agree that the wizard framework is useful. A potential for developing alternative WOz studies was identified that allows the wizard for getting more involved directly in the study context where the mobile tool does not uncover him immediately as a human wizard. Anyway, a majority agreed that the wizard framework approach makes only sense for more complex studies. For simple studies they would take the manual approach if possible. Also a pre-condition is that the framework needs to work reliably.

5 Conclusions and Outlook

In this paper we presented a modular rapid prototyping framework. The framework described provides a configurable middleware tailored for context measurement and simulation using real-world as well as virtual actuators at the same time. The framework is configurable and adjustable in real-time and provides the usability engineer with Wizard of Oz (WOz) functionalities, allowing him to control the environment during a study. For integrating the WOz method into the framework we developed a set of "contextual Wizard of Oz tools" as part of the presented framework. The high flexibility of the framework and the protocol allows to use these tools in high-complex dynamic contexts in laboratory and field settings. The combination of context simulation and the WOz method enables the evaluator to conduct advanced user study setups while at the same time developing semi-functional prototypes where the non-functional part is taken over by a human wizard. In order to show the applicability and usefulness of our framework we applied it in a study and conducted structural interviews with developers using the framework to develop study setups and prototypes as well as with evaluators using the framework during user studies. The results of the interviews reflected the potential of using such a framework for user studies and pointed out important requirements that still need to be regarded during future development cycles of the whole framework. Potential problem areas identified through the interviews mainly focus on the overhead of learning and applying the contextual interaction framework in small studies. Also the current user interfaces need to be optimized in order to allow proper usage of the framework also for non-technicians (i.e. evaluators).

Future work comprises the concurrent application of the presented framework in contextual WOz studies in order to show its flexibility and applicability within a multitude of contextual situations. Such studies will cover laboratory and field settings. We are currently iterating the framework in order to be able to provide the evaluators and developers of study setups with a library of bundles containing the current state of the art sensors and actuators in order to allow them to use the framework in a broad range of studies. Nevertheless, future work will imply a steady enhancement of the available bundles which shall improve the framework by integrating additional sensors, actuators, and context reasoning tools. Future work also addresses the contextual WOz tools. They need to be iterated based on the results of the interviews where a main focus is also on the easy configuration of studies and thus the wizard utilities. Further, we will provide interfaces to integrate data logged during studies into existing observation software. The goal here is to enhance the framework through being able to provide and analyze data logged during and after a particular study.

Acknowledgements. The financial support by the Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged. The work described in this paper was supported by the Christian Doppler Laboratory for "Contextual Interfaces" and the COMET K-Project "AIR – Advanced Interface Research". Further we want to thank our colleagues Michael Humer who supports us with the implementation of the framework and during user studies, as well as Astrid Weiss, Ewald Strasser, and Sebastian Osswald who supported us throughout the user study described in this paper.

References

1. Ardito, C., Buono, P., Costabile, M.F., Lanzilotti, R., Piccinno, A.: A tool for wizard of oz studies of multimodal mobile systems. In: Proceedings of the 2nd Conference on Human System Interactions, HSI 2009, pp. 341–344. IEEE Press, Piscataway (2009)
2. Barton, J.J., Vijayaraghavan, V.: UBIWISE, A Ubiquitous Wireless Infrastructure Simulation Environment. Technical report (2002)
3. Diaper, D.: The wizard's apprentice: a program to help analyse natural language dialogues. In: Proceedings of the Fifth Conference of the British Computer Society, Human-Computer Interaction Specialist Group on People and Computers V, pp. 231–243. Cambridge University Press, New York (1989)
4. Dow, S., MacIntyre, B., Lee, J., Oezbek, C., Bolter, J.D., Gandy, M.: Wizard of Oz Support throughout an Iterative Design Process. IEEE Pervasive Computing 4, 18–26 (2005)
5. Gu, T., Pung, H.K., Zhang, D.Q.: Toward an OSGi-Based Infrastructure for Context-Aware Applications. IEEE Pervasive Computing 3, 66–74 (2004)
6. Kurz, M., Ferscha, A.: Sensor Abstractions for Opportunistic Activity and Context Recognition Systems. In: Lukowicz, P., Kunze, K., Kortuem, G. (eds.) EuroSSC 2010. LNCS, vol. 6446, pp. 135–148. Springer, Heidelberg (2010)

7. Lee, M., Billinghurst, M.: A Wizard of Oz Study for an AR Multimodal Interface. In: Proceedings of the 10th International Conference on Multimodal Interfaces, ICMI 2008, pp. 249–256. ACM, New York (2008)
8. Liu, L., Khooshabeh, P.: Paper or Interactive? A Study of Prototyping Techniques for Ubiquitous Computing Environments. In: CHI 2003 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2003, pp. 1030–1031. ACM, New York (2003)
9. Maulsby, D., Greenberg, S., Mander, R.: Prototyping an Intelligent Agent through Wizard of Oz. In: Computer Human Interaction, pp. 277–284 (1993)
10. Newman, M.W., Ackerman, M.S., Kim, J., Prakash, A., Hong, Z., Mandel, J., Dong, T.: Bringing the field into the lab: supporting capture and replay of contextual data for the design of context-aware applications. In: Proceedings of UIST 2010, pp. 105–108. ACM, New York (2010)
11. Obrenovic, Z., Martens, J.-B.: Sketching interactive systems with sketchify. ACM Trans. Comput.-Hum. Interact., 18:4:1–4:38 (2011)
12. Salber, D., Coutaz, J.: Applying the Wizard of oz Technique to the Study of Multimodal Systems. In: Bass, L.J., Unger, C., Gornostaev, J. (eds.) EWHCI 1993. LNCS, vol. 753, pp. 219–230. Springer, Heidelberg (1993)
13. Schlögl, S., Doherty, G., Karamanis, N., Luz, S.: WebWOZ: A Wizard of Oz Prototyping Framework. In: Proc. of EICS 2010, pp. 109–114. ACM, New York (2010)
14. Strasser, E., Weiss, A., Grill, T., Osswald, S., Tscheligi, M.: Combining Implicit and Explicit Methods for the Evaluation of an Ambient Persuasive Factory Display. In: Paternò, F., de Ruyter, B., Markopoulos, P., Santoro, C., van Loenen, E., Luyten, K. (eds.) AmI 2012. LNCS (LNAI), vol. 7683, Springer, Heidelberg (2012)
15. Suchman, L.A.: Plans and Situated Actions: The Problem of Human-Machine Communication. Cambridge University Press, New York (1987)

Recognizing the User Social Attitude in Multimodal Interaction in Smart Environments

Berardina De Carolis, Stefano Ferilli, and Nicole Novielli

Dipartimento di Informatica, University of Bari, Italy
{decarolis,ferilli,novielli}@di.uniba.it

Abstract. Ambient Intelligence aims at promoting an effective, natural and personalized interaction with the environment services. In order to provide the most appropriate answer to the user requests, an Ambient Intelligence system should model the user by considering not only the cognitive ingredients of his mental state, but also extra-rational factors such as *affect, engagement, attitude*, and so on. This paper describes a study aimed at building a multimodal framework for recognizing the social response of users during interaction with embodied agents in the context of ambient intelligence. In particular, we describe how we extended a model for recognizing the social attitude in text-based dialogs by adding two additional knowledge sources: speech and gestures. Results of the study show that these additional knowledge sources may help in improving the recognition of the users' attitude during interaction.

1 Introduction

Ambient Intelligence (AmI) systems should promote an effective, natural and personalized experience in interacting with services provided by the environment. For this reason, these systems should introduce novel means of communication between humans and the surrounding environment [1]. Multimodal interaction paradigms that combine several modalities are a powerful approach to promote natural and effortless experience [2]. Moreover, multimodal dialog systems allow AmI systems to build user models that, besides modeling cognitive ingredients of the user's mental state (interests, preferences, etc.), also consider extra-rational factors such as *affect, engagement, attitude*, and so on. In our opinion these factors become of particular importance when the interaction takes place in everyday life environments.

In this paper we describe a study aimed at building a framework for recognizing the social response of users in natural multimodal interaction in the context of ambient intelligence. In particular, social agents endowed with human-like behavior have been used as interfaces in smart environments, especially in assisted living contexts [2,3,4,5,6]. In fact, according to Reeves et al.'s work [7] on the media equation, in which they reported that people react to media as if they were social actors, there is a growing interest in on-screen agents and robotic characters that are endowed with human-like behavior.

Embodied Conversational Agents (ECAs) and Social Robots, if properly designed and implemented, may improve the naturalness and effectiveness of interaction

F. Paternó et al. (Eds.): AmI 2012, LNCS 7683, pp. 240–255, 2012.

between users and smart environment services. They have the potential to involve users in human-like conversations using verbal and non-verbal signals for providing feedback, showing empathy and emotions in their behavior [8,9]. Thanks to these features, embodied agents can be exploited in the AmI domain, where it is important to settle long-term relations with the user. In fact, social embodied agents are commonly accepted as a new metaphor of Human Computer Interaction in several application domains since they make interaction more effective and enjoyable [10] also providing emotional, appraisal and instrumental support.

A natural way for humans to interact with such agents is using a combination of speech and gestures. Thus, in developing models for recognizing the user social attitude in the AmI context, we propose a framework that integrates the analysis of the linguistic component of the user's communicative act with the analysis of the acoustic features of his spoken sentences and of his gestures. The idea underlying our approach is that the combination of these different input modalities may improve the recognition of multimodal behaviors that may denote the openness attitude of the users towards the embodied agent.

We built our framework as a Dynamic Bayesian Network (DBN) [11], due to the ability of this formalism to represent uncertainty and to its graduality in building an image of the user's cognitive and affective state of mind. The dynamic recognition, during a dialogue, of social signs in language, prosody and gestures enables not only to estimate the overall social attitude value but also to adapt the dialogue plan accordingly.

To this aim we designed an experimental setting to collect a corpus of multimodal conversations with an ECA, displayed on a wall of a smart environment, in a Wizard of Oz (WoZ) simulation study [12,13]. Then, after carefully examining our corpus and considering suggestions from the studies about verbal and non-verbal expression of social attitude [14-17], we annotated each multimodal user dialogue move in the corpus according to the social attitude it conveyed. Then, we used the resulting dataset to extract the knowledge necessary to build and tune the model. In order to test the resulting model in real settings, we performed another experiment aimed at automatically recognizing the user's social attitude during the dialog. Specifically, as regards the new modalities, we used a voice classifier able to recognize the valence and arousal in the user's spoken sentence, while for gesture recognition we used Microsoft Kinect [18] with Kinect DTW (Dynamic Time Warping) [19].

This paper is structured as follows: in Section 2 we provide an overview on the background and the motivation of this research. In Section 3 we summarize which are the signals that can be considered as relevant for modeling the social attitude by using linguistic, acoustic and body features. In Section 4 the experiment for collecting the corpus and the dynamic modeling approach used for integrating the results of the multimodal analysis are presented; then, Section 5 describes the results of the evaluation of the model and provides a sample dialogue that demonstrates the functioning of our framework. Conclusions and future work directions are reported in Section 6.

2 Background and Motivation

There are two prevailing views for allowing communication between users and the environment [20]. In the former, interaction is made through a multitude of task-specific information appliances [21]. The other is based on the use of a centralized and intelligent interface that proactively tries to anticipate users' needs [22]. In the latter case, according to Reeves et al.'s work on the media equation [7] in which they reported that people react to media as if they were social actors, there is a growing interest in on-screen agents and robotic characters that are endowed with human-like behavior. These issues become even more relevant in the AmI context since the main goal of these systems is both to allow monitoring the wellness of the user and to provide natural and pleasant interfaces to the smart environment services. One consequence of embedding technology in our everyday life environments is that it will be continuous in time and will involve a disparate range of devices providing support for different types of services [3]. According to the ubiquitous computing vision, these devices will disappear in the background and the user will relate dialogically to the environment. Since people already have a tendency to attribute human-like properties to interactive systems [7], it is expected that implementing human-like properties in such environment dialog systems will have an important impact on the user–system interaction. Thus, the use of embodied agents could give to the users the feeling of cooperating with a human partner or having a companion rather than just using a tool [23]. However, if the social agent's behavior is not properly designed and implemented there could be the risk of creating unrealistic expectations on the part of the users, and to lead to wrong mental models about the system's functionality and capacity [24]. On the other hand, several studies report successful results on how expressive ECAs and robots can be employed as interaction metaphors in the assisted living and in other ones [4,5,6] where it is important to settle long-term relations with the user [25]. Conversational agents, ranging from voice agents to embodied on-screen animated characters and physically embodied robots, build on this concept. These agents use speech, gaze, gesture, intonation and other non-verbal modalities to emulate the experience of human face-to-face conversation with their users [9]. In order to achieve successful interaction, embodied agents must be able to model not only the cognitive ingredients of the user's mental state (interests, preferences, beliefs), but they should also consider extra-rational factors such as *affect, engagement, attitude*, and so on. After several forms of 'anthropomorphic behavior' of users towards technologies were demonstrated [7], various terms and concepts have been employed to denote this behavior and describe it. Paiva [10] talks about *empathy*, Hoorn and Konijn [26] address the concept of *engagement, involvement*, *sympathy* and their contrary, *distance*. Cassell and Bickmore [27] adopt Svennevig's theory of *interpersonal relations*. We refer to Scherer's concept of *interpersonal stance* as a category which is *"characteristic of an affective style that spontaneously develops or is strategically employed in the interaction with a person or a group of persons, coloring the interpersonal exchange in this situation (e.g. being polite, distant, cold, warm, supportive, contemptuous)"*. In particular, in referring to the social response of users to ECAs, we distinguish warm/open from cold/close/distant *social attitude*, according to the Andersen and Guerrero's definition of interpersonal warmth [15] as "*the pleasant, contented, intimate feeling that occurs during positive interactions with* *friends, family, colleagues and romantic partners*". A large variety of verbal and non-verbal markers of interpersonal stance have been proposed: body distance, memory, likeability, physiological data, task performance, self-report and others [15,16,17,28,29].

3 Signals of Social Attitude in Multimodal Interaction

Social interfaces are designed to interact with the users in a way that takes advantage of principles from social interaction between humans in order to achieve more 'natural' and intuitive interaction with complex systems [30]. In this communicative process humans use several communication channels. We can distinguish between verbal communication (words, sentences) and non-verbal communication (gesture, body language or posture, facial expression and eye contact, prosodic features of speech such as intonation and stress, and so on). Therefore in developing a model for recognizing the user social attitude in multimodal interaction with an embodied agent placed in a Smart Environment we decided to consider not only the linguistic content of the message but also the acoustic features of the user spoken sentence, and the gestures. In the following sections we describe the multimodal indicators of social attitude we employ in our approach.

3.1 Signals of Social Attitude in Language

Recognition of signals of social attitude in the language is performed according to the approach described in [14]. In particular, we defined a taxonomy of signals for analyzing social communication in text-based interaction which employs *affective, cohesive and interactive indicators* (similarly to [15]):

- personal address and acknowledgement (using the name of the persons to which one is responding, restating their name, agreeing or disagreeing with them),
- feeling (using descriptive words about how one feels),
- paralanguage (features of language which are used outside of formal grammar and syntax, which provide additional enhanced, redundant or new meanings to the message),
- social sharing (sharing of information not related to the discussion),
- social motivators (offering praise, reinforcement and encouragement),
- value (set of personal beliefs, attitudes),
- negative responses (disagreement with another comment),
- self-disclosure (sharing personal information).

3.2 Signals of Social Attitude in Speech

According to several studies [31-33] the linguistic analysis is not enough to properly interpret the real user's communicative intention and the attitude of the user. For instance, the user can pronounce the same sentence with different emotional attitudes in order to convey different communicative intents. While words still play an important role in the recognition of communicative intents, taking into account the user attitude while speaking adds another source of knowledge that is important for resolving ambiguities and compensate for errors [32, 34]. Research in emotional speech has shown that acoustic and prosodic features can be extracted from the speech signal and used to develop models for recognizing emotions and attitudes [35]. In fact, the effects of emotion in speech tend to alter the pitch, timing, voice quality, and articulation of the speech signal and reliable acoustic features can be extracted from speech that vary with the speaker's affective state. In order to classify the social attitude from the analysis of the spoken utterance we used an approach similar to [34,36]. We decided to use the valence of the sentence in terms of a *negative/cold* style vs. a *positive/friendly* one and the arousal from *low* to *high* in a three-points scale as signals of social attitude. Recognizing the value of only these two dimensions is justified since the valence indicates a failure/success in the achievement of the user's goal and, if it is related to the arousal, it allows to distinguish for instance a negative/cold attitude towards the agent from sadness related to a personal mental state. Moreover, a classification model based on simple features allows handling online analysis of the user's attitude [36].

3.3 Signals of Social Attitude in Gestures

In order to endow an embodied agent with the ability of recognizing the social attitude also from gestures, we considered those, involving arms and hands position, that denote an attitude of openness or closure [37-40]. In fact, our gestural habits can convey several kinds of information regarding our personality, cultural background and mood [41,42]. In doing so, we analyzed the literature on the topic and we classified gestures according to the arms and hands position. Arms are quite reliable indicators of mood and feeling, especially when interpreted with other signals. Arms act as defensive barriers when across the body, and conversely indicate feelings of openness and security when in open positions, especially combined with open palms. Hands are also involved in gesture recognition. They are very expressive parts of the body and they are used a lot in signaling consciously - as with emphasizing gestures - or unconsciously - as in a wide range of unintentional movements which indicate otherwise hidden feelings and thoughts. For instance, picking nose denotes a social disconnection, inattentiveness or stress, neck scratching denote doubt or disbelief. Even if the position of the legs cannot be considered as a part of gesture, in evaluating the social attitude we take into account whether the legs are crossed or not. This information can be used to support the corresponding arms signals, for example in conjunction with crossed arms they allow to recognize a strong closure or rejection.

4 Towards a Model for Multimodal Social Attitude Recognition

As explained in the previous section, researchers proposed a large variety of markers of social attitude related to verbal and nonverbal behavior. In developing a social agent able to interact with user of a smart environment, we decided to study whether and how the analysis of a combination of the communication modalities could help us in building an accurate user model in terms of social attitude recognition. We collected a corpus of dialogues with a set of Wizard of Oz studies and annotated the linguistic, acoustic and gestural components of the user input. The resulting annotated dataset has been used to build the multimodal framework for social attitude recognition.

4.1 Collecting the Corpus

The experiment was conducted in fitness center. This choice is justified by the fact that one of the main limitations of this kind of experiments may derive by the non-voluntary nature of the interaction and lack of motivations in interacting with the ECA, since subjects may not be necessarily interested to the dialogue topic. In order to avoid this bias, we hypothesize that customers of a gym could be naturally interested in receiving suggestions about healthy eating. For this reason, 2 weeks preceding the experiment, we advertised people in the gym about the event. Thus, subjects involved in the study spontaneously requested to interact with the ECA to receive suggestions about nutrition and healthy eating.

The study involved 10 subjects aged between 19 and 28 years old, equally distributed by gender and background (they were all Italian undergraduate students attending the fitness center). We assigned to each subject the same goal of information seeking: *getting information about a correct diet in order to stay in shape*. To obtain this information, subjects could dialogue with an ECA displayed on the wall of one of the rooms of the fitness center, playing the role of an expert in the healthy eating context. Before starting the experiment we administered to each subject a simple on-line questionnaire aiming at collecting some personal data (age and gender), understanding their background (degree of study), their interest and experience in the healthy eating topic. After this assessment phase, subjects were asked to give their consent to video-record their behavior during the experiment. Then, subjects could start interacting with the agent. Each user move was video-recorded. After the experiment the subject was fully debriefed.

We collected 10 multimodal dialogues with 204 moves overall. After the first analysis of the corpus we could notice that different signals of social attitude could be observed by looking at verbal and non-verbal expressions. Each move was then annotated by a human rater with respect to the perceived user social attitude, conveyed by the turn. The annotation experiment has been performed by three researchers in the field of human-computer interaction. Each move received a final label for the social attitude using a majority agreement criterion. The annotation language is reported in Table 1.

The following are examples of annotation:

Example 1a: System: *Hi' my name is Valentina, I'm here to suggest you how to eat well.* User: *Ok.* (L: nl – A: NI – G: Closure – Satt: slightly negative)

Example 1b: System: *Hi' my name is Valentina….* User: *Hi Valentina, my name is Carla!* (L: fsi – A: PI – G: Open – Satt: positive)

According to the result of the annotation process we defined the model and we conducted a preliminary evaluation for tuning its structure and parameters.

Table 1. – Labels for signals of social attitude

Signals of social attitude in language	
Label	**Signals**
Friendly self-Introduction (**fsi**)	Greetings and Self introduction
Colloquial style (**cstyle**)	Paralanguage, Terms From Spoken Language, Dialectal and Colored Forms, Proverbs and Idiomatic Expressions, Diminutive or Expressive Forms
Talks about self (**talks**)	First person pronouns, first person auxiliary, knowledge, attitude, ability, liking or desiring verbs.
Questions about the agent (**qagt**)	Second person pronouns, first person auxiliary, knowledge, attitude, ability, liking or desiring verbs.
Positive or Negative comments (**poscom / negcom**)	Generic comments, Expressions of agreement or disagreement, Message evaluation, Evaluation of agent's politeness, Evaluation of agent's competence, Remark about agent's repetitivity, Evaluation of agent's understanding ability
Friendly farewell (**ffwell**)	Expressions of farewell and Thanking
Neutral (**nl**)	Neutral language expressions
Acoustic Signals of social attitude	
Label	**Signals**
Neutral Intonation (**NI**)	neutral intonation (neutral valence and low or medium arousal)
Negative intonation (**NegI**)	negative intonation (negative valence)
Positive Intonation (**PI**)	friendly positive intonation (positive valence)
Signs of Social Attitude in Gestures	
Label	**Signals**
Open attitude (**Open**)	palm(s) open, knees apart, elbows away from body, hands not touching, legs uncrossed,…
Closure attitude (**Closure**)	crossed arms, gripping own upper arms, crossed legs,…
Negative attitude (**Negative**)	adjusting cuff, watchstrap, tie, etc., using an arm across the body, touching or scratching shoulder using arm across body, picking nose, pinching bridge of nose, neck scratching, …

4.2 Dynamic Modeling of the User Social Attitude

The modeling of the user's cognitive and affective state is a key factor for the development of affect-aware systems. User modelling allows socially intelligent systems to adapt to the users' behavior by constantly monitoring it and by continuously collecting their direct and indirect feedback.

In deciding which approach to adopt in order to model the user social attitude we analyzed some of the relevant research work in the domain and we noticed due to the uncertainty typical of emotion modeling, probabilistic approaches are very often used for user affect modeling tasks. For instance Prendinger et al. [43] used a probabilistic decision model in the scope of the Empathic Companion project. A probabilistic model (Dynamic Decision Network) is used also by Conati [44] to monitor a user's emotions and engagement during the interaction with educational games. Sabourin et al. [45] present a study about designing pedagogical empathic virtual in which a cognitive model structured as a Bayesian network is adopted. A different approach is used by Caridakis et al. [46] for recognizing the user emotional state in naturalistic interaction. In this case, recognition is performed using neural networks, due to their efficacy in modeling short term dynamic events that characterize the facial and acoustic expressivity of users. A detailed review on methods and approaches adopted in literature for performing user emotion modeling may be found in [47].

In our framework, the user modeling procedure integrates (i) language analysis for linguistic cues extraction, (ii) prosodic analysis and (iii) gesture recognition into a Dynamic Belief Network (DBN) [11]. DBNs, also called time-stamped models, are local belief networks (called time slices) expanded over time; time slices are connected through temporal links to constitute a full model. The method allows us to deal with uncertainty in the relationships among the variables involved in the social attitude estimation. The DBN formalism is particularly suitable for representing situations that gradually evolve from a dialog step to the next one. The DBN (Figure 1) is employed to infer how the social attitude of the user evolves during the dialog according to signals expressed in the verbal and non-verbal part of the communication.

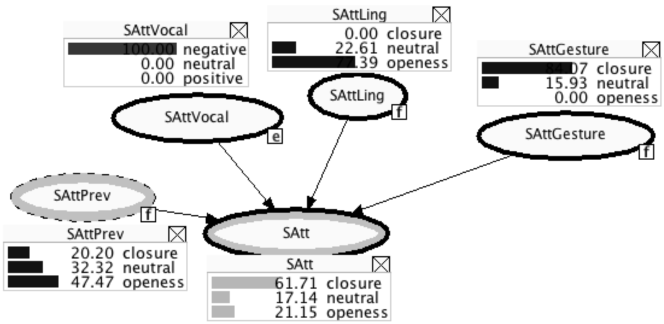


Fig. 1. DBN modelling the user social attitude

The social attitude (SAtt) is the variable we want to monitor, which depends on *observable* ones, i.e. the recognized significative signals in the user move deriving from the linguistic, acoustic and gestural analysis. These nodes may correspond to a simple variable, as in the case of the SAttVocal variable, or to a nested belief network as in the case of the SAttLing and SAttGesture variables whose probability distribution is calculated by the correspondent belief networks. Links among variables describe the causal relationships among stable characteristics of the users and their behavior, via intermediate nodes. DBNs, as employed in this paper, are said to be 'strictly repetitive models'. This means that structure and parameters of individual time slices is identical and temporal links between two consecutive time slices are always the same. We use a special kind of strictly repetitive model in which the Markov property holds: the past has no impact on the future given the present. In our simulations, every time slice corresponds to a user move and temporal links are established only between dynamic subject characteristics in two consecutive time slices.

At the beginning of interaction, the model is initialized; at every dialog step, knowledge about the evidences produced by the multimodal analysis are entered and propagated in the network: the model revises the probabilities of the social attitude node. The new probabilities of the signs of social attitude are used for planning the next agent move, while the probability of the social attitude node supports revising high-level planning of the agent behavior. In particular, Figure 2 shows the model for inferring the social attitude from linguistic signals. More details about this model can be found in [14].

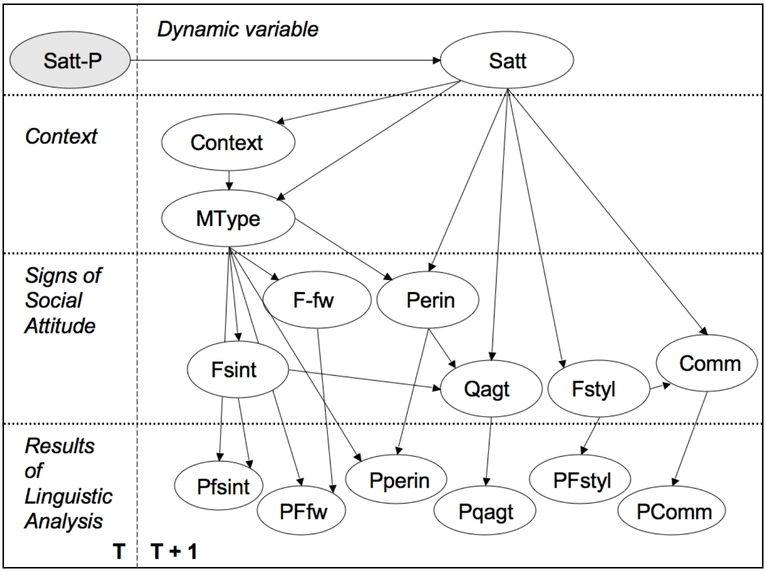


Fig. 2. User Model for the Social Attitude for Linguistic Analysis, a generic time-slice

As far as the voice analysis is concerned, our model classifies the spoken utterance according to the recognized valence and arousal. In parallel with the manual annotation process, the audio files relative to the moves in the corpus were analyzed using Praat functions [48] for extracting from the audio file features related to: i) the variation of the fundamental frequency (pitch minimum, mean, maximum and standard deviation, slope); ii) the variation of energy and harmonicity (min, max and standard deviation); iii) the central spectral moment, standard deviation, gravity centre, skewness and kurtosis and iv) the speech rate. Following an approach similar to [36], our classifier exploits several algorithms, the J48 and NNge [49] showed to be the most accurate ones (85% and 89% validated using a *10 Fold Cross Validation* technique).

Using simple rules, from the values of valence and arousal classified by this module we set the evidence of the variable SAttVocal in the model. For instance, when the valence is positive and the arousal is high, as in the example reported in Figure 3, the system set the attitude to positive.

As far as recognizing signals of social attitude from gestures is concerned, since in our system gesture recognition is performed using Microsoft Kinect, we had to consider only those nodes in the skeleton that the Kinect SDK is able to detect. Moreover, at present the Kinect skeleton does not includes nodes for detecting the position of fingers, for this reason at present we consider only the subset of gestures. Figure 3 shows a generic time-slice of the BN for modeling the user social attitude from gestures. In particular, the gesture recognized with Kinect becomes the evidences of the roots nodes of the network. Then these evidences are propagated in the net and the probabilities of the SAttGesture node is computed given the probabilities of intermediate nodes, Hands, Arms and CrossedLegs, denoting the social attitude expressed by each of them.

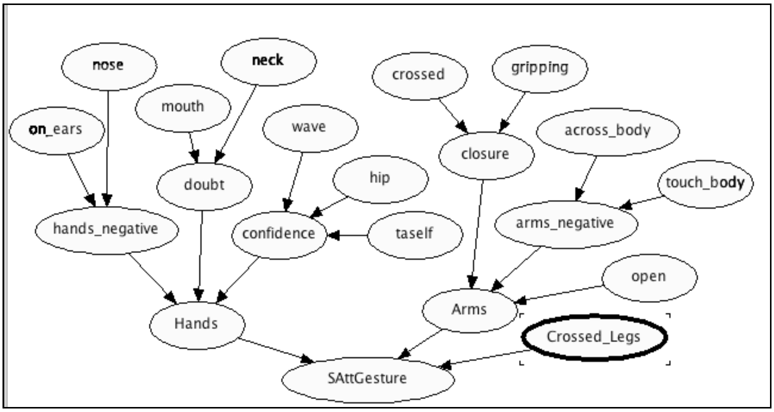


Fig. 3. User Model for the Social Attitude from Gestures, a generic time-slice

5 Evaluation of the Model

In order to evaluate the model, we performed another experiment for understanding whether a variation in the threshold of the probability of the monitored variable (SAtt) correctly affects sensitivity and specificity of the model in recognizing this feature. As in the previous experiment, we performed a Wizard of Oz simulation study in the same fitness center. This time the user moves were automatically annotated in terms of social attitude using the three models described above. Participants involved in this study, 8 in total, were selected among the people that asked to participate to this second round having a background consistent with the one of subjects involved in the previous experiment. The experiment setting and procedure was the same than in the previous experiment except for the fact that the subject gestures were detected with Microsoft Kinect and recognized using KinectDTW (Dynamic Time Warping) [19] that we previously trained for recognizing signals concerning arms, hands and legs. Moreover, the subject was wearing a wireless microphone whose output was sent to the speech processing system. This time we collected a corpus of about 200 moves that were annotated by three human raters with respect to the perceived user social attitude, conveyed by the turn. Each move received a final label for the social attitude using a majority agreement criterion. Then we automatically annotated each move using the modules defined for the language and speech analysis and the DTW for the gestures. These evidences constitute the features describing each user move, according to the signals variables in the previously described models. According to the result of the annotation experiments (both manual and automatic) we conducted a preliminary evaluation for tuning the structure and the probabilities of our multimodal framework. Results of this experiment show that the model is quite accurate since it agrees with the human annotation at 89% in recognizing the overall social attitude.

5.1 An Example of Dialogue

In this section we show an example of a dialogue and how the proposed framework recognized the social attitude in each move. In this experiment the wizard followed the same standard dialog plan for all the subjects and did not apply any kind of strategy for preventing or recovering a negative social attitude. This issue is part of our future work. In the following dialogue (translated from Italian) the agent and user moves are denoted respectively by Vi and Uj. The letters L, A and G denote signals of social attitude in the language, voice and gestures respectively. *Satt* indicates the probability value of recognized social attitude in the user move. Where no annotation is provided, our framework did not recognize any relevant signal of social attitude.

U0: (L:-, A: -, G: crossed arms) – Satt: 0.51 (initial attitude of the user)
V1: Hi, I'm Valentina! I'm here to give you information about correct eating habits.
U1: Hi, my name is Carla. (L:friendly self introduction- A: positive valence, medium arousal - G: crossed arms) – Satt: 0.61
V2: Hi Carla, nice to meet you. How can I help you?
U2: I would like to learn more about a correct diet! (L: neutral – A: positive valence, medium arousal – G: open arms) – Satt: 0.8
V3: Good.Let's start with a pratical example.What did you eat for breakfast this morning?
U3: Coffee and some cookies.
V4: And for lunch?
U4: Just a sandwich
V5: What did you eat for dinner yesterday night?
U5: Meat and salad.
V6: Good. Your diet is not so bad. However, like you, most of people forget about the importance of eating more fruit and vegetables.International Research about foods proves that fruit and vegetables are very important in a healthy diet…(benefits of vitamins and fibers). Now, try to think again about your diet, is it healthy?
U6: Well, I should change something (L: talk about self – A: neutral valence, low arousal – G: touching nose) - Satt: 0.48

V7: Good, let's try to do something then. Did you ever think to follow a correct diet?
U7:Yes, sometimes I did it and I did not get very good results (L: talk about self- A: negative valence, medium arousal, G: using an arm across the body). – Satt: 0.21
V8: Why?
U8: I don't know. (A: negative valence, low arousal- G: crossed arms). – Satt: 0.08
V9: Could you tell me why, so that I can help you?
U9: I do not want to answer. (L: - A:negative valence, medium arousal – G: Crossed arms and crossed legs) Satt: 0.02
V10: OK I understand that sometimes talk about self can be difficult….

In the present prototype, linguistic cues of *friendly style* and *talks about self* are detected and evidences about these signals contribute to increase the overall likelihood of observing a warm social attitude of the user from the linguistic point of view.

Valence and arousal are detected from acoustic parameters using a classifier and gestures are detected using Kinect and a library for recognition of signals related to gestures based on DTW are given as evidences to the model for recognising a more closed or open attitude.

For instance, in the move U7, the linguistic analysis provides an evidence of Talk_about_Self (the Perin variable in Figure 2) to allow the recognition of a warm social attitude through linguistic analysis. However, the acoustic analysis classifies the valence as negative and the arousal as medium and the recognized gesture is touching nose, thus denoting a negative/cold attitude. The model instance showed in Figure 1 illustrates this particular case.

Then the agent, in the move V8, asks which is the reason of this result. In the next move U8 the user says that she does not want to answer with a negative prosody and crosses her arms. These signals provide evidences of a negative/cold social attitude (Figure 4).

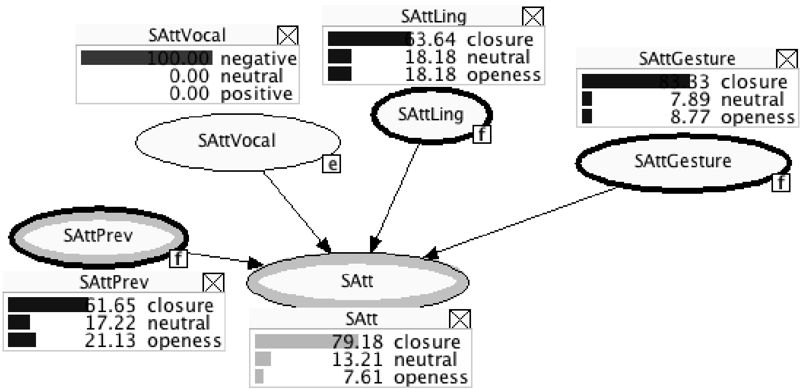


Fig. 4. Social Attitude Recognition for Move U8

The subsequent move V9, in which the agent keeps asking for the reason, causes a further closure since the user move, U9, is recognized linguistically as a Negative Comment, acoustically as having a negative valence and from gesture and legs position as a strong closure (Figure 5 and 6).

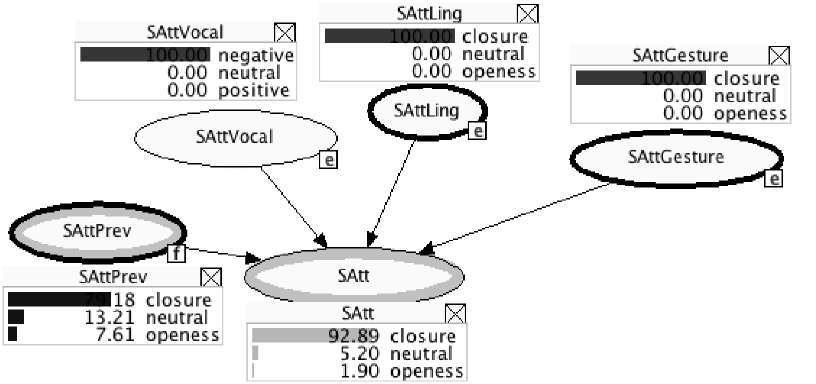


Fig. 5. Social Attitude Recognition for Move U9

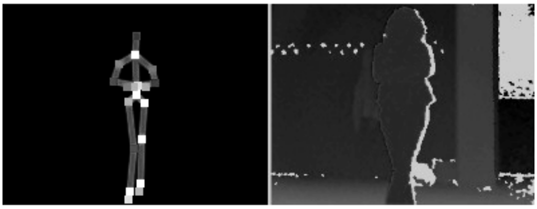


Fig. 6. KinectDTW recognition of crossed arms and legs

6 Conclusions

This research builds on prior work on affect modeling and dialog simulation [14]. In this paper we enrich the model for recognizing social attitude with the analysis of signals regarding not verbal communication: prosody and gesture in particular. The two approaches to social attitude modeling using speech and language analysis have been validated also in our previous research, with satisfying results. In this paper we have proposed an extension of the multimodal analysis to gestures, according to the meanings that psycholinguistic researchers attach to gestures in conversations.

The first consideration regards the flexibility of the proposed approach. In particular, we are able to attach/remove new modules for the analysis of different users behavior with respect to the changes occurring in the interaction scenario (e.g. linguistic analysis would be enough in case of textual interaction, while multimodal analysis is of great help when the user interacts with a smart environment, in which sensors are able to capture his speech and body language).

In our future research we plan to improve the gesture recognition analysis. We are currently testing the proposed approach with the FUll Body Interaction (FUBI) framework [50] since it allows for hands recognition. However, we do not see this as a limitation of our approach since new devices, like the new Kinect 2, should allows for a better gesture recognition, thus improving the precision of the model. In the near future, we plan to perform more evaluation studies in order to test the robustness of our framework for the social attitude recognition in different scenarios and with respect to different interaction modalities with both ECAs and Social Robots. Moreover, we plan to test different strategies for recovering the dialog when the social attitude starts decreasing by extending the model with contextual features in order to use it in a prognostic way.

Acknowledgments. We thank our students for their work on the prototype and the Body Energy fitness center in Mola di Bari for kindly enabling us to perform the new WoZ experiments.

References

[1] Riva, G., Vatalaro, F., Davide, F., Alcaniz, M. (eds.): Ambient Intelligence: the Evolution of Technology, Communication and Cognition Towards the Future of Human-Computer Interaction Emerging Communication (2005)

[2] López-Cózar, R., Callejas, Z.: Multimodal Dialogue for Ambient Intelligence and Smart Environments. In: Nakashima, H., Aghajan, H., Augusto, J.C. (eds.) Intelligence, p. 93808. Springer US (2010)

[3] Nijholt, A., de Ruyter, B., Heylen, D., Privender, S.: Social Interfaces for Ambient Intelligence Environments. In: Aarts, E., Encarnaçao, J. (eds.) True Visions: The Emergence of Ambient Intelligence, ch. 14, pp. 275A–289A. Springer, New York (2006)

[4] Ortiz, A., del Puy Carretero, M., Oyarzun, D., Yanguas, J.J., Buiza, C., Gonzalez, M.F., Etxeberria, I.: Elderly Users in Ambient Intelligence: Does an Avatar Improve the Interaction? In: Stephanidis, C., Pieper, M. (eds.) ERCIM UI4ALL Ws 2006. LNCS, vol. 4397, pp. 99–114. Springer, Heidelberg (2007)

[5] Cesta, A., Cortellessa, G., Giuliani, M.V., Iocchi, L., Leone, G.R., Nardi, D., Pecora, F., Rasconi, R., Scopelliti, M., Tiberio, L.: The RoboCare Assistive Home Robot: Environment, Features and Evaluation, The RoboCare Technical Report, RC-TR-0906-6 (2006)

[6] van Breemen, J.N.: iCat: a generic platform for studying personal robot applications. Paper Presented at the IEEE SMC, Den Haag (2004)

[7] Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, Cambridge (1996)

[8] Bickmore, T., Cassell, J.: Social Dialogue with Embodied Conversational Agents. In: van Kuppevelt, J., Dybkjaer, L., Bernsen, N. (eds.) Advances in Natural, Multimodal Dialogue Systems, pp. 1–32. Kluwer Academic, New York (2005)

[9] Cassell, J., Prevost, S., Sullivan, J., Churchill, E.: Embodied Conversational Agents. MIT Press, Cambridge (2000)

[10] Paiva, A. (ed.): Empathic Agents. Workshop in conjunction with AAMAS 2004 (2005)

[11] Jensen, F.V.: Bayesian Networks and Decision Graphs. Springer (2001)

[12] Whittaker, S., Walker, M., Moore, J.: Fish or Fowl: A Wizard of Oz evaluation of dialogue strategies in the restaurant domain. In: Language Resources and Evaluation Conference (2002)
[13] Oviatt, S., Adams, B.: Designing and Evaluating Conversational Interfaces With Animated Characters. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.) Embodied Conversational Agents. The MIT Press (2000)
[14] Novielli, N., de Rosis, F., Mazzotta, I.: User Attitude Towards an Embodied Conversational Agent: Effects of the Interaction Mode. Journal of Pragmatics 42(9), 2385–2397 (2010) ISSN: 0378-2166, doi:10.1016/j.pragma.2009.12.016
[15] Andersen, P.A., Guerrero, L.K.: Handbook of Communication and Emotions. Research, theory, applications and contexts. Academic Press (1998)
[16] Polhemus, L., Shih, L.-F., Swan, K.: Virtual interactivity: the representation of social presence in an on line discussion. In: Annual Meeting of the American Educational Research Association (2001)
[17] Swan, K.: Immediacy, social presence and asynchronous discussion. In: Bourne, J., Moore, J.C. (eds.) Elements of Quality Online Education, vol. 3. Sloan Center For Online Education, Nedham (2002)
[18] `http://kinectforwindows.org`
[19] KinectDTW, `http://kinectdtw.codeplex.com/`
[20] Markopoulos, P.: Designing ubiquitous computer human interaction: the case of the connected family. In: Isomaki, H., Pirhonen, A., Roast, C., Saariluoma, P. (eds.) Future Interaction Design. Springer (2004)
[21] Norman, D.: The Invisible Computer. MIT Press, Cambridge (1998)
[22] Pentland, A.: Perceptual intelligence. Communications of the ACM 43(3), 35–44 (2000)
[23] Reilly, W.S.N.: Believable Social and Emotional Agents. PhD thesis (1996)
[24] Brahnam, S., De Angeli, A.: Special issue on the abuse and misuse of social agents. Interacting with Computers 20(3), 287–291 (2008)
[25] Bickmore, T.: Relational Agents: Effecting Change through Human-Computer Relationships. PhD Thesis, Media Arts & Sciences, Massachusetts Institute of Technology (2003)
[26] Hoorn, J.F., Konijn, E.A.: Perceiving and Experiencing Fictitional Characters: An integrative account. Japanese Psychological Research 45, 4 (2003)
[27] Cassell, J., Bickmore, T.: Negotiated collusion: modelling social language and its relationship effects in intelligent agents. User Modelling and User-Adapted Interaction 13, 1–2 (2003)
[28] Bailenson, J.N., Aharoni, E., Beall, A.C., Guadagno, R.E., Dimov, A., Blascovich, J.: Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. In: Proceedings of 7th Annual International Workshop on PRESENCE, pp. 216–223 (2004)
[29] Bailenson, J.N., Swinth, K.R., Hoyt, C.L., Persky, S., Dimov, A., Blascovich, J.: The independent and interactive effects of embodied agents appearance and behavior on self-report, cognitive and behavioral markers of copresence in Immersive Virtual Environments. Presence 14(4), 379–393 (2005)
[30] Cramer, H., Kemper, N., Amin, A., Wielinga, B., Evers, V.: Give me a hug: the effects of touch and autonomy on people's responses to embodied social agents. Comp. Anim. Virtual Worlds 20, 437–445 (2009), doi:10.1002/cav.317
[31] Mozziconacci, S., Hermes, D.J.: Role of intonation patterns in conveying emotion in speech. In: Proceedings, International Conference of Phonetic Sciences, San Francisco (1999)

[32] Bosma, W.E., André, E.: Exploiting Emotions to Disambiguate Dialogue Acts. In: Nunes, N.J., Rich, C. (eds.) Proc. 2004 Conference on Intelligent User Interfaces, Funchal, Portugal, January 13, pp. 85–92 (2004)

[33] Litman, D., Forbes, K., Silliman, S.: Towards emotion prediction in spoken tutoring dialogues. In: Proceedings of HLT/NAACL (2003)

[34] De Carolis, B., Cozzolongo, G.: Interpretation of User's Feedback in Human-Robot Interaction. Journal of Physical Agents 3(2), 47–58 (2009) ISSN: 1888-0258

[35] Sundberg, J., Patel, S., Björkner, E., Scherer, K.R.: Interdependencies among voice source parameters in emotional speech. IEEE Transactions on Affective Computing 2(3) (2011)

[36] Vogt, T., André, E., Bee, N.: EmoVoice — A Framework for Online Recognition of Emotions from Voice. In: André, E., Dybkjær, L., Minker, W., Neumann, H., Pieraccini, R., Weber, M. (eds.) PIT 2008. LNCS (LNAI), vol. 5078, pp. 188–199. Springer, Heidelberg (2008)

[37] Allan, Pease, B.: The Definitive Book of Body Language. Bantam Books (2006)

[38] Feldman, R.S.: Applications of Nonverbal Behavioral Theories and Research. Lawrence Erlbaum Associates, Inc., New Jersey (1992)

[39] Knapp, M., Hall, J.: Nonverbal communication in human interaction. Holt, Rinehart & Winsten, Inc., Orlando (1992)

[40] Malandro, L.A.: Non-verbal Communication, 2nd edn. Random House, New York (1983,1989)

[41] Efron, D.: Gesture, Race and Culture. Mouton and Co. (1972)

[42] Rehm, M., Nakano, Y., André, E., Nishida, T.: Culture-Specific First Meeting Encounters between Virtual Agents. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 223–236. Springer, Heidelberg (2008)

[43] Prendinger, H., Mori, J., Ishizuka, M.: Recognizing, Modeling, and Responding to Users' Affective States. In: Ardissono, L., Brna, P., Mitrović, A. (eds.) UM 2005. LNCS (LNAI), vol. 3538, pp. 60–69. Springer, Heidelberg (2005)

[44] Conati, C.: Probabilistic assessment of user's emotions in educational games. Applied Artificial Intelligence 16, 555–575 (2002)

[45] Sabourin, J., Mott, B., Lester, J.: Computational Models of Affect and Empathy for Pedagogical Virtual Agents. Standards in Emotion Modeling, Lorentz Center International Center for Workshops in the Sciences (2011)

[46] Caridakis, G., Karpouzis, K., Wallace, M., Kessous, L., Amir, N.: Multimodal user's affective state analysis in naturalistic interaction. Journal on Multimodal User Interfaces 3(1-2), 49–66 (2009)

[47] Li, X.: Integrating User Affective State Assessment in Enhancing HCI: Review and Proposition. The Open Cybernetics and Systemics Journal 2, 192–205 (2008)

[48] Boersma, P., Weenink, D.: Praat: doing phonetics by computer (version 4.5.15) [computer program] (2007), http://www.praat.org/ (retrieved February 24, 2007)

[49] Martin, B.: Instance-Based learning: Nearest Neighbor With Generalization, Master Thesis, University of Waikato, Hamilton, New Zealand (1995)

[50] Kistler, F., Endrass, B., Damian, I., Dang, C.T., André, E.: Natural interaction with culturally adaptive virtual characters. Journal on Multimodal User Interfaces

Evolutionary Feature Extraction to Infer Behavioral Patterns in Ambient Intelligence

Leila S. Shafti[1], Pablo A. Haya[2], Manuel García-Herranz[3], and Eduardo Pérez[3]

[1] Computer Science and Engineering Dept., Universidad Carlos III de Madrid
lshafti@inf.uc3m.es
[2] Knowledge Engineering Institute, Universidad Autónoma de Madrid
pablo.haya@iic.uam.es
[3] Computer Engineering Dept., Universidad Autónoma de Madrid
{manuel.garciaherranz,eduardo.perez}@uam.es

Abstract. Machine learning methods have been applied to infer activities of users. However, the small number of training samples and their primitive representation often complicates the learning task. In order to correctly infer inhabitant's behavior a long time of observation and data collection is needed. This article suggests the use of MFE3/GA^{DR}, an evolutionary constructive induction method. Constructive induction has been used to improve learning accuracy through transforming the primitive representation of data into a new one where regularities are more apparent. The use of MFE3/GA^{DR} is expected to improve the representation of data and behavior learning process in an intelligent environment. The results of the research show that by applying MFE3/GA^{DR} a standard learner needs considerably less data to correctly infer user's behavior.

Keywords: Intelligent Environments, Behavioral Inference, Machine Learning, Genetic Algorithms, Constructive Induction, Feature Construction.

1 Introduction

A challenging problem in ambient intelligence is to effectively infer activities of people in the intelligent environment. People tend to have a pattern of behavior, often hard to predict [1]. Learning the user's behavior can help to improve the interaction between the ambient intelligence system and the users. With the observation of actions, patterns of user behavior can be learned and future operations can be anticipated. Activity inference has been applied to monitor activities of inhabitants of an intelligent home and detect abnormal behavior of them in order to provide security and health, especially in case of mild cognitive impairment of the inhabitants [2,3]. It has also been used in adaptive smart environments to anticipate the next actions of the users in order to improve the response of the system to their needs and free them from manual control of the environment [4,5].

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 256–271, 2012.

Several works have been conducted to infer user's activities by analyzing audio visual data [6]. However, the use of audio visual devices in an intelligent environment is considered as an intrusion into the private life of the inhabitants. Other systems suggest the use of wearable sensors such as accelerometers and gyroscopes [7,8] or RFID tags and readers [9]. Although these sensors do not violate the privacy, they are not recommended for user comfort and battery power consumption. Considering these problems, the use of low-cost sensors based on infrastructure that detect the interactions of users with environment is suggested for analyzing person's behavior. These sensors are usually located in places where they are not visibly noticeable and, therefore, do not disturb the user. Also, they do not violate the user's privacy.

However, inferring activities using these sensors has three main difficulties. First, data collected from simple sensors are primitive data represented by low-level attributes; therefore, the relevant information is hard to discover. Second, many sensors based on infrastructure are needed to effectively infer interactions of users; thus, the training data is represented by a large number of attributes. Third, since the goal is to learn the user's behavior in the shortest possible time, few training data instances are available for learning. Analyzing such data to learn the behavior of the person is a challenging task. The few training data collected from a huge set of simple low-cost sensors and, therefore, represented by a large number of low-level attributes, suggests the use of techniques such as Constructive Induction (CI) to ease the task of knowledge inference.

CI aims to transform the low-level representation space of data into a new one where the regularities are more apparent to improve learning. This goal is achieved by selecting relevant attributes and removing irrelevant ones (feature selection) as well as constructing more relevant attributes or *features* (feature construction) from original attributes. Feature Selection (FS) and Feature Construction (FC) proved to have an outstanding impact on learning results [10].

MFE3/GA^{DR} is a CI method introduced by Shafti [11,12] that uses an evolutionary algorithm to find relevant attributes and construct functions over them. The advantages of MFE3/GA^{DR} when applied to primitive data with complex relations among attributes have been previously shown [12,13]. This article suggests the use of this method to infer inhabitants' behavior from data collected from simple sensors. Although several machine learning methods have been used for activity inference [14,15,16,17], one of the main weaknesses of these learners is that a large number of training data collected during various days of users observation is needed to effectively infer user's behavior. In this work we show that by applying MFE3/GA^{DR} a standard learner needs considerably less data to correctly infer user's behavior.

The next section explains the problem of primitive data representation and the need for an evolutionary CI method. Section 3 reviews MFE3/GA^{DR}. This article empirically evaluates application of this method to infer behavior of inhabitants. Since collecting data from an intelligent environment is costly and time consuming a data simulator is used for empirical evaluations. The preparation of data with this simulator is explained in Section 4. Sections 5 and 6 report data preprocessing and experiments, and Section 7 discusses the results. Conclusions are summarized in Section 8.

2 Primitive Data Representation and CI

Most learners assume attribute independence and consider attributes one by one. These methods achieve high accuracy when available domain knowledge provides a data representation based on highly informative attributes, as in many of the UCI Databases used to benchmark machine learning algorithms [18]. Since most real-world data are not prepared for machine learning purposes, their representations are often not appropriate for learning target concepts; they are represented by primitive low-level attributes, and several irrelevant and redundant attributes are included. Learning systems face more difficulty when a small set of such low-level primitive data is provided for training. Data collected from sensors of an intelligent environment is an example of such data, as will be seen in experiments (Sections 6 and 7).

The primitive representation facilitates the existence of attribute interactions whose complexity makes the relevant information opaque to most learners. *Interaction* exists among attributes when the relation between one attribute and the target concept is not constant for all values of the other attributes [19]; hence, there appears to be no relation between an individual attribute and the target concept. Considering attributes individually does not help to uncover the underlying patterns that define the concept. When complex interaction exists, the values of all interacting attributes need to be considered simultaneously to predict the concept. Thus, the interaction complexity augments with an increasing number of interacting attributes.

Interaction becomes a stronger hindrance for learners when data contains irrelevant attributes, since interacting attributes can be easily mistaken as irrelevant ones [19]. FS methods have been designed to tackle attribute interaction problem [20,21,22]. These methods consider several attributes together and apply heuristics to distinguish subsets of interacting attributes from subsets of irrelevant attributes. However, when interactions are complex, identifying relevant attributes may not be sufficient for learning. Due to the complex interaction, target concept is dispersed and regularities are less prominent. Therefore, even if interacting attributes are identified correctly, the underlying structure of the concept is still difficult to detect. Interactions need to be highlighted in order to discover underlying complex patterns that define the target concept. This problem worsens when few training data samples are available.

Moreover, when primitive attributes are provided for representing data, the concept description that is to be generated using such attributes tends to be large and complex. Therefore, it is likely that the learner makes mistakes in constructing such description.

FC plays an important role in a CI method when data representation is primitive. FC aims to generate more relevant and predictive attributes derived from the set of primitive attributes to improve the performance of a particular learning system [10]. When FC is applied to concepts with complex interactions, it aims to discover opaque information about the relations between a group of subsets of interacting attributes and the target concept. The discovered information is abstracted and encapsulated into a constructed *feature*. The new feature groups data samples of the same class, which could be scattered in the original data space, and, hence, could not be found similar by simpler methods. An FC method either replaces original attributes with constructed features or adds features as new attributes to the attribute set. If FC finds the appropriate features, such a change of representation makes the instance space highly regular. Each new feature works as an intermediate concept, which forms part of the theory that highlights interactions in primitive data representation. With the new representation a learning system learns the concept easier.

Many progresses have been achieved by FC methods; nevertheless, learners still face serious difficulties to succeed when confronted with complex attribute interactions [13,19]. Most FC methods perform a local search to find interacting attributes one by one. As mentioned earlier, due to complex interaction, each attribute may be considered as an irrelevant attribute when it is evaluated individually. Thus, it becomes necessary to search the space of subsets of attributes. Since the search space of attribute subsets grows exponentially with the number of original attributes and has high variation, a global search such as evolutionary algorithms are preferred for a FC method to find interacting attributes.

Additionally, some FC methods construct and evaluate features one by one. Thus, the construction of each new feature depends on the earlier constructed features. Therefore, if features are constructed incorrectly in the primary steps, the successive features will be irrelevant. This problem worsens when concepts are composed of several complex interactions. Evolutionary algorithms provide the ability to avoid this problem by representing a group of features as one individual in the evolving population and, therefore, constructing and evaluating several features simultaneously. Recent works [23,24] show that a CI method based on an evolutionary algorithm is more likely to be successful in searching through intractable and complicated search space of interacting attributes.

3 Multi-Feature Extraction by GA

MFE3/GA^{DR} introduced by Shafti and Pérez [12] is a preprocessing CI method. It receives as input the training data and the original attribute set. It distinguishes the relevant attributes from irrelevant ones and constructs a set of functions that encapsulate the relations among relevant attributes in order to highlight them to the learner. Constructed functions are then used as the new set of features (new attributes) to replace the original attributes; and data are redescribed. After that, the new representation of data is given to a standard learner to proceed with the machine learning. This section briefly presents MFE3/GA^{DR}. For more details about this method, the reader is referred to [12].

MFE3/GA^{DR} uses a Genetic Algorithm (GA) to find relevant attributes and construct functions over them. Each GA's individual in MFE3/GA^{DR} is a set of attribute subsets represented by bit strings. Each subset is associated with a function defined over attributes in the subset and induced directly from data. Functions are represented in non-algebraic (operator-free) form; that is, no algebraic operator is used for representing them. Each function is a vector of binary values that represent the outcome of the function for each combination of attribute values. Non-algebraic representation has been used successfully for expressing constructed functions by other FC methods [25,26]. This form of representation permits extracting part of the function's description from data and inducing the rest. GA aims to converge the population members toward the set of attribute subsets and their corresponding functions that best represent relations among attributes.

Genetic operators are performed over attribute subsets. However, each attribute subset is associated with a function defined over it; thus, changing a subset in an individual by operators implies changing the associated function. Therefore, genetic operators indirectly evolve functions to construct better ones. The fitness measure evaluates the inconsistency and complexity of constructed functions based on MDL principle [27]. The current version of MFE3/GA^{DR} assumes that the class labels are binary and a discretization preprocessing algorithm, such as [28,29], has been applied to transform all continuous attributes to nominal ones before running the system.

Note that MFE3/GA^{DR} constructs a set of maximum K Boolean features where K is a user defined parameter ($K = 5$ by default). Replacing the original N nominal attributes, each of n_i possible values, with K Boolean features by MFE3/GA^{DR} reduces the representation complexity from $\prod_{i=1}^{N} n_i$ to 2^K, or less since the actual number of constructed features is bound by K. Thus, MFE3/GA^{DR} reduces data size.

MFE3/GA^{DR} successfully captures the relations among attributes and encapsulates them into a set of new features. Since the new features are highly informative, the new data representation makes regularities apparent. Therefore, a learner such as C4.5 can exploit the new representation to easily learn concepts that it could not learn accurately by using only the original representation. Thus, MFE3/GA^{DR} increases the learning accuracy after data reduction. Data collected from sensors of an intelligent environment is an example of such poor data representation. The following sections explain in more detail about these data and how MFE3/GA^{DR} can be applied to improve their representation.

4 Ambient Intelligence Simulated Data for Living Room

Intelligent environments present a complex scenario when sensor data is analyzed. Attributes of real-world data are often redundant (e.g. the light's status and the luminosity sensor value of the room are entangled), or irrelevant to a particular analysis (e.g. outdoor temperature is normally irrelevant when analyzing the state of bathroom lights), since the environment is populated with many primitive low-level sensors and devices. Thus, prediction is a daunting task.

We have been working on Intelligent Environments for the last decade, applying different technologies to set up real spaces designed to study home environments [30,31,32]. While real-spaces in laboratories are good enough to analyze the performance of many applications and perform some basic end-user studies, obtaining large amount of real-world data needs a real space with people living for a long time, requiring large amounts of investment and time. In addition, when talking about home environments, a tryout home results in biased data (e.g. not knowing where things are or not having real preferences over the environment). Using our experience in Intelligent Environments, we have designed a scenario to replicate a plausible small home. It should be noted that our main goal is not to exactly mimic a real setting, but to capture the complexity of the interactions between inhabitants and sensors. The scenario has been designed to consider those special problems of real data, such as redundancy or irrelevant data, as explained in Section 2.

Scenario: The house consists of four rooms (i.e. a kitchen, a bedroom, a toilet and a living room) and three inhabitants (i.e. Tom, Dick and Harry). All rooms are connected through a hallway which has to be crossed to go from one room to another. The living room is equipped with several sensors and devices allowing, respectively, to gather context information and change the world status. Thus, the house is equipped with a temperature sensor, the indoor and outdoor luminosity sensors, and the following devices: a main light, a reading light, a television, a phone, and a blind. In addition, inhabitants wear sensors to identify the room in which they are located.

As it will be shown, there is a direct relation between the sensors and devices and the attributes used in the learning process. These attributes can be classified either as dependent or independent. The former comprises those whose value depends on the value of other attributes and time, meanwhile the value of the latter depends only on time. Table 1 summarizes the attributes used in the living room. The independent attributes of the scenario are $temp$, $loc1$, $loc2$, $loc3$, and $lumOut$, being the dependent ones $tvSt$, $tvVol$, pho, bld, $lumInt$, $lightR$, and $lightC$. The simulation process consists in obtaining the attributes' values during several days. Thus, changes in attributes emerge from changes in sensors and devices and from their relations with other attributes. Note that each attribute has a corresponding *trigger*, indicating whether the status of the attribute at a particular moment (time=t) has changed, compared to the previous moment (time=$t-1$). Therefore, there are in total 25 attributes (12 pairs of sensor values and triggers, and an attribute that represents the time).

The simulation time is measured in seconds. A simulated day starts at zero (0:00:00) second and finishes at the 86399^{th} second (23:59:59), having each day 86400 seconds, that is called SEG_DAY. Every second, all attributes are checked for changes, checking first independent attributes, and then, dependent attributes. A new case is obtained for every change in such a way that each case represents only one attribute change. The occurrence of several cases at the same time is possible. Thus, a different number of cases can be obtained for each simulated second.

While attributes are related to changes in the sensors' values, these values depend on four temporal parameters that must be set up before running a one day simulation: wake up, sleep, sunrise and sunset time. Wake up ($WAKEUP\_TIME$)

Table 1. Attributes Summary

Name	Value	Description
time	integer	time of day in seconds [0.. 86399]
t_temp	binary	trigger of *temp*
temp	continuous	temperature [15..20]
*t_loc*1	binary	trigger of of *loc*1
*loc*1	In/Out	location of the first person (*Tom*)
*t_loc*2	binary	trigger of *loc*2
*loc*2	In/Out	location of the second person (*Dick*)
*t_loc*3	binary	trigger of *loc*3
*loc*3	In/Out	location of the third person (*Harry*)
t_tvSt	binary	trigger of *tvSt*
tvSt	On/Off	TV status
t_tvVol	binary	trigger of *tvVol*
tvVol	continues	TV volume [0..100]
t_pho	binary	trigger of *pho*
pho	HangUp/Ringing/Busy	phone state
t_bld	binary	trigger of *bld*
bld	Up/Down	blinding status
t_lumInt	binary	trigger of *lumInt*
lumInt	discrete	indoor luminosity [*vlow*, *low*, *med*, *high*, *vhigh*]
t_lightR	binary	trigger of *lightR*
lightR	On/Off	reading light status
t_lumOut	binary	trigger of *lumOut*
lumOut	discrete	outdoor luminosity [*vlow*, *low*, *med*, *high*, *vhigh*]
t_lightC	binary	trigger of *lightC*
lightC	On/Off	main light status

and sleep time (*SLEEP_ TIME*) define the daily routine of the inhabitants, while sunrise and sunset times represent the environmental conditions. The simulation can run for as many days as desired, nevertheless, parameters and devices are initialized to their original values every new day.

In the following, we describe how the values of each attribute in the simulation are obtained, and how they are related to the room sensors and devices. The **temperature** attribute (*temp*) reflects the outside temperature and its value is calculated according to a function with maximum (*TEMP_ MAX*) and minimum (*TEMP_ MIN*). The function to obtain the mean temperature is a sinusoidal of time [33]. $MeanTemp(t) = A_0 * \sin(\omega * t - \pi) + TEMP_0$ where A_0 is the peak deviation of the function from its center position, that is, $(TEMP\_MAX - TEMP\_MIN)/2.0$. The angular frequency (ω) specifies how many oscillations occur in a second. It is calculated as $2 * \pi * \mathtt{F}$ where F is the frequency, that is, $1/SEG\_DAY$. Finally, $TEMP_0$ is the mean of the maximum and minimum temperature of the day. Once the mean temperature is obtained, the temperature at this instant is figured using a Gaussian distribution with the calculated mean and a standard deviation of 0.01.

A **Location** attribute (such as *loc*1, *loc*2, *loc*3) is associated to each inhabitant, taking a room identity as a value. A Markov chain based model is used to compute their location in each moment. Every inhabitant has two different Markov chains, one for nighttime and another for daytime. Markov chains are weighted directed graphs where each node represents a location, and edges reflect whether two rooms are connected. The edge weight represents the probability of an inhabitant moving from one room to another; thus, nighttime and daytime chains have the same connections but different transition probabilities. The wake up time ($WAKEUP\_TIME$) and the sleep time ($SLEEP\_TIME$) determine when to use each chain. Each inhabitant has its own transition probabilities. The location attribute is summarized as a binary value (In/Out) representing whether the inhabitant is in the living room or not.

The living room is equipped with a **phone** device. The phone state (*pho*) can take three different values (hanged up, ringing and busy) with four different state transitions: (1) from hanged up to ringing, (2) from ringing to hanged up or (3) to busy, and (4) from busy to hanged up. The transition type (1) represents incoming calls. An exponential distribution is used to obtain the time interval between two transitions of type (1). The mean time between incoming calls is determined as four hours, assuming there are no incoming calls during night. Once the phone is ringing, an exponential distribution is used to compute the time to remain ringing. After that time, the transition (2) occurs if there is at least one inhabitant in the living room to pick up the phone, otherwise transition (3) occurs instead. The phone ringing mean time is ten seconds. The duration of the call is computed as an exponential distribution with five minutes as mean service time.

Television is another device in the room. Two attributes are used to model it: status ($tvSt$) and volume ($tvVol$). Status is a binary value (On/Off) and volume is a real value ranging from zero to 100. Changes in both attributes can only happen if someone is present in the room. The time between every two modifications is computed using an exponential distribution with one hour as mean change time for status and two hours for volume. Status always changes to the opposite value while volume takes a new random value.

The **outdoor luminosity** depends on the time of the day. Represented as a real value, it ranges from $LIGHT\_MIN$ to $LIGHT\_MAX$ [20..900]. The mean luminosity during nighttime is a constant value greater than but close to $LIGHT\_MIN$. During daytime, the mean luminosity varies according to a triangular function, linearly increasing from sunrise time until midday and, then, linearly decreasing until sunset time. The final luminosity value (both for daytime and nighttime) is calculated with a Gaussian distribution according to the mean value and with standard deviation of 15. Finally, the outdoor luminosity attribute ($lumOut$) is figured as a discretization of the previous value in five equal-length intervals (*vlow*, that is, very low, *low*, *medium*, *high*, and *vhigh*, that is, very high).

Three factors affect the indoor luminosity: *blind*, *reading light* and *main light* status. The **blind** (*bld*) is modeled as a binary attribute with two possible values (Up and Down). The blind is opened up at wake time and closed down at sleep time. The **reading light** ($lightR$) has a binary status (On/Off) and its change frequency depends on the number of inhabitants in the living room: the probability of changing the light is 0.03 for each person in the room. The **main light** status attribute is explained further down.

The **indoor luminosity** sensor value is based on a Gaussian distribution. The mean value for the distribution is obtained as follows: if any of the living room lights is turned on, the mean indoor luminosity is set to the maximum. Otherwise, if the blind is up, its mean takes the outdoor luminosity value. In any other case, the mean indoor luminosity is set to the minimum. This behavior has been devised as a simplification of a real experience. The luminosity value is calculated as a real value according to a Gaussian distribution over the previous mean and with a standard deviation of 15. The measured luminosity is assigned to the indoor luminosity attribute ($lumInt$), discretized as one of the five possible luminosity values ($vlow$, that is, very low, low, $medium$, $high$, and $vhigh$, that is, very high).

Finally, the most important behavior is that of the **main light** ($lightC$), which is the one the system tries to predict. The behavior of the inhabitants according to the status of the main light is guided by a simple rule: *if somebody is in the room then turn on the light and if nobody is in the room, turn off the light*. Obviously, in addition to this rule, inhabitants have more preferences. For example they may like to turn off the light when TV is on and enough luminosity is in the room. The aim is to predict the status of the main light after observing the behavior of the inhabitants in the living room and inferring their preferences. In simulated data it is supposed that the inhabitants' behavior is coherent; i.e., they follow a logical rule as follows.

Inhabitants' Preference: the inhabitants' rule regarding the main light ($lightC$) is as follows. If there is a lot of luminosity, i.e, $lumOut = vhigh \vee lumOut = high \vee lumOut = medium$ or $lumOut = low \wedge (tvSt = On \vee lightR = On)$ or $lumOut = vlow \wedge tvSt = On \wedge lightR = On$, then the main light must be off. Otherwise (i.e, there is low luminosity), if there is only one person or nobody in the room turn off the light and if there is more than one person turn on the light.

5 Data Preprocessing

Section 6 empirically shows the advantages of applying MFE3/GA^{DR} to data before inferring user's behavior with C4.5 [34], a classical machine learning system. The reason C4.5 is selected is the interpretability of its results. In an intelligent environment, it is necessary to model the behavior using an interpretable language so that the user can interact with the intelligent system and modify rules. The intelligent environment should be able to easily explain which sensors were the cause of a particular action whenever the user requested it. Thus, a human-friendly form of representing the inferred model such as decision trees and rules is preferred to more powerful models with less interpretable representation such as those obtained by SVMs, Artificial Neural Networks, Hidden Markov Models and other similar methods. Simulated data described in Section 4 are used for these experiments. Some data preprocessing is needed in order to prepare data for experiments.

Since C4.5 is a supervised learner, samples in training data must be classified and labeled before learning. Thus, in order to predict the status of the main light (*lightC*), it is necessary to assign a class label to each sample indicating whether the light should be on or off with the current conditions in the room. Considering that samples of each day are stamped and ordered by time, the class label assigned to each sample is the value of attribute *lightC* in the following sample.

MFE3/GA^{DR} accepts only discrete attribute values. Thus, continuous attributes, *temp* and $tvVol$, should be discretized. A simple approach is applied for these experiments. The range of each attribute value is divided into five equal-width intervals and a discrete numeric value is associated with each interval. Attribute *time* which is an incremental value from zero to $(60*60*24)-1 = 86399$ is not used for learning.

One aspect in the training data that makes the learning process more difficult is that the ratio of the number of negative cases to the total number of cases in data is significantly high (approximately 98.5% on average). This causes a learning system to incorrectly classify all cases as negative and achieve 98.5% accuracy. To reduce this problem some negative cases which are not relevant to the learning process should be removed to decrease the majority class percentage. First, for this purpose, all cases in which nothing changes in the room are removed (that is, when all triggers are zero). It is supposed that if nothing in the room changes, there is no reason to change the status of the main light. Thus, these cases are not needed for the learning process. Second, it is established that the learning process for each day starts when the first person enters the living room during the day and finishes at the end of the day. This is reasonable because when nobody has entered the room, no activity is performed; and therefore, the status of the light does not change. Thus, cases belonging to the first period of each day, when nobody has entered the room, are removed. These two modifications in the data helped to reduce the number of negative cases (when *lightC* is off) in the data and, therefore, reduce the majority class percentage to 89.9% on average.

6 CI and Learning Experiments

Although C4.5 is preferred to other machine learning methods due to readability of its outcome, the application of this method alone is not enough for accurately inferring the user's behavior. MFE3/GA^{DR} is designed to improve C4.5 when applied to complex concepts with few training data and interactions among attributes [12] (see Section 2). These experiments aim to highlight that such a CI method will be useful in the context of learning behaviors from a few data samples collected from sensors, since the data has a primitive representation with interactions among attributes.

For experiments, 10 sets of training data are generated by simulator corresponding to 10 days of learning. For evaluating the learning results, a set of data

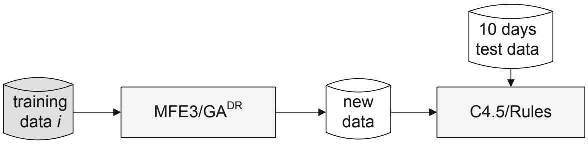


Fig. 1. An experiment's process: data i corresponds to i days of training, $i = [1..10]$

Table 2. Average error over 20 experiments

# days	# train.	C4.5	C4.5Rules	MFE3/GA^{DR}+C4.5
1	815	**362.9(4.41%)**	**293.2(3.56%)**	121.6(1.48%)
2	1623	**197.8(2.41%)**	**125.8(1.53%)**	5.2(0.06%)
3	2449	**112.4(1.37%)**	**33.5(0.41%)**	7.0(0.09%)
4	3266	**60.1(0.73%)**	6.1(0.07%)	0.0(0.00%)
5	4094	**44.9(0.55%)**	**9.3(0.11%)**	0.0(0.00%)
6	4901	**18.3(0.22%)**	1.7(0.02%)	0.7(0.01%)
7	5714	**14.6(0.18%)**	1.5(0.02%)	0.0(0.00%)
8	6528	3.7(0.05%)	0.0(0.00%)	0.0(0.00%)
9	7342	2.8(0.03%)	0.0(0.00%)	0.0(0.00%)
10	8140	0.0(0.00%)	0.0(0.00%)	0.0(0.00%)

corresponding to another 10 days is generated and kept unseen to be used as test data. The aim is to learn the concept after each day using training data and measure the accuracy of the inferred concept after each day using the set of test data. Thus, 10 experiments are performed using one to 10 sets of training data for learning (see Fig. 1). For each experiment, MFE3/GA^{DR} is used to extract new features; the original attributes are replaced with new features; and then, training and test data are redescribed. Next, new sets of data and attributes are given to C4.5 as a learning system for inferring the rules. The result is tested using unseen test data. The whole process of 10 days of training and testing is repeated 20 times over 20 independently generated sets of training and test data and the average accuracies are calculated. The results are compared to the average accuracies of C4.5 and C4.5Rules [34] using original data. All methods are used with their default parameters as a black-box.

Table 2 summarizes the results. The first column shows the number of days used for training. The second column gives the average training data size (over 20 sets) after each day of collecting data. The average number of test data samples corresponding to 10 days is 8193. The results show the average number of misclassified cases (errors) and the percentage of error. Results of C4.5 and C4.5Rules with original data are reported in columns three and four. Column five shows results of C4.5 after constructing features with MFE3/GA^{DR}. The same results are obtained using C4.5Rules after FC and, therefore, are not reported in the table. A Bold means that with a significant level of 0.05 the result obtained after FC (last column) is better than this result, using t-distribution test.

It can be seen from Table 2 that C4.5 needs data collected during eight days to learn the behavior of users with almost no error, using original data set. With less data, due to poor representation, there is not enough information to see the regularities. But when MFE3/GA^{DR} is used to change the representation, the learning accuracy of C4.5 is significantly improved even when few training samples are used for learning. C4.5Rules requires slightly fewer data samples than C4.5 for training using original data set. It needs data collected during at least six days to approximate the number of errors to zero. However, MFE3/GA^{DR} still significantly outperforms C4.5Rules. With four days or more of data collection, MFE3/GA^{DR} constructs features that perfectly highlight information and improve the representation of data; therefore, after changing the data representation, C4.5 produces no error on test data (except in case of training in six days where a small error is produced). The learning process of MFE3/GA^{DR}+C4.5 takes 102 seconds on average using four days of training data on a pc with 3.33GHz Dual CPU and 3GB memory running on Ubuntu 9.1 operating system.

7 Discussion

It is interesting to analyze features constructed by MFE3/GA^{DR} to see how this method improves learning. In all experiments MFE3/GA^{DR} successfully detects relevant attributes. However, when one to four days are used for FC, constructed features are often not good enough to represent the whole concept due to lack of information in training data. When more training data is provided, more predictive features are constructed. If data related to four or more days are used, MFE3/GA^{DR} often replaces 24 original attributes with two features that perfectly highlight all the interactions among relevant attributes to the learner. One of the two features is defined over *loc*1, *loc*2, and *loc*3, and indicates whether there is more than one person in the living room. The other one is a function defined over *tvSt*, *lightR* and *lumOut* and is true when TV or reading light are off and the outside luminosity is very low (i.e., very low or no luminosity) or when both TV and reading light are off and outside luminosity is low (i.e., low luminosity). After representing data with the new features, C4.5 generates a decision tree representing the conjunction of two features; that is, if both attributes are true then turn on the light; otherwise turn off the light. MFE3/GA^{DR} generates two functions that successfully encapsulate and highlight regularities in the target concept that were difficult to uncover directly from original set of attributes.

When data of two or three days are used for FC, this method constructs either one or two features over relevant attributes. Since training data size is small, the constructed features do not represent exactly all the relations; and therefore, C4.5 after FC produces some error. When data of the first day is used for FC, this method often replaces original attributes with three features. One of the features is a function defined over attributes *loc*1, *loc*2, and *loc*3, which indicates whether there is more than one person in the room. The other one is defined over *tvSt* and *lightR*, and is false when both TV and reading light are on, otherwise it is true. The last one is defined over *tvSt* and *lumOut* and is true when *lumOut* is off. Each function represents part of the concept. However,

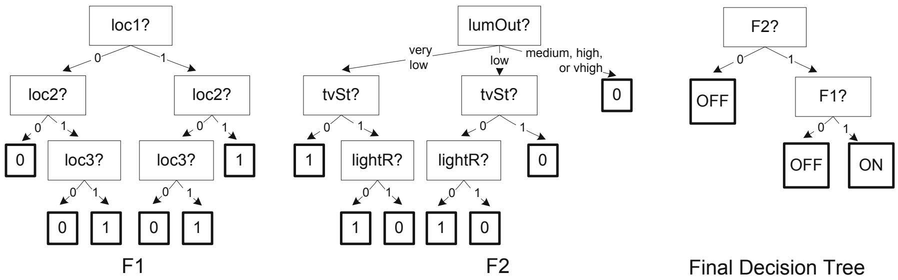


Fig. 2. Decision tree representation of constructed features after four days of training and the final decision tree generated by C4.5 using the new features

they are not good enough to represent the complex underlying structure of the whole concept.

It is interesting to analyze the classifiers generated by C4.5 and C4.5Rules with original data representation to see why they fail to learn the concept correctly. In the case of few training data the main reason for their failure is that they select attributes incorrectly. As mentioned in Section 2, because of the interactions among attributes and few training data, these learners confound irrelevant attributes with interacting ones and, consequently, construct an imperfect classifier. With four to seven training days they select relevant attributes correctly. However, the underlying concept is still complex and hard to generate even with seven days of training data in case of C4.5 and five days of training data in case of C4.5Rules. The classifiers generated by these learners are an approximation to the target concept and, therefore, produce error. This demonstrates that the underlying concept is complex and cannot be discovered using relevant attributes. Highlighting interactions among attributes in this concept is necessary to improve learning and produce no error.

With more than seven days of training data, C4.5 and C4.5Rules produce almost no error. However, the classifiers generated by these learners are complex and hard to interpret by users of the intelligent system (a decision tree of 44 nodes in case of C4.5 and a set of 18 rules in case of C4.5Rules). Since inhabitants of the intelligent environment need to have access to the inferred classifier and modify it in case that it is necessary, a simpler classifier is preferred. MFE3/GA^{DR} generates a pair of features, each encapsulating a part of the user's behavior. Each feature represented by non-algebraic form (truth table) can be easily transformed to a decision tree that is small and interpretable. This process is done by passing the truth table as training data to C4.5 to obtain the corresponding decision tree. Fig. 2 shows decision trees corresponding to the features. The final decision tree generated by C4.5 after FC represents a conjunction of these two features which is considerably small and simple. Decomposing the complex rule into smaller parts and representing them by smaller decision trees facilitates the ability to express the rule in a more user friendly language. This process is necessary as the first step towards communicating with inhabitants of the environment. However, more study is going on to convert the decision trees into a language closer to the natural language and represent them through an interface for further modifications by an inexpert inhabitants.

8 Conclusion

The problem of primitive data representation of intelligent environments is highlighted in this article. When complex interactions exist among attributes and the only information provided about the concept is a few primitive training data, regularities are opaque to the learner. Then a CI method is needed to construct features that abstract and encapsulate such occulted information into new features in order to highlight it. Each constructed feature works as an intermediate concept, which forms part of the theory that highlights interactions in primitive data representation.

MFE3/GA^{DR}, as an evolutionary CI method, is shown to improve the low-level primitive data representation of concepts. An application of the method to an intelligent environment to infer the inhabitants' behavior is presented in this article. Data collected from sensors that register interactions of inhabitants with the environment has a low-level representation. Thus, the concept is hard to be learned directly from such representation.

Experiments show that the CI method, MFE3/GA^{DR}, with the help of a genetic algorithm successfully detects interacting attributes and constructs highly informative features that abstract interactions. Thus, it significantly improves learning while it reduces the data size. It was observed that with this method fewer data samples are needed in order to produce no error comparing to the results obtained without CI. After one day of registering the inhabitants' behavior, this method transforms the data representation into a new one where regularities are more apparent and, therefore, reduces the error rate. After four days of training, it perfectly abstracts and encapsulates regularities into few new features. Thus a learning system can easily learn the behavior of the users with no errors.

Obtaining real data is costly and time consuming. One alternative is to use a public data set. Several data sets are available [35] that include sensor data gathered from real environments. However, an external real setting involves dealing with unknown sources of bias, which can be overcame using data annotated by ourselves. This is particularly important in our case as we require to know in advance not only the users' behavior but also their preferences so that we can use them to assess the inferred rules. Thus, we preferred using a simulation as it provides a controlled framework where it is easier to validate the results. As for a future work, MFE3/GA^{DR} will be evaluated with data collected from a real environment as well.

Acknowledgments. This work is partially funded by the Spanish Government (projects GO-LITE, TIN2011-24139, and ASIES, TIN2010-17344).

References

1. Philipose, M., Fishkin, K., Perkowitz, M., Patterson, D., Fox, D., Kautz, H., Hähnel, D.: Inferring activities from interactions with objects. IEEE Pervasive Computing, 50–57 (2004)
2. Campo, E., Chan, M.: Detecting abnormal behavior by real-time monitoring of patients. In: AAAI Workshop on Automation as Caregiver, pp. 8–12. AAAI Press (2002)
3. Tapia, E., Intille, S., Larson, K.: Activity Recognition in the Home Using Simple and Ubiquitous Sensors. In: Ferscha, A., Mattern, F. (eds.) PERVASIVE 2004. LNCS, vol. 3001, pp. 158–175. Springer, Heidelberg (2004)
4. Mozer, M.: An intelligent environment must be adaptive. Intelligent Systems and Their Applications 14, 11–13 (1999)
5. Das, S., Cook, D., Battacharya, A., Heierman, L.T.Y.: The role of prediction algorithms in the mavhome smart home architecture. IEEE in Wireless Communications 9, 77–84 (2002)
6. Aggrawal, J., Cai, Q.: Human motion analysis: A review. Computer Vision and Image Understanding 73(3), 428–440 (1999)
7. Bourke, A., Lyons, G.: A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor. Medical Engineering and Physics 30(1), 84–90 (2008)
8. Ravi, N., Dandekar, N., Mysore, P., Littman, M.: Activity recognition from accelerometer data. American Association for Artificial Intelligence 3, 1541–1546 (2005)
9. Huang, P.C., Lee, S.S., Kuo, Y.H., Lee, K.R.: A flexible sequence alignment approach on pattern mining and matching for human activity recognition. Expert Systems with Applications 37, 298–306 (2010)
10. Liu, H., Motoda, H. (eds.): Feature Extraction, Construction and Selection: A Data Mining Perspective. The International Series in Engineering and Computer Science, SECS, vol. 453. Kluwer Academic Publishers, Norwell (1998)
11. Shafti, L.S.: Multi-feature Construction based on Genetic Algorithms and Non-algebraic Feature Representation to Facilitate Learning Concepts with Complex Interactions. PhD thesis, Escuela Politecnica Superior, Universidad Autonoma de Madrid (2008)
12. Shafti, L.S., Pérez, E.: Evolutionary multi-feature construction for data reduction: A case study. Journal of Applied Soft Computing 9(4), 1296–1303 (2009a)
13. Shafti, L.S., Pérez, E.: Feature Construction and Feature Selection in Presence of Attribute Interactions. In: Corchado, E., Wu, X., Oja, E., Herrero, Á., Baruque, B. (eds.) HAIS 2009. LNCS, vol. 5572, pp. 589–596. Springer, Heidelberg (2009)
14. Logan, B., Healey, J., Philipose, M., Tapia, E.M., Intille, S.: A Long-Term Evaluation of Sensing Modalities for Activity Recognition. In: Krumm, J., Abowd, G.D., Seneviratne, A., Strang, T. (eds.) UbiComp 2007. LNCS, vol. 4717, pp. 483–500. Springer, Heidelberg (2007)
15. Guan, D., Yuan, W., Lee, Y., Gavrilov, A., Lee, S.: Activity recognition based on semi-supervised learning. In: The 13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications, pp. 469–475. IEEE Computer Society (2007)
16. Lustrek, M., Kaluza, B.: Fall detection and activity recognition with machine learning. Informatica 33, 205–212 (2009)
17. van Kasteren, T.L.M., Englebienne, G., Kröse, B.J.A.: An activity monitoring system for elderly care using generative and discriminative models. Personal and Ubiquitous Computing 14, 489–498 (2010)

18. Blake, C., Merz, C.: UCI repository of machine learning databases (1998)
19. Freitas, A.A.: Understanding the crucial role of attribute interaction in data mining. Artificial Intelligence Review 16(3), 177–199 (2001)
20. Saeys, Y., Inza, I., Larranaga, P.: A review of feature selection techniques in bioinformatics. Bioinformatics 23(19), 2507–2517 (2007)
21. Liu, H., Motoda, H. (eds.): Computational Methods of Feature Selection. Data Mining and Knowledge Discovery Series. Chapman and Hall/CRC, New York (2007)
22. Zhao, Z., Liu, H.: Searching for interacting features in subset selection. Intelligent Data Analysis 13, 207–228 (2009)
23. Estébanez, C., Valls, J.M., Aler, R.: GPPE: a method to generate ad-hoc feature extractors for prediction in financial domains. Applied Intelligence 29(2), 174–185 (2008)
24. Zhang, Y., Rockett, P.I.: Domain-independent feature extraction for multi-classification using multi-objective genetic programming. Pattern Analysis and Applications 13 (2009)
25. Zupan, B., Bratko, I., Bohanec, M., Demsar, J.: Function Decomposition in Machine Learning. In: Paliouras, G., Karkaletsis, V., Spyropoulos, C.D. (eds.) ACAI 1999. LNCS (LNAI), vol. 2049, pp. 71–101. Springer, Heidelberg (2001)
26. Alfred, R.: DARA: Data summarisation with feature construction. In: Asia International Conference on Modelling and Simulation, pp. 830–835. IEEE Computer Society, Kuala Lumpur (2008)
27. Grunwald, P.D.: The Minimum Description Length Principle. MIT Press (2007)
28. Boulle, M.: Khiops: A statistical discretization method of continuous attributes. Machine Learning 55, 53–69 (2004)
29. Kurgan, L.A., Cios, K.J.: CAIM discretization algorithm. IEEE Transactions on Knowledge and Data Engineering 16(2), 145–153 (2004)
30. García-Herranz, M., Haya, P.A., Montoro, G., Esquivel, A., Alamán, X.: Easing the smart home: Semi-automatic adaptation in perceptive environments. Journal of Universal Computer Science 14, 1529–1544 (2008)
31. Esquivel, A., Haya, P.A., García-Herranz, M., Alamán, X.: Managing Pervasive Environment Privacy Using the "fair trade" Metaphor. In: Meersman, R., Tari, Z. (eds.) OTM-WS 2007, Part II. LNCS, vol. 4806, pp. 804–813. Springer, Heidelberg (2007)
32. Haya, P.A., Montoro, G., Alamán, X.: A Prototype of a Context-Based Architecture for Intelligent Home Environments. In: Meersman, R., Tari, Z. (eds.) CoopIS/DOA/ODBASE 2004. LNCS, vol. 3290, pp. 477–491. Springer, Heidelberg (2004)
33. Pidwirny, M.: Daily and annual cycles of temperature, fundamentals of physical geography, 2nd edn. (2006), http://www.physicalgeography.net/fundamentals/7l.html
34. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)
35. BoxLab Wiki Page, http://boxlab.wikispaces.com/

Personalization of Content on Public Displays Driven by the Recognition of Group Context

Ekaterina Kurdyukova, Stephan Hammer, and Elisabeth André

University of Augsburg, Universitaetsstrasse 6a,
86159 Augsburg, Germany
{kurdyukova,hammer,andre}@hcm-lab.de

Abstract. Personalization of content on public displays relies on the knowledge of spectator interests and real-time recognition of social context. In busy public places, with numerous individuals circulating daily, the knowledge of individual interests becomes unrealistic. This paper presents an approach for automatic personalization which, instead of individual profiles, relies on *group context*. The system recognizes the constellation of spectators in front of a public display, based on their disposition and gender. Thus, the approach provides an important prerequisite for a completely automated personalization, requiring no input from the spectator side, neither for training, nor for real-time content adaptation. The experiment conducted in a public area showed that the presented approach can successfully identify the differences in the content observation of various groups. Moreover, the approach provides an insight into the diversity of circulating groups, and gives a hint about spectators' emotional and conversational response to the content.

Keywords: Public displays, Adaptation, Social Context.

1 Motivation

Personalized content on public displays offers clear advantages: the users get direct access to the information of their interest. A challenging task, however, is to *learn the interests* of the users and to *offer matching content* in real time. These tasks become even more complex if the displays are installed in busy public places, where numerous individuals circulate every day. Such displays have to learn the interests of a huge amount of users; moreover, they need to combine the interests of distinct individuals, when a group of several individuals observes the content.

Another challenge relates to the implicit nature of the personalization mechanism: the content adaptation should happen automatically, without any input from the user side. Indeed, a manual input, such as activation of an online profile or switching on a mobile client, is hardly acceptable in a busy public place. People have no time and attention for the manual input; moreover, they may be just unaware of the input possibility.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 272–287, 2012.

Therefore, there is a need in a mechanism that would *learn the interests* of the spectators and *adapt the content* completely *automatically*. Such a mechanism should distinguish between *different profiles* of spectators, however, requiring *no explicit input* from spectator side.

The system proposed in this paper meets the described requirements. In order to illustrate the concept, imagine the following scenario. A busy train station is equipped with a large public display. The display provides incoming passengers with city-related advertisement and useful tips. When a mother with two kids passes by the display, the screen advertises leisure activities for families. When a single lady passes by, the display shows a trendy perfume shop. For a passing-by couple, the display shows tips on romantic cafés. Finally, for a group of teenage boys, the display advertises an adventure attractions park.

The learning of spectators' interests and the real-time adaptation are enabled by the recognition of *group context*. By means of a camera mounted on the display, the system scans the composition of the observing groups: the number of individuals, their mutual disposition, and gender. Additionally, the system registers positive emotions of the individuals and whether they have a conversation.

In the learning phase the system tracks visual attention of the groups and relates it to the popularity of the content categories. During the real-time adaptation, the system recognizes the group standing in front of the display and shows the content which has the highest popularity within the given group.

In the current work we focus on the *learning phase*. By means of an experiment we demonstrate how the system can be used to identify differences in the visual attention (interest) of different groups. Important to mention, the notion of interest in this work equals to a spectator's "visual interest". We consider a person interested in the content, if he/she spends some time observing the content. Such an approach eliminates the confusion with person's intrinsic interest or commercial interest. Although the question of a person's real interest versus "visual interest" is definitely an important point, it is out of the scope of this work. The methods to measure audience's interest can be found in the related literature [1]. An argument against the "visual interest" can be the necessity of people to stand in front of the display, for instance, while waiting. The frontal orientation of the observing face, however, does imply the visual interest of the observer who keeps his or her face oriented to the content.

After an overview of related research, we describe the personalization mechanism in detail. In order to prove the usefulness of the system, we present the experimental deployment conducted in a public area of a university. Although this public space is definitely less busy than, for example, a train station, it does represent a valid public space with active and irregular circulation of diverse individuals. Therefore, this space is suitable to prove the performance of our system. The results of the experimental deployment show that the system can be successfully used to tag the content according to the group-based observation patterns (visual interest). Moreover, it provides interesting insights into the diversity of circulating groups and their emotional and conversational activity while observing the presented content.

2 Personalization of Content

In order to provide personalized content in real-time, a system should (1) learn the interests of the spectators and (2) be able to present the right content in the real-time according to the learnt interests. The first task is usually achieved by *tagging*. The second task refers to *real-time adaptation*.

2.1 Tagging

By means of tagging, the content elements are labeled with the spectators' interests. The interests can be retrieved in an explicit or implicit way.

For *explicit tagging*, a sample of potential users is asked to rank the presented content manually [2]. Such a tagging is usually done in laboratory conditions. The users rate each element of the content according to a given schema, e.g. a linear scale. Although the method delivers precise results, these results might not reflect the user interests. Since the lab setting is not natural, the ranking may deviate from the preferences the users would express in a real setting. Moreover, explicit ranking requires a significant effort from the user. The tiredness caused by the ranking routine, therefore, may also impact the number of provided ratings and the reliability of the results [2, 3].

Explicit tagging is barely applicable for the content of displays installed in a public place. The ranking reflects the interests only of some distinct individuals. Therefore, it can significantly deviate from the interests of the numerous other individuals circulating in the crowd. Another disadvantage refers to the extraction of group interests. Complex algorithms must be applied to derive the group interests from the interests of distinct individuals [4]. The complexity grows in busy public places where the group compositions are usually very diverse.

Implicit tagging is a more adequate ranking approach for public places. The method usually exploits crowd monitoring. An illustration of implicit tagging can be found in the work by Müller and colleagues [5]. The method counts the number of frontal faces registered for every content element (e.g. a slide). The element which has accumulated the largest number of the faces is considered to be of the highest interest. In the subsequent real-time adaptation process, the "most interesting" element will be set to the highest priority in the content schedule.

The approach replicates the real behaviour of the spectators: people look at the content when they are interested in it. However, the approach is not flexible enough to distinguish between groups of spectators. What if one content element was observed only by numerous single persons, and never – by couples? The approach, however, will prioritize the content element for both groups: singles and couples.

Another disadvantage of the method is the assumption that the frontal look equals interest. Although visual attention is an important hint to derive interest, it is not sufficient. In fact, the spectators can look at the display for many other reasons [6, 7]. For instance, the display is oriented frontally to the spectators' path or the colours of the content subconsciously attract attention. Such effects do not necessarily imply interest. Therefore, for automatic personalization more contextual cues should be used to support the assumption of interest.

To summarize, implicit tagging is the most suitable approach for public spaces. Such tagging requires no user input and reflects the natural setting. The tagging mechanism must distinguish the compositions of various spectator groups.

2.2 Real-Time Adaptation

The second part of the personalization process refers to the real-time content adaptation. The adaptation is based on the results obtained by the tagging. The real-time adaptation can rely on user *contributions* or work completely *automatically*.

The *contributions*-based approach requires a certain registration from the user side. For instance, the spectators switch on their Bluetooth devices and transmit the pre-set profiles [8, 9, 10]. The display receives the profiles and adjusts the content according to the profile interests. By means of dedicated strategies [11] the group interests can be derived from the individual profiles [12]. Although the contributions-based approach provides a precise overview about the present spectators, it is hardly applicable in a crowded public place. The numerous visitors of the public place may not possess the required devices. And even those who possess them may simply forget to switch the device on. As a result, the display will retrieve an incomplete or a wrong picture about the surrounding spectators.

The alternative *automatic* approach can utilize the identity of the user. For example, by means of face recognition the system can understand who stays at the display, and thus automatically adjust the content [13]. Although such an approach does not require any user contribution, it has to carefully learn the user profiles in advance. This requirement is not realistic in a busy public place with numerous individuals. Müller and colleagues [5] used face detection as a trigger for real-time adaptation. Once a face is detected in front of the display, the most popular content appears on the screen. The method is more suitable for public places than identification: it eliminates the unrealistic knowledge of each single individual. However, it does not distinguish between the interests of various individuals. For instance, the content popular among women does not necessarily match the content popular among men. Moreover, the method cannot does not take into account the composition of the spectator groups. For example, the same content will be displayed for two teenage girls, a couple, a group of elderly men or a mother with three kids.

All in all, it is a challenging task to recognize the interests of spectators and provide the right content in the real-time. Ideally, the system should know the individual profiles of the spectators, which is difficult to realize in a busy public place. Moreover, the system should work automatically, not requiring any input from the spectator side. A trade-off would be an approach which approximates to the individual profiles, but does not require the spectators to explicitly provide the system with their profiles.

In this paper we present a group-based personalization approach, focusing on the tagging phase. Based on the gender and disposition of detected spectators, the system classifies the spectators into distinct groups, for instance, a couple, two men or a single woman. Thus, the system registers the interest not of distinct individuals, but of distinct groups. Since the detection of the spectators and the recognition of their gender are done with a camera, the approach works fully automatically eliminating any user input.

3 Approach for Group-Based Personalization

Based on the analysis of the existing approaches for tagging and real-time adaptation, we came up with a set of requirements for personalization systems in busy public places:

- the system must work **completely automatically**, not expecting any contribution from the user (spectator) side.
- the system must **capture the group composition**, instead of identities of distinct individuals. The unique groups should be defined by the characteristic features of the group members, e.g. gender or age.
- the system must be **robust against the traffic of a busy public place**, taking into account the diversity of passing-by individuals and possible groups.

The system presented in this paper meets the requirements of automatic personalization. Below we describe how it can be used for automatic tagging and real-time adaptation.

Completely Automatic System

Content tagging and subsequent adaptation are based on the monitoring of spectators. A camera installed on the top of the display detects the faces in the proximity of the display (see Fig.1); the microphone detects the noise level.

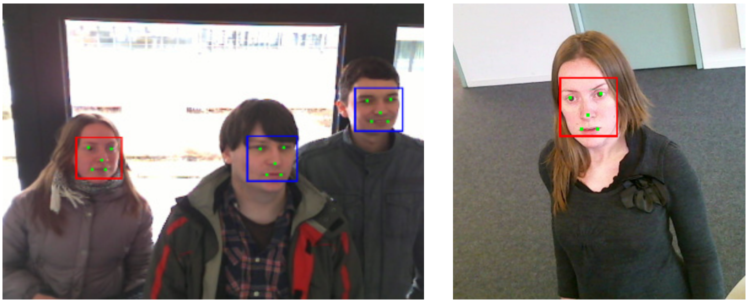


Fig. 1. Recognition of the group structure: red and blue rectangular indicate female or male faces

The system runs face detection algorithm enabled by the SHORE [14] and SSI [15] software. SHORE provides robust face detection and gender recognition. For each detected face, it delivers the following parameters:

Gender. The SHORE software delivers a probability percentage for the face being male or female. Our system uses the threshold of 80% to accept SHORE's decision on gender.

Position. The SHORE software provides coordinates of outlining rectangle for each detected face. From the size of the rectangle and its position on the x-axis we can infer the spectators' proximity to the display, location, and mutual disposition.

Emotional state. The SHORE software also provides emotion recognition, relying on facial expressions[1]. The software delivers percentage on probability for four emotions: “happy”, “angry”, “surprised”, and “sad”. If the probability is lower than a predefined threshold, the emotional state is registered as “neutral”. If the probability of an emotion is higher than the set threshold (we used 70%), the face is registered as expressing the given emotion.

Besides information on detected faces, the tagging system registers whether spectators have a *conversation*. A microphone integrated into the camera registers the volume level when spectators are present in front of the screen. The volume level is classified into three ranges: silent (almost no sound), moderate (moderate discussion between several individuals), and loud (active discussion).

The SSI software helps synchronize the content slide show with the data delivered by SHORE. The SHORE and SSI software run in the background, leaving in the foreground solely the content. Thus, the system runs the group recognition in a complete automatic way. It neither requires any input from the spectators, nor does it disturb the observation process.

Capturing the Group Composition

Having the data on present faces, their gender, and disposition, the system can conclude about the group composition. The spatial disposition of the faces enables us to determine whether the present spectators belong to the same group or are standing alone individuals.

Robust against the Traffic of a Public Place

Generally, the number of SHORE-detected faces is not limited. The software, however, allows defining a minimal size of the face outline as a percent of the entire camera field. We set the smallest face to be 2,5% of the field covered by the camera. Smaller faces refer to distant persons who cannot see the content properly; thus, they are not considered as valid spectators.

The recognition of spectator’s interest can be supported by additional cues, such as positive emotions or discussion of the content. Our personalization mechanism registers emotions of each group member, as well as the volume level of the conversation. However, the reliability of these additional cues has to be proven experimentally. Emotions and conversations are not necessarily caused by the display content. Therefore, we consider emotional and conversational response as a secondary hint to the spectator’s interest.

Speaking about emotional response, it is important to mention that the interest does not always imply positive emotions. For instance, a person can be highly concentrated on the content (thus, interested), but have a neutral facial expression. A positive emotion therefore is not equivalent to “relevant”, but is a contextual condition that influences how “relevant” the content is.

---

[1] The details on implementation are given by the SHORE authors in the related literature [14].

4 Accuracy of Recognition

Before the deployment of the system, we tested the accuracy of face detection and emotion recognition delivered by the SHORE. Important to emphasize, our goal was not to verify the accuracy of the SHORE algorithms. This question has been elaborated by the authors of SHORE and can be found in the related literature [14]. Our goal was solely to test how accurate the SHORE software performs in our experimental conditions.

The test was conducted in the public area of the university, where the main experiment took place, employing the same displays as in the main experiment. For the test we presented some arbitrary photos on the displays. All spectators were video-recorded; a note informed them about the recording fact. Simultaneously with the video recording, the tagging system was running in the background. It logged the detected faces, their gender, and emotions.

In total, we collected 16 hours of video, containing 128 female and 120 male faces. The video material was manually annotated, registering the recorded number of female and male faces, group constellation, and emotions (based on subjectively estimated facial expressions). The annotation was compared with the log data, yielding the accuracy of recognition.

About 95% of all the faces of people standing more than 1.5 seconds in front of the display were captured by SHORE. The faces further than the specified observation distance (face rectangle covering less than 2.5% of the camera view) as well as the faces of passers-by who just glanced at the display without stopping were not detected by SHORE as faces. This limitation is in line with our definition of spectators: people within a close proximity to the screen, who do stop to watch the content.

Gender Recognition of SHORE showed accuracy rate of 96% for males, and 92% for females. The system needed about 0.3 seconds on average to decide on the gender.

Emotion Recognition of SHORE, i.e. for “happy” emotion, showed an accuracy rate of 90% for male spectators and 92% for female spectators. The recognition of “surprised”, “sad”, and “angry” emotions showed less reliable results, yielding only 60-65% of accuracy. Moreover, these emotions were recognized on the video rarely: usually, spectators either smiled or didn’t express any emotions.

To summarize, gender recognition with SHORE can be reliably used for the group-based personalization. As for the emotion recognition with SHORE, we can reliably apply only the recognition of the classes “happy” vs. “not happy” emotion (in other words, “smile” vs. “no-smile”).

5 Experimental Deployment

The goal of the experimental deployment was to see how well the system can be applied for group-based personalization. In particular, we aimed to answer the following questions:

- Can the system identify differences in observation patterns (content preferences) of distinct spectator groups?
- What insights into spectator groups can be gained with the system?
- Does information on emotional and conversational response deliver reliable hints on spectator interests?

5.1 Deployment Set-Up

The system was deployed during three weeks on the displays in a university public area. Three displays were involved in the deployment: one display situated in a lobby and two in a passage (see Fig. 2). All displays have non-touchable screens of 62 inches and 45 inches in diagonal.

Fig. 2. The lobby display (left) and the passage displays (right) used in the experiment

The circulation of people on the premises of the university is moderate. Besides the main "inhabitants", consisting of about 30 researchers, the experiment area is used by students and visitors. The passage area is often used as a short cut to the university canteen, the parking lot or other places within the university. During the experimental weeks two events took place at the area adjacent to the experiment public place; bringing in total about 30 visitors from outside the university.

The aim of the personalization system was to tag the content newly created for the university displays. The content was compiled in a slide show; the personalization system ran on the background. The content topics were proposed by researchers of the university. Within a brainstorming session, the researchers came up with four content categories: "Team", presenting the members of the research team, "News", informing about recent info, e.g. upcoming events or lectures, "Department Life", presenting events of research unit, "Quiz", posting a tricky question about a research unit, followed by the correct answer. The researchers found these categories relevant for the university life. However, we needed to find out whether the content would also attract our students and visitors.

The design of the content was kept consistent, in order to exclude distractions caused by visual design (see Fig. 3). Each content slide stayed on the screen for 10 seconds.

Fig. 3. Examples of the content: “Department Life” and “Quiz”

5.2 Tagging Procedure

In order to tag the content categories according to the group interests, the tagging algorithm was launched in the background of the slide show. Each display was supplied by a camera with an integrated microphone. The cameras were installed on the top side of the display frame. The SHORE software was processing the images captured by the camera. With the frequency of 15 frames per second, SHORE delivered information on each detected face: gender estimation, coordinates of outlining rectangular, and emotion estimation.

This data was processed to make an entry to the log file. Based on the number of detected faces and the gender data, the **<group composition>** was calculated. Based on the coordinates of the rectangular, we calculated **<position>** of each group member. Position reflected the user location at the display (left, centre, right) and the proximity to the display (near, middle, far). From this information we could estimate whether the spectators belong to the same group (stand next to each other) or are several distinct individuals. Based on the probability of each emotion, we registered the resulting **<emotion>**. The emotion having probability more than 70% was entered to the log. Finally, the microphone provided data on estimated **<volume level>**.

As a result, a log entry consisted of the general description of the social context and the detailed description of each face:

<timestamp> <group composition> <volume level>
<face1><gender><position><emotion>
<face2><gender><position><emotion>

An entry was added to the log every time the social context was changing, for instance, people joined the group, people left, or emotional context changed. The following lines illustrate an example of a log entry (F stands for female, M stands for male):

<15:29:00> <2F + 1M> <loud>
<face1><F><left near><neutral>
<face2><F><left near><neutral>
<face3><M><left near><happy>

The log files were created for every day, separately for each display. The tagging system did not capture any raw video and audio signals.

After the experiment, the log files were parsed. We summarized how frequently the groups observed each content category, which emotions were expressed, and the volume of the conversations. Additional information, such as the number of all groups, the total number of females, etc. could also be derived from the log files.

6 Experiment Results

In total, 324.2 hours and 4727 detected faces were recorded in the log files. The analysis of the log files enabled us to answer all questions posted to the experiment. First, we proved that the system is able to recognize the interests of distinct spectator groups. Second, we obtained interesting insights for the groups circulating in the public area. Finally, we could conclude whether emotional and conversational context can support the evidence of spectators' interests. Below we provide the detailed results.

The system successfully identified the differences in observation patterns (visual interests) among distinct groups. Figure 4 illustrates the distribution of visual interests among groups of two or three spectators. The illustration clearly shows the differences in observation patterns: for instance, topic "News" was more frequently observed by the group "1 Male + 1 Female" than by other groups. Topic "Quiz" was more often observed by homogeneous groups, "2 Males" or "2 Females". The distribution refers to the data obtained at the lobby display; very similar distribution patterns were observed on the passage displays. The figure reflects the interests of only composite groups; the interests of single spectators (one male or one female) were distributed similarly to the interests of the respective groups of two (two males or two females).

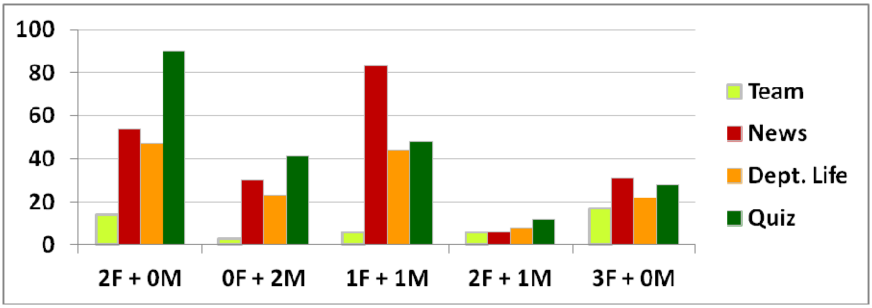


Fig. 4. Distribution of group interests (F stands for female, M – for male). Y axis indicates the number of times the visual interest of the group was detected

In total, the system detected 10 different kinds of groups. The majority of the detected spectators were single individuals (see Fig. 5). This finding was quite surprising for us, since many meetings and collaborations take place at the university.

Observing the behaviour of spectators, we realized that in spite of the gathering in meeting rooms or lecture halls, the transitional public places (such as the passage and the lobby) people mostly pass alone. Detected composite groups consisted mostly of two persons.

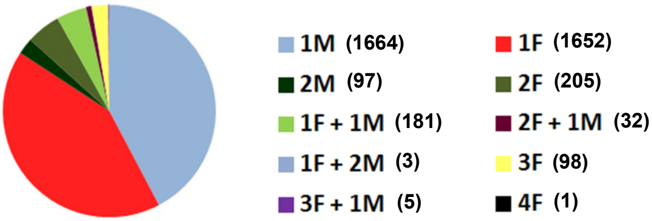


Fig. 5. Distribution of spectator groups. M stands for male, F - for female

During the experiment the system registered a solid number of positive emotions. For the analysis, we considered only positive emotions, since the SHORE software yields reliable recognition results only for “happy” vs. “not happy” emotions (see Section 4).

Figure 6 illustrates the distribution of positive emotions among single female and single male spectators. We conducted the analysis on single males and females, since they were the most represented spectator types. The analysis for other groups can be done similarly.

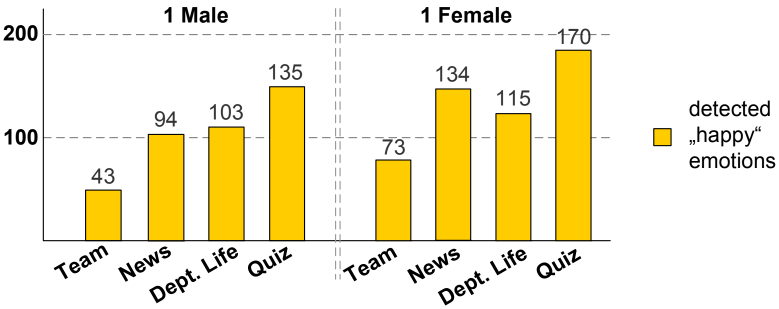


Fig. 6. Distribution of “happy” emotions

From the first sight, Figure 6 uncovers clear differences in the frequency of positive emotions expressed for different content topics. However, calculating conditional probabilities for each topic (considering how frequently each topic was observed by either group) we didn’t find any noticeable differences.

Male spectators showed almost equal emotional response to all content topics: “Team” (0.34), “News” (0.34), “Department Life” (0.36), “Quiz” (0.33). Females had a lightly more frequent positive response to “Team” (0.53) and “News” (0.45); however, quite a similar response to “Department Life” (0.4) and “Quiz” (0.39). Generally, we found that males expressed positive emotions slightly less frequently than females (0.34 and 0.43).

The analysis of the conversational activity was done in a similar way. For the analysis we considered the groups of two spectators. We chose these groups for the analysis, since the system mostly detected conversations between two persons. For each group we calculated conditional probabilities: how often and in which volume a group had conversations while observing the content.

The analysis did not reveal any noticeable differences. Most of the conversations were done in moderate volume, independently on the content topic.

Homogeneous groups (only males or only females) were slightly more silent when observing the content “Quiz”. Mixed groups had generally slightly more conversations when observing “News”. These observations can be explained by the nature of the content. “Quiz” posts the spectator a question, substituting a real conversation and thus making people silent. “News” provokes a discussion about some urgent events. However, the conversations could also be not related to the content.

7 Discussion

Below we provide the interpretation of the experiment results, addressing the research questions posted above. We discuss limitations of the study, further steps, and possible applications of the presented approach.

7.1 Content Preferences of Distinct Spectator Groups

The experiment has shown that the group-based personalization mechanism can successfully extract the differences in observation patterns of distinct spectator groups.

The main interest differences can be observed between homogeneous groups (only males or only females) and mixed groups (a male and a female). Homogeneous groups mostly preferred “Quiz” category, whereas the mixed groups were more interested in “News”.

The phenomenon can be explained by the relationships within homogeneous and mixed groups. Observing our spectators, we noticed that homogeneous groups often represent close friends. They meet at the university not only for study-related occasions, but also for socializing, chatting or spending a free time slot together. Therefore, they are likely to involve into such an entertaining occasion as a quiz. Mixed groups often represent study fellows, connected not by a friendship, but rather by a common studying activity. They meet at the university for a certain study-related occasion, e.g. to work on a project or prepare for an exam. Therefore, they are not likely to spend time for a “Quiz”, but would rather pay attention to the study-related “News”.

The overall majority of spectators showed more interest to the content “Quiz” and “News”. The preference to “News” relates to its informative content: people tried not to miss relevant and important facts. The preferences to “Quiz” can be explained by its interactive nature. The quiz questions were related to the university stories. Therefore, the quizzes not only challenged the spectators, but also gave them some curious facts.

Comparing these results with the observations of Rist and colleagues [16], who evaluated various contents at university displays, we may see slight contradictions. The authors reported that people *generally* have lower interest in entertainment content and higher interest in news. However, in their work, entertainment related to the games which demonstratively uncover user participation. Unlike games, our entertaining quiz allows the users to participate unnoticeably, with no demonstration of success or failure. Such unnoticeable interaction is known to be appreciated by people in public locations [17, 18].

7.2 Gaining Insight into Spectator Groups

Detected groups contained slightly more females than males. We found this fact surprising: statistically our technical institute counts more males than females. One explanation of this phenomenon can be the natural curiosity of women and their ability to notice the surrounding objects better than men [19].

Analyzing the log files we could see that single spectators were often joined by other persons, creating a group. Such behavior is known as the "honey pot effect" [20, 21]. People are not courageous enough to demonstrate their interest in public. Thus, they feel more comfortable to join an existing spectator.

Among the 10 detected groups, only 5 groups were presented in the passage area. The circulation of people in the lobby is indeed higher, since it is a large recreation room where people usually gather. The passage, on the contrary, is a narrow corridor. People usually pass it quickly, heading to a certain room or to the canteen. The lower number of spectators in the passage can also be explained by the orientation of the displays. As mentioned by Müller et al. [22], the displays oriented at 180 degrees to the user trajectory attract less attention than the displays oriented at 90 degrees. This observation applies to the orientation of our displays: the passage displays are oriented at 180 degrees, and the lobby display – at 90 degrees to a typical passer-by trajectory.

7.3 Emotional and Conversational Response

Observing arbitrary spectators, we noticed that positive emotions and conversations are often not related to the content. They are usually brought from a dialogue preceding the display observation. Therefore, our experiment results do not give enough evidence that detected positive emotions and conversations were *provoked* by the content.

7.4 Application in Other Public Spaces

Although the experiment was conducted in a public space with rather moderate circulation of people, it demonstrates that the system can be deployed in other public spaces. Apart from the university public space scenario, the system can be applied in an environment with a brighter diversity of groups.

The system installed at a large *shopping mall* can recognize the interests of different customer groups. Unlike existing ambient technologies facilitating shopping experience [23], our system is able to learn the interests of the customers. Based on the learnt shopping interests, the system can advertise the matching content immediately when customers approach the display. In a similar way, the system can be deployed at a *travel agency*. It will help to recognize trends in vacation destinations among couples, single travelers, families, etc.

The system can give an insight into the tastes of the people. Imagine the system installed at a *picture gallery* or a photo exhibition. Tracking how visitors observe the art pieces, we can conclude which authors and which genres are popular among different visitor groups. Such information could facilitate planning of the future exhibitions.

Finally, the system has a potential to impact the tastes of the people. Imagine the system to be installed at a university *"Open Doors" day*. The "Open Doors" day is an annual event organized by universities, aimed to orient school students in the choice of their future education. A current problem of engineering faculties is a low ratio of female students. The problem is partially caused by gender stereotypes, but partially by insufficient awareness of school girls about the engineering career. A display recognizing social context could increase their awareness. Once girls are recognized in front of the screen, the display can switch to the Engineering content.

7.5 Real-time Adaptation

The experiment illustrated how the system can be used for content tagging. The next step, real-time adaptation, can be achieved by integrating the extracted preferences into the adaptive content schedule. Once a group approaches the display, the system recognizes the group structure and switches to the content preferred by the group.

In order to validate the system performance for real-time adaptation, a more realistic *public* setting is necessary. The experiment presented in this work does show that the group-based approach can be successfully applied in a public setting. However, we found that the groups presented at the university environment are not that diverse. A real busy public place, such as a train station or a shopping mall, would be more appropriate to test the system in real-time adaptation mode.

8 Conclusion

The paper presented a system for group-based personalization on public displays. The system can be used for the *tagging of content* according to spectators' interests and for the real-time *content adaptation*.

The advantage of the proposed invention over existing systems is its completely *automatic* adaptation mechanism. The extraction of interests, as well as the real-time adaptation is performed automatically, without any input from spectator side. This requirement is critical in busy public places: the passers-by are unlikely to have time, attention or means for an input.

Another advantage of the proposed system is its capability to distinguish between spectator profiles. Instead of the retrieval of individual profiles (which is hardly realistic in a busy public place), the system extracts *groups profiles*. The groups are defined according to the number of spectators, their disposition, and gender. For example, the display distinguishes between two women, a couple, a girl or a group of boys.

An experimental deployment, conducted in a real public space, proved that the system can successfully identify the differences in observation patterns (visual interests) of different spectator groups. Moreover, the experimental results gave us insights into the circulating groups: what constellations of spectators are typical, which groups circulate in different public areas, what is the proportion of female and male spectators.

Finally, the experiment enabled us to conclude whether the tagged data on the emotional and conversational response can be correlated with the displayed content. The results gave us no evidence that positive emotions and conversations are directly related to the content. Often they are caused by events preceding the display observation. However, we have shown that the system can reliably tag positive emotions. We believe that a more entertainment-oriented content (such as a photo exhibition) can reveal differences in emotional response. The content chosen for the experiment was rather emotionally neutral; it addressed the topics relevant to the public area – a university environment.

As the next step we are planning to extend the definition of groups by recognition of *age*. This advance will enable us to distinguish, for example, between two teenage boys, an old couple or a mother with a kid. These groups are likely to have very different content preferences.

Moreover, we are planning to run a study on a real-time content adaptation, in order to see how spectators accept the automatic adaptation. However, the study has to be run in a more real public place, with diverse group profiles.

Acknowledgements. This research is partly sponsored by OC-Trust project (FOR 1085) of the German research foundation (DFG).

References

1. Müller, J., Alt, F., Michelis, D.: Pervasive Advertising. Springer (2011)
2. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: Recommender Systems: An Introduction. Cambridge University Press (2010)
3. Rashid, A.M., Albert, I., Cosley, D., Lam, S.K., McNee, S.M., Konstan, J.A., Riedl, J.: Getting to know you: learning new user preferences in recommender systems. In: 7th International Conference on Intelligent User Interfaces, pp. 127–134. ACM Press, New York (2002)
4. Masthoff, J.: Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. User Modeling and User-Adapted Interaction 14(1), 37–85 (2004)
5. Müller, J., Exeler, J., Buzeck, M., Krüger, A.: ReflectiveSigns: Digital Signs That Adapt to Audience Attention. In: Tokuda, H., Beigl, M., Friday, A., Brush, A.J.B., Tobe, Y. (eds.) Pervasive 2009. LNCS, vol. 5538, pp. 17–24. Springer, Heidelberg (2009)
6. Huang, E., Koster, A., Borchers, J.: Overcoming Assumptions and Uncovering Practices: When Does the Public Really Look at Public Displays? In: Indulska, J., Patterson, D.J., Rodden, T., Ott, M. (eds.) PERVASIVE 2008. LNCS, vol. 5013, pp. 228–243. Springer, Heidelberg (2008)
7. Müller, J., Alt, F., Michelis, D., Schmidt, A.: Requirements and design space for interactive public displays. In: 9th International Conference on Multimedia, pp. 1285–1294. ACM Press, New York (2010)

8. Alt, F., Balz, M., Kristes, S., Shirazi, A.S., Mennenöh, J., Schmidt, A., Schröder, H., Goedicke, M.: Adaptive User Profiles in Pervasive Advertising Environments. In: Tscheligi, M., de Ruyter, B., Markopoulus, P., Wichert, R., Mirlacher, T., Meschterjakov, A., Reitberger, W. (eds.) AmI 2009. LNCS, vol. 5859, pp. 276–286. Springer, Heidelberg (2009)
9. Karam, M., Payne, T., David, E.: Evaluating BluScreen: Usability for Intelligent Pervasive Displays. In: 2nd International Conference on Pervasive Computing and Applications, pp. 18–23. IEEE Press, New York (2007)
10. Mahato, H., Kern, D., Holleis, P., Schmidt, A.: Implicit Personalization of Public Environments using Bluetooth. In: 26th Annual Conference on Human Factors in Computing Systems, pp. 3093–3098. ACM, New York (2008)
11. Jameson, A., Smyth, B.: Recommendation to Groups. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 596–627. Springer, Heidelberg (2007)
12. Villar, N., Kortuem, G., Van Laerhoven, K., Schmidt, A.: The Pendle: A Wearable Mediator for Mixed Initiative Environments. In: International Workshop on Intelligent Environments, pp. 173–181. IEEE Press, New York (2005)
13. Vogel, D., Balakrishnan, R.: Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In: 17th Annual ACM Symposium on User interface Software and Technology, pp. 137–146. ACM, New York (2004)
14. Küblbeck, C., Ernst, A.: Face detection and tracking in video sequences using the modified census transformation. Image and Vision Computing 24(6), 564–572 (2006)
15. Wagner, J., Lingenfelser, F., André, E.: The Social Signal Interpretation Framework (SSI) for Real Time Signal Processing and Recognition. In: 13th Annual Conference of the International Speech Communication Association, pp. 3245–3248. Springer, Heidelberg (2011)
16. John, L., Rist, T.: xioScreen: Experiences Gained from Building a Series of Prototypes of Interactive Public Displays. In: Ubiquitous Display Environments. Springer (2012)
17. Ning, T., Müller, J., Walter, R., Bailly, G., Wacharamanotham, C., Borchers, J., Alt, F.: No Need To Stop: Menu Techniques for Passing by Public Displays. In: Workshop on Large Displays in Urban Life at International Conference on Human Factors in Computing Systems (2011)
18. Holleis, P., Rukzio, E., Otto, F., Schmidt, A.: Privacy and Curiosity in Mobile Interactions with Public Displays. In: Workshop on Mobile Spatial Interaction at International Conference on Human Factors in Computing Systems (2007)
19. Pease, A., Pease, B.: The Definitive Book of Body Language: How to Read Others' Attitudes by Their Gestures. Sheldon Press, London (1988)
20. Mathew, A., Rogers, Y., Lloyd, P.: Post-it note art: evaluating public creativity at a user generated art installation. In: 8th ACM Conference on Creativity and Cognition, pp. 61–70. ACM Press, New York (2011)
21. Müller, J., Walter, R., Bailly, G., Nischt, M., Alt, F.: Looking Glass: A Field Study on Noticing Interactivity of a Shop Window. In: International Conference on Human Factors in Computing Systems, pp. 297–306. ACM Press, New York (2012)
22. Müller, J., Wilmsmann, D., Exeler, J., Buzeck, M., Schmidt, A., Jay, T., Krüger, A.: Display Blindness: The Effect of Expectations on Attention towards Digital Signage. In: Tokuda, H., Beigl, M., Friday, A., Brush, A.J.B., Tobe, Y. (eds.) Pervasive 2009. LNCS, vol. 5538, pp. 1–8. Springer, Heidelberg (2009)
23. Meschtscherjakov, A., Reitberger, W., Mirlacher, T., Huber, H., Tscheligi, M.: AmIQuin - An Ambient Mannequin for the Shopping Environment. In: Tscheligi, M., de Ruyter, B., Markopoulus, P., Wichert, R., Mirlacher, T., Meschterjakov, A., Reitberger, W. (eds.) AmI 2009. LNCS, vol. 5859, pp. 206–214. Springer, Heidelberg (2009)

Towards the Generation of Assistive User Interfaces for Smart Meeting Rooms Based on Activity Patterns

Michael Zaki and Peter Forbrig

University of Rostock, Department of Computer Science
Rostock, Germany
{michael.zaki,peter.forbrig}@uni-rostock.de

Abstract. The main purpose of a given smart meeting room is to increase the efficiency of the meetings taking place in this room by assisting the resident actors performing their tasks and thus offering them the opportunity to focus on the exchange of information among each other. However, a proper assistance should be based on a clear understanding of the nature of tasks the users are performing in the environment. Therefore, in this paper we present an attempt to base the design of the assistive system to be operated in smart meeting rooms on activity patterns collected in the analysis stage and resulting from the perception of the human behavior in those environments. The end goal is to tailor individualized user interfaces for each actor depending on his/her current role and user profile.

Keywords: smart meeting room, task model, team model, task pattern, precondition, post-condition, assistive user interface.

1 Introduction and Background Information

As a subset of ambient intelligent environments, smart environments gained a lot of attention in the last decade. The concept of smart environments has been introduced by Weiser [1] and then further elaborated by Cook and Das when they defined a given smart environment to be "*a small world where different kinds of smart devices are continuously working to make inhabitants' lives more comfortable*" [2]. This definition insists on the fact that supporting the users performing their daily life tasks is of primary concern in such environments.

Nowadays various types of smart environments exist (e.g. Smart kitchens, smart offices, smart meeting rooms, smart homes…etc.). While the nature of tasks the user may perform differs from one domain to another, all smart environments share the same tenet which is the ability to provide proper assistance in favor of the resident actors to make their life easier and their experience in the environment enjoyable.

In our research, we focus on the domain of smart meeting rooms [3] where the main goal for which a group of people may gather is the exchange of information. Consequently, our ultimate goal is to help the users achieve their goals while being offered a convenient assistance delivered by the room. In order to take benefit of the assistance

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 288–295, 2012.

offered by the room, two interaction techniques between the user and the system are conceivable. The user may explicitly interact with the system through a user interface (UI), or alternatively the system can try to infer the current task the user is executing and based on that information, the needed assistance can then be offered to the user. We refer to those interaction techniques as explicit and implicit [4] interaction correspondingly. Actually, both interaction paradigms should be taken into consideration to design an optimal system [5]. While the implicit interaction technique seems to minimize the burden on the user since it makes the system totally invisible to him, experiments show that usually the user wants to have control over the environment and can have a negative experience about it if it does not work in the expected way. Thus, giving the user the opportunity to adjust the system explicitly is highly desirable. Whereas task models have usually been used as a tool to elicit requirements in the early development stages, nowadays they are playing a more influencing role as an appropriate starting point for interactive processes development. For example, [6] presents an attempt to create a first draft of the user interface based on task trees. Also Blumendorf et al. [7] suggests the usage of dynamic task models in order to build adaptive user interfaces. Whereas several notations for task models exist, the concur task trees (CTT) [8] is widely known and is well-suited for our purpose given the rich set of temporal operators it provides and which helps to precisely determine the order of execution of the tasks the user wants to perform in the environment. In this paper, we aim to discuss our approach suggesting a hybrid interaction technique for smart meeting rooms. Thus, we discuss an approach for generating adaptive user interfaces based on activity patterns that we collected from the perception of the behavior of actors in our smart lab. While the term "pattern" was initially introduced in urban architecture by C. Alexander [9], patterns have known their way to the software engineering [10] as well as the HCI area [11]. We define "activity pattern" to be *"A sequence of actions the user usually follows in order to perform a given activity in the environment"*. In other words, a given activity can be performed in several ways. However, by observing the actors in the environment, one can notice that for every activity there is a dominant sequence of actions the user is usually following to achieve his goal. We consider this sequence to be a pattern for performing that activity. Thus, in our work we compile a set of activity patterns extracted from the domain of smart meeting rooms. Additionally, we employ those patterns as a starting point for the design of the user interface displaying the assistance dedicated to the user by the environment. Therefore, we further refine the compiled task patterns until we reach the finest and most-detailed task model which can be used for the design of our system. Afterwards, we derive dialog graphs out of the resulting task models and finally concrete user interfaces can be tailored out of those dialog graphs. The paper is structured as follows: In the next section we discuss the methodology we followed in the analysis stage in order to compile our patterns. In section 3 we elaborate on the transition we made to achieve a model in the design level out of our patterns gathered in the analysis stage. Afterwards, the usage of the resulting model to generate the desired assistive user interfaces is presented and a brief example highlighting the main contribution of our approach is discussed. Finally, we conclude our ideas and we give a brief overview of future work in that direction.

2 User Activity Patterns in Smart Meeting Rooms

Smart meeting rooms are collaborative environments, in which a group of actors are gathering in order to achieve a final shared goal. We refer to this goal in the following as the "*team goal*". In order to achieve the desired team goal, every actor within the environment should successfully play his/her individual role within the room. Consequently, the behavior of every user in the environment can be justified by the role he/she plays.

Therefore, in order to compile meaningful useful patterns, we had first of all to look into the bigger picture and to answer the question: Why would a group of people gather in a given smart meeting room? What can be the team goal they want to achieve? After investigating the domain of smart meeting rooms, we collected five main distinguishable team goals which are formulized as final states in the following: *"Conference session performed", "lecture given", "work defended", "parallel presentation session performed" and "video watched"*. The roles to be played by the resident actors differ from one team goal to another. Thus for each team goal, different role-based patterns are to be compiled. In Fig.1 the *"Conference Session"* team pattern is depicted. As a notation for the tasks to be performed by all the actors in the room, we employ the same notation which has been introduced by the collaborative task modeling language (CTML) [12].

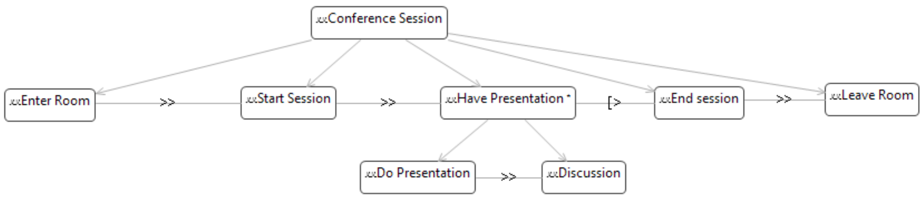


Fig. 1. Task model illustrating the "Conference Session Team Pattern"

In the above figure, we can see a task model representing the tasks the group of people in the room need to execute in order to successfully perform a conference session. First of all, all actors have to enter the room. Then, the team task "Start Session" corresponds to the fact that the chairman is introducing the session while everybody else is sitting in the audience zone. Afterwards, we have the iterative team task "Have Presentation" where the current presenter is presenting his talk while the audiences are listening to the presentation and taking notes. Once all presenters are finished with their talks, the "End Session" team task disables the previous one and the chairman summarizes the session and ends it. Finally, everybody leaves the room.

To summarize, for every task within the conference session team pattern, corresponding tasks in the included roles are to be executed. Usually, the users playing those roles stick to a sequence of actions execution to accomplish their goal. Those sequences constitute our task patterns or namely "*role-based patterns*". An example of one of those patterns is the "presenter" role to be played within a conference session and which is visualized in Fig.2.

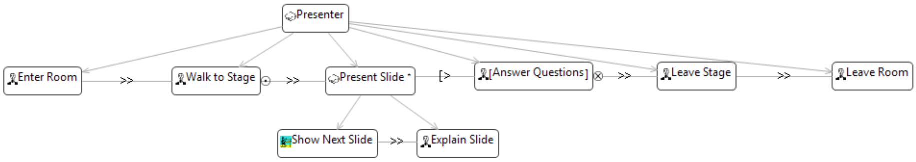


Fig. 2. The "presenter" role-based pattern for a conference session

Unlike the previous pattern, the one above is simply illustrating the tasks to be executed by an individual who is playing the role "presenter".

In order to use those patterns for the design of our assistive system, further transformations should take place until all technical characteristics of our system are well specified within the model. In the next section we thoroughly discuss this issue.

3 Towards Individualized Assistive User Interfaces

3.1 From Analysis to Design Stage

As task models enable us to focus on the tasks the user is performing in the environment, they can be further used as a starting point for the design of the application. However, those models should obey to various transformations so that we can achieve a holistic specification of the application to be developed. Briefly, the development process goes through three distinguishable stages, namely the analysis, the system requirements specifications and the system design stages.

The role-based patterns discussed in the previous section are structurally task models which have been gathered in the analysis stage. Thus, they specify the way users are performing their daily life tasks in the environment without any intervention from an external system. Once we gather those models, a transformation to the second stage is feasible. In this stage the functional requirements of the system under construction should be integrated within the model while abstracting from the technical details of interaction between the user and the application. Finally, after validating the models at that level, further transformations take place to assimilate the design decisions taken by the developer and which encapsulate the exact way in which the user is going to interact with the system.

Nevertheless, the transformation of a given model from one abstraction level to another should be validated in order to ensure that the resulting model did not violate any of the main requirements of the system to be designed. There are already various techniques to accomplish the validation process. Among those techniques we can mention the bisimulation equivalence and the trace equivalence [13] techniques. A suitable refinement and validation technique for task models from our point of view is performed using the so-called meta-operators [14]. Briefly, those operators enable the designer to attach the desired level of restriction to every task within the task model to be transformed to a lower level of abstraction. In that way, those operators are acting as a communication language between the analyst and the designer so that the analyst can express which tasks should remain persistent in all subsequent levels, which tasks could be removed and which tasks could be further refined. Furthermore, a set of guidelines guiding the assignment of those operators to the different tasks have been presented by the authors in [15]. Due to lack of space, Fig.3 illustrates only a part of the role "presenter" in the design level, resulting from the initial task pattern collected in the analysis stage and which is depicted in Fig. 2. In those task models, the benefit of the usage of meta-operators is exemplified. For example, in Fig 2 a shallow binding operator is bound to the user task "Walk to Stage". The semantics of this operator forces the designer to keep that task but gives him the opportunity to further refine it with subtasks in subsequent levels as we can see in Fig. 3. However, the "Answer Questions" user task has been bound with the deep binding operator which means that this task should remain the same without any changes in all subsequent levels. Please notice that the task model presented in Fig. 3 should be further refined by adding more technical details concerning the exact way those interactive tasks should take place. Thus, this is a primary model in the design level and not the very final model based on which the design can be built.

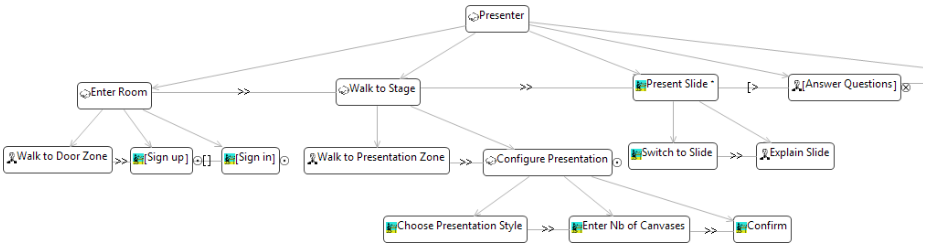


Fig. 3. The "presenter" task model in the design level

3.2 From Task Models to Assistive UIs for Smart Meeting Rooms

In this subsection, we briefly discuss the transformation of the task model in the design level to the final individualized UI we aim for. There have been already various attempts to generate user interfaces out of task models [16, 17]. We follow a methodology which is similar to the one presented in [18] in order to bridge the gap between the obtained task model and the final required UI. We start by mapping our task models to dialog graphs. Afterwards, another transformation takes place where the resulting dialog graph is transformed to an abstract user interface targeting multiple final user interfaces. Finally, depending on the device on which the user interface has to run, the final user interface is built. However, we believe that the optimal user interface to be operated in a given smart environment should not have a fixed set of supportive tasks. The kind of support to be displayed for the user should be relevant to the task the user is currently performing in the environment. Thus, whenever the system recognizes that the actor is finished with a given task and has started executing another one, the user interface visualized by that actor should be updated and has to offer a set of options which are directly related to the new task being performed. In that way, the user is not confused by a crowded user interface with a big set of useless assistance. For example, if a given user is finished with her presentation and walked from the presentation zone to the audience zone, it does not make sense anymore to display the presentation styles offered by the room to that user since currently this user is playing the role “listener”. Thus, the user interface should offer her more information about the next talk for example. In other words, our idea is to merge the explicit and the implicit interaction techniques so that we can optimize the quality of assistance offered by the room. We implicitly infer the most probable current task the user is executing, and based on that information we update the user interface where the user can explicitly ask for help. We employ environmental pre and post-conditions in order to recognize the task the user is currently performing. To make our ideas more concrete, we take here a very brief overview of the system and the exact way the application has to work in the room. We are currently developing the application as a plugin using eclipse IDE. Thus, we foster our explanation here with some figures of that tool. Due to lack of space, we do not use figures to illustrate all steps the user has to follow while using the application. First of all, the application needs to be configured according to the scenario which is about to run in the room. This configuration should be realized by the responsible for the room or the chairman by interacting with a dialogue in which he can choose the team goal for which the users are about to gather in the room. Let us assume we are about to have a conference session in the room, so that we can stick to the same example we adopted in the paper. Consequently, on the user interface of the chairman, he will be able to see the corresponding team pattern for the scenario chosen and which has been depicted in Fig.1. As already discussed, for every team pattern, corresponding role-based patterns have been compiled. An example of such a pattern (The presenter role-based pattern) has also been shown in Fig.2. For every given task within the team pattern, there is a characteristic post-condition signalizing that this task has been successfully executed. Our experimental smart meeting room in the university is equipped with various types of sensors (e.g. RFID tags, Ubisense, etc.) which enable us to check the state of any entity in the environment. That is how we can check whether a given post-condition has been realized. In Fig. 4, two of the team tasks included in a conference session team pattern are presented. The figure shows the mapping between every team task and the tasks to be accomplished by the individuals playing the corresponding roles. Those tasks can be seen under the column “During Task Execution”. Moreover, one can see the pre and post-conditions for both team tasks. Whenever we infer that the pre-condition of a given task is true, then our system makes the assumption that this task has started. Thus the user interfaces for the different related roles should then be updated.

Task	Pre-condition	During Task Execution	Post-condition
Start Session	Door. closed=true	Listener. Sit and Listen to Talk & Listener. Take Notes Chairman. Introduce Session	Chairman. leave stage=true
Do Presentation	Presenter.isAtStage=true	Presenter. Present Slide Listener. Sit and Listen & Listener. Take Notes Chairman. Manage Talk	Presentation. allSlides. explained=true

Fig. 4. Team tasks of the conference session scenario with corresponding individual tasks

Let us consider the case of a given presenter within this scenario. He will start by registering for the application and creating his own user profile. The patterns we collected are adaptable according to several user characteristics. For example, in case of a visually impaired user, the model is adapted by integrating the user-oriented accessibility patterns presented by the authors in [19]. When for example the precondition "*Presenter.isAtStage*" is true, then the user interface dedicated to that presenter will display the various styles of presentations offered by the room, so that he/she can pick the preferred style for the talk, and the number of projectors (tracks) to be used as it is shown in Fig.5. To summarize, the user interface is tailored according to the exact needs in terms of user profile and also to the current task being executed by the user.

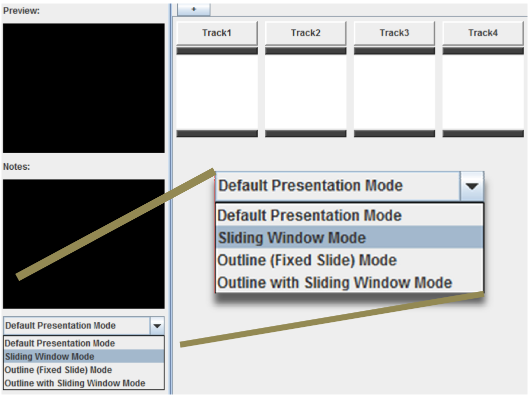


Fig. 5. Multiple presentation techniques offered for the presenter

4 Conclusion

In this paper we tackled the notion of having tailored and individualized assistive user interfaces in favor of the users in smart meeting rooms. In order to achieve our goal, we started by identifying the several collaborative goals for which a group of people may gather in a given meeting room. We experimentally analyzed the behavior of the actors within our smart meeting room while accomplishing those team goals so that we can compile our collection of role-based patterns. We discussed in the paper, how we gradually evolved our models to make the transition to the lowest level of abstraction by integrating the technical characteristics of the system we want to design for the room within our models. After that, we discussed how we want to cover the gap between the resulting model and the final user interface we aim for. We made it clear that we strive for a combined interaction means between the user and the final system. We aim to implicitly infer which task the user is performing and adapt the user interface with which he/she explicitly interacts depending on that information. We briefly explained the concept of recognizing the team tasks based on environmental pre and post-conditions after mapping each of those tasks with corresponding tasks to be accomplished by the individuals playing the needed roles. In the future, we aim first of all to finalize the development of our application and optimize its usage. After that, we have to validate our application and evaluate it against the different challenges it was developed for. The evaluation of such a system is not a trivial process. We formulize the questions we would like to answer by the evaluation in the following:

- How efficient is the assistance provided by the user interface?
- How flexible are those user interfaces?
- Can they be employed in all kinds of smart meeting rooms?

Based on the experiments we have so far, we truly believe to get a positive feedback out of the planned evaluation.

References

1. Weiser, M.: The Computer for the 21st Century 265, 94–104 (1991)
2. Cook, D., Das, S.: Smart Environments: Technology. Protocols and Applications. Wiley Series on Parallel and Distributed Computing (2004)
3. James, A., Miyazaki, J.: Building A Smart Meeting Room: From Infrastructure To The Video Gap (Research And Open Issues) (2005)
4. Schmidt, A.: Implicit Human Computer Interaction Through Context. Personal and Ubiquitous Computing 4 (2000)
5. Ju, W., Leifer, L.: The Design of Implicit Interactions: Making Interactive Systems Less Obnoxious. Design Issues 24(3), 72–84 (2008)
6. Feuerstack, S., Blumendorf, M., Albayrak, S.: Bridging the gab between Model and Design of User Interfaces. In: Proc. GI Jahrestagung (2), pp. 131–137 (2006)
7. Blumendorf, M., Lehmann, G., Albayrak, S.: Bridging Models and Systems at Runtime To Build Adaptive User Interfaces. In: Proc. of the 2nd ACM SIGCHI Symposium on Engineering Interactive Computing Systems, pp. 9–18 (2010) ISBN: 978-1-4503-0083-4
8. Paterno, F., Mancini, C., Meniconi, S.: ConcurTaskTrees: A diagrammatic Notation for Specifying Task Models. In: IINTERACT 1997, IFIP TC13, pp. 362–369 (1997)
9. Alexander, C., Ishikawa, S., Silverstein, M.: A Pattern Language. Towns, Buildings,Construction. Oxford University Press (1977)
10. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley, Reading (1994)
11. Borchers, J.: A pattern approach to interaction design. In: DIS 2000 Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 369–378 (2001) ISBN: 1-58113-219-0
12. Wurdel, M., Sinnig, D., Forbrig, P.: CTML: Domain and Task Modeling for Collaborative Environments. J.UCS 14, 3188–3201 (2008)
13. Bergstra, J.A.: Handbook of Process Algebra. Elsevier Science Inc. (2001)
14. Wurdel, M., Sinnig, D., Forbrig, P.: Task Model Refinement with Meta Operators. DVS-IS (2008)
15. Zaki, M., Wurdel, M., Forbrig, P.: Pattern Driven Task Model Refinement. In: Abraham, A., Corchado, J.M., González, S.R., De Paz Santana, J.F. (eds.) DCAI. AISC, vol. 91, pp. 249–256. Springer, Heidelberg (2011)
16. Luyten, K.: Dynamic User Interface Generation for Mobile and Embedded Systems with Model-Based User Interface Development, phd Thesis (2004)
17. Pribeanu, C.: An Approach to Task Modeling for User Interface Design. World Academy of Science, Engineering and Technology (2005)
18. Wolff, A., Forbrig, P.: Deriving User Interfaces from Task Models. In: MDDo-AUI 2009, Sanibel Island, USA (2009)
19. Zaki, M., Forbrig, P.: User-Oriented Accessibility Patterns for Smart Environments. In: Jacko, J.A. (ed.) HCI International 2011, Part I. LNCS, vol. 6761, pp. 319–327. Springer, Heidelberg (2011)

Reducing Dementia Related Wandering Behaviour with an Interactive Wall

Saskia Robben, Kyra Bergman, Sven Haitjema,
Yannick de Lange, and Ben Kröse

Amsterdam University of Applied Science, DMCI Create-IT, Duivendrechtsekade 36-38, 1096 AH, Amsterdam, The Netherlands
{s.m.b.robben,kyra.bergman,sven.haitjema,
yannick.de.lange,b.j.a.krose}@hva.nl
http://www.digitallifecentre.nl/

Abstract. People suffering from dementia often have problems with way finding and feel restless. In this paper we present an interactive wall developed for decreasing the amount of wandering behaviour of people suffering from dementia. The installation aims at making these people feel more at home in the nursing homes by guiding them with a motion triggered audio path. This leads them to a wall with large windows displaying images and short movie tracks from their hometown. The results of an observation study show that the interactive wall succeeds in attracting people and thus reducing the wandering behaviour. Remarks of the elderly as well as their family and caretakers support this conclusion.

Keywords: Dementia, Wandering, Elderly People, Alzheimer, Interactive Wall.

1 Introduction

With the population of the world aging and the number of people with dementia growing exponentially [1], a predicted 65.7 million in 2050, the amount of research in the field of dementia is also gaining traction. On the one hand scientists are looking for causes of dementia [2]. One the other hand there are scientists that focus on the possibilities of reducing the effects of dementia or assisting the elderly with (interactive) technology [3]. Dementia is a degenerative condition in the brain that mostly occurs as people are getting older. It progressively reduces a person's ability to remember, think logically, communicate effectively and care for themselves. Although it is known that dementia is caused by structural and chemical changes in the brain that eventually lead to the death of brain cells, it is hard to prevent or cure dementia [4]. Studies [5,7] researching the wandering behaviour of elders with dementia show that they often feel lost and out of place. They can feel locked in their own environment and start to wander, looking for the way out or the path to their own destination, e.g. their home, work, family or spouse. 63% of all the people with dementia wanders and 70% of the caretakers see the wandering as a risk for the care of these people [5]. They have the risk of falling or getting lost or fatigued.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 296–303, 2012.

Multiple interventions have been described for calming the elderly including multi-sensory stimulation [12]. Though such interventions engage the elderly in an experience and reduce stress and wandering behaviour, often supervision by a caretaker or family member is necessary.

There is a need for an installation with which wandering elderly can interact independently. Can the use of such an interactive installation be effective in reducing the wandering behaviour of the elders with dementia?

In this paper we present such an interactive installation that engages the elders with dementia in a pleasant experience and distracts them from their wandering. A motion triggered sound path provides a direction or a goal for people who feel lost and are wandering through the hallways of an restricted psychogeriatric ward. The path leads them to a wall which aims at offering peace by providing a recognizable view on the world outside of the home. For evaluating the effect of the interactive installation observation studies will be performed accompanied by interviews of the caretakers and family.

2 Related Work

Several nursing homes in the Netherlands as well as in the UK have already build an (interactive) installation in an attempt to reduce the wandering of the elderly residents and inducing a feeling of comfort and peace. Three notable examples are the Train Wagon ('Coupé') in Delft [8], the Beach Room in Vreugdehof [9] and Millhouse in Cheshire [5]. The Train Wagon and the Beach Room provide the elderly with a room or setting in which they can relax, emerge in a experience or have social contact with their fellow travellers. The residents with dementia in both nursing homes have responded positively to the installations. There is less wandering around and especially the Train Wagon also assists the residents in their path finding, as it may be perceived as a way to reach their own destination. Whereas both installations are single rooms within a larger complex and only the Train Wagon can be operated independently by the elder, the recently developed nursing home in Cheshire [5] has been completely designed for accommodating the elderly with dementia as comfortably as possible. The doors to the rooms of elderly look like real porches, providing a sense of ownership and independence. Non-accessible doors and gates have been painted the same colour as the wall, extending skirting boards and handrails. The house aims at giving the residents a feeling of freedom and independence, while not alerting or stressing them with opportunities of leaving the premises, and thus reduce the wandering.

3 Product

The interactive wall consists of two main elements: direction and experience. The directive part of the installation consists of an interactive sound path that reacts to the presence of an elder and plays sounds varying between music from the 1960's and the audio that normally comes from a living room. This path of sound, which travels along with the movement of the elder, leads the elder to the interactive wall, the experience.

As can be seen in figure 1, this wall features three interactive 'windows' that serve as a portal to the outside world. As dementia progresses the elders sense of time rewinds and they feel like they are living in their younger years, therefore all the ornaments on the wall are designed in the style of a Dutch 1960's house. On the screens movies are shown from several cities in Holland. The cities are currently chosen based on the amount of people in the institute that have resided there. Both the path and the wall emerge the elders with dementia in an experience. The sounds played are hits from their younger years and the wall acts as a possibility of looking outside and beyond the borders of the hallways of the nursing home.

3.1 Requirements

Studies [7,10,11] into the requirements for designing products for people with dementia provide valuable insights. Any interaction required with the product should be kept to a minimum, the installations should feel safe and secure and any experience should connect with the goals and destination of the elderly. Orpwood et al. [10] also indicate that any interaction that involves a chain or sequence of events can cause much difficulty for people with poor working memory such as those with dementia. These difficulties can often lead to withdrawal from engaging in such activities because of their experiences of repeated failure.

Furthermore interviews with the caretakers of the psychogeriatric ward in Naarderheem combined with several observations of the current situation brought to light that, although the elderly act often confused and as living in their own reality, they still act very curious. They look into every door and room and especially in the early stages of dementia can also be very suspicious and alert. In the development of the interactive audio path and wall proposed in this paper these user characteristics were taken into account. There is a balance between creating an installation that stimulates the senses while emerging the user in a relaxing experience that does not demand any additional interaction besides their presence or is to fragile to handle by the residents.

3.2 Implementation

The centre of the installation is a computer which connects and synchronizes all of the devices. Using digital time switches the installation automatically starts up at 10 am in the morning and shuts down at 9 pm while requiring no additional actions from the caretakers. When the installation is off, only a light bulb is on.

The audio path consists of four connected printed circuit boards with each having a passive infrared motion sensor (PIR) and a speaker with different volume settings. The sound begins when a motion is detected and then plays for ten seconds, this provides a sense of direction.

The three LCD screens are each connected to a media player and are protected by transparent plates of polycarbonate. The placement of one of the windows was optimized for viewing from a wheelchair perspective while the other two are for standing people. The movies shown on these screens are shot in a specific way, three cameras were positioned in exactly the same way as the windows on the wall so the resulting images are coherent with reality. To this end a special tripod was developed.

Both the path and wall have built-in tools and connections that enabling altering and extension of the system. Though currently not in use, the wall also features motion sensors for allowing specific actions when the elderly are approaching.

Employees of the nursing home without any specific technical knowledge can alter or replace movies and audio material. Also all systems are prepared for time and season dependent sounds and images.

Fig. 1. Left the interactive wall with the three 'windows' is displayed. On the right there is a close-up of the small fireplace which is one of the ornaments on the wall.

4 Experiment

The aim of the observation study was measuring the response on the interactive wall.

The interactive wall and audio path are installed in the psychogeriatric ward in healthcare centre Vivium Zorggroep Naarderheem in Naarden. The ward consists of four connected hallways around a squared courtyard, with a living room on each of the four corners. Each of the residents belongs to a living room with specific caretakers, but are free to walk on the premises. For evaluating the wall and its effect on the wandering behaviour of the elderly an observation study was conducted.

As the target group is hard to interview and often inconsistent in their answers or reactions, we primarily used the knowledge of the caretakers and family to provide us with greater insight in the usage of the installation by the residents and the experiences of these elders with the installation.

4.1 Observation

The number of demented people passing by were measured as well as reactions such as looking to the wall, slowing down, stopping in front of the wall, talking and smiling.

For a baseline comparison the same wall is observed, but with the interaction shut off. Because the wandering behaviour in the hallways is not continuously the same, a window of several hours was used to measure possible responses. The baseline observation was three hours and the actual observation six hours.

The number of elders that wandered past the installation were measured and any noticeable responses to the installation were listed. For standardization of the observation and analysis a observation table was used with predefined fields for (anonymous) ID of the resident, the time and any visible or audible reactions.

4.2 Caretakers

Besides the observation study, the caretakers of the living room closest to the wall were asked to fill out a questionnaire about the wall after each shift. The questionnaire contained both closed questions about the number of elders they spotted walking past the wall as well as room for their own remarks and suggestions. Besides this formal feedback the caretakers and family also provided informal feedback, which was taken into account.

5 Results and Conclusion

5.1 Observation

The results of the observation study can be found in table 1. The observation held when both the wall and path were turned on showed that 20 of the counted residents (N=75) looked at the new installation by turning their heads. Twelve of the observed residents actually stood still and redirected their attention from wandering to the sound and images. This in comparison to the baseline situation, where none of the passers stopped. Besides noticing the installation and stopping in front of it, also a number of smiles where observed. This indicates

Table 1. Results of Observation Study

	Baseline Observation	Observation
Walking	25	60
Wheelchair	11	15
Look	0	20
Slow down	3	20
Stopping	0	12
Talking	0	2
Smile	0	20

that the residents like the wall, which is confirmed with some of their remarks. In comparison to the baseline situation the residents showed a far greater interest in the wall and, when asked, started describing the images and relating them to their own destination and or experiences from their past: 'That's the great church of Naarden!' - 'I know that village!' - 'A wedding, those children always run around'.

5.2 Caretakers

The response to the questionnaires was quite low. During the first week the installation was turned on only the caretakers of one shift filled in the questionnaires, resulting in six returned questionnaires. They report an average of ten people walking past the installation of whom seven stopped in front of it. No additional feedback was given on the forms. However, interviews with the caretakers provided us with insight in the reception of the wall.

Although the first interaction with the installation was often invoked out of own interest from the wanderer, when asked by the caretakers or family additional motivation or recognition was often extracted. Some of the wanderers also tend to walk in pairs, vividly discussing the images shown on the screen or quietly singing along with the music played by the path. Many of the wanderers, when asked, started telling about their own past, the objects they recognized or responded to the presence of the new installation in the hallway.

According to the caretakers some of the noticeable quotes of the residents that stood in front of the windows of the wall were: 'I know that place, it's Heerhugowaard, I have lived there for many years' - 'It's magnificent!' - 'They even make sound, finally some action in the hallways!'.

Often family of the elder asked about the possibilities to also show video from the hometown of their father or mother or add their picture into one of the frames on the wall. This engagement of the family with the wall was an unforseen side-effect but valued high by caretakers. The family and also the caretakers who were interviewed unanimously agreed that the new installation is a great addition to the hallways, allowing the elder to enjoy some music or talk among each other about the video images shown on the screen.

6 Discussion and Future Work

Overall, during the evaluation, the interactive path and wall was experienced positively by the wandering elders and the installation proved to be an improvement in attracting the elders attention compared to the old empty environment of the hallways.

However, our study has some limitations, one of which is that this study did not differentiate between people with different wandering patterns. Another one is that it is not determined yet if the wall has a similar positive effect after a longer period of time, because it might be possible that there will be a habituation effect. But this study shows that the proposed installation initially succeeds in reducing the amount of wandering of the elders and emerges them in an experience, e.g. by bringing up childhood memories.

While the caretakers at Naarderheem provided us with valuable insight into the residents in interviews, the response rate to the questionnaires was quite low. The response rate could very well have been influenced by the workload and a lack of time. Worth mentioning is that only the caretakers operating in one of the four living rooms, the one closest to the path and wall, were asked to fill in the questionnaire. The residents of the other living rooms, further from the installation, do wander and walk by the wall but their caretakers confide to their own spaces and hallways and therefore are never observe any use of the installation.

The results from the study also provided ideas that can further improve the interactive wall. As the people have a positive response to familiar content, adding recognition of the wanderer to the wall would provide the opportunity for displaying content familiar for a specific user. For example RFID technology provide the possibility for displaying custom user content. Another added value would be a connection to the internet. If a web-server is set up on the system, family members could log on to the web-server and upload family pictures or home video's. When a resident approaches the wall, the computer reads his RFID tag and matches the appropriate content and displays it on the appropriate screen. Secondly the content displayed on the screen could be adjusted as the head of the elder moves, using eye trackers. This enables a far wider viewing angle per screen, enabling the possibilities to view a whole view in one screen. It should be noted that such an experience (image moving along) should be tested with the elders, as it might create extra confusion.

The interactive wall proves to be a successful installation in reducing dementia related wandering behaviour. An advantage to other interventions is that the wall can be approached independently by the elderly. Additionally it allows the family of the elderly to be engaged by providing personalized content.

Acknowledgements. This research was supported by the 'Health-lab' project in the Dutch Pieken in de Delta program and the 'Smart Systems for Smart Systems' project in the SIA-RAAK program. We thank the people from Vivium Zorggroep Naarderheem, in particular Marco Wisse, Rob Hooft, Gerard Veenendaal and Diane de Graaf, for their abundant suggestions and assistance and also we thank the residents the psychogeriatric ward and their caretakers.

References

1. Wimo, A., Winblad, B., Aguero-Torres, H., Strauss, E.: The Magnitude of Dementia Occurence in the World. Alzheimer Disease & Associated Disorders 17, 63–67 (2003)
2. Cooper, J.K., Mungas, D., Weiler, P.: Relation of cognitive status and abnormal behaviors in Alzheimer disease. J. Am. Geriatric Society 38, 867–870 (1990)

3. Duh, H.B.L., Do, E.Y.L., Billinghurst, M., Quek, F., Chen, V.H.H.: Senior-Friendly Technologies: Interaction Design for Senior Users. In: Proceedings of the 28th of the International ACM Conference on Human Factors in Computing Systems, pp. 4513–4516 (2010)
4. Crips, H.: Spotlight on dementia care: A Health Foundation improvement report. The Health Foundation UK (2011)
5. Utton, D.: The design of housing for people with dementia. Journal of Care Services Management 3 (4), 380–390 (2009)
6. Robinson, L., Hutchings, D., Corner, L., Beyer, B., Dickinson, H., Vanoli, A., Finch, T., Hughes, J., Ballard, C., May, C., Bond, J.: A systematic literature review of the effectiveness of non-pharmacological interventions to prevent wandering in dementia and evaluation of the ethical implications and acceptability of their use. Health Technology Assessment 10(26) (2006)
7. Klein, D., Steinberg, A., Galik, E., Steele, C., Sheppard, J.M., Warren, A., Rosenblatt, A., Lyketsos, C.: Wandering Behaviour in community-residing persons with dementia. International Journal of Geriatric Psychiatry 14, 272–279 (1999)
8. Droge Wendel, Y., Hellings, L.: De Coupé, De Bieslandhof Utrecht (2008), http://classic.skor.nl/artefact-3538-en.html (accessed June 22, 2012)
9. IDe: Tot rust komen op privestrand in verpleeghuis (2011), http://www.innovatiekringdementie.nl/Nieuws/Tot-rust-komen-op-privestrand-in-verpleeghuis.aspx (accessed June 22, 2012)
10. Orpwood, R., Sixsmith, B., Torrington, J., Chadd, J., Gibson, G., Chalfont, G.: Designing technology to support quality of life of people with dementia. Technology and Disability 19, 103–112 (2007)
11. Margot-Cattin, I., Nygård, L.: Access technology and dementia care: Influences on residents' everyday lives in a secure unit. Scandinavian Journal of Occupational Therapy 13, 113–124 (2006)
12. Lancioni, G.E., Cuvo, A.J., O'Reilly, M.F.: Snoezelen: an overview of research with people with developmental disabilities and dementia. Disability & Rehabilitation 24(4), 175–184 (2002)

Gesture Based Semantic Service Invocation for Human Environment Interaction

Carsten Stocklöw and Reiner Wichert

Fraunhofer Institute for Computer Graphics Research,
Fraunhoferstr. 5, 64283 Darmstadt, Germany
{carsten.stockloew,reiner.wichert}@igd.fraunhofer.de

Abstract. The assistance of users in their activities of daily life by a smart environment is the main goal of Ambient Assisted Living (AAL). In this case, interaction is of particular interest since some users are very familiar with modern technology and for some users this technology is very challanging so that poorly designed interaction metaphors will lead to a low acceptance. Additionally, AAL has to cope with the challenges of open systems in which at any time new devices and functionalities can appear. This paper presents a gesture based approach to control devices and their functionalities in a smart environment at a semantic level to issue a command or to set a level. Redundant functionalities are filtered out before presenting the list of functions to the user. This concept is validated by a demonstrator that uses the semantic AAL platform universAAL.

Keywords: Gesture based Interaction, Semantic Services, Ambient Assisted Living, Human Environment Interaction.

1 Introduction

Ambient Assisted Living (AAL) and Active and Healthy Ageing (AHA) both target the assistance of users in their activities of daily life. Although the main focus group of AAL is often associated with elderly or disabled people, the goal to increase the quality of life is valid for all humans at every age. One of the biggest challenges in AAL is the interaction of users with devices in the environment. Over the years a number of devices became available for the consumer market to control not only the well-known technical multimedia devices (like TV or hi-fi system), but also devices that were previously not technically controllable (like window, blinds, or heating). Using standardized protocols like KNX[1] these appliances can be connected to a platform that allows a unified interaction method and thus eases the usage of new equipment. If the platform is designed as an *open system* and functionality is described in a semantic way, then potentially any new device can be attached to the platform and made available to other components without changing existing components.

[1] http://www.knx.org/

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 304–311, 2012.

Explicitly controlling devices in the environment can be achieved intuitively by using hand gestures. As humans use gestures quite often in everyday life, they are expected to be an intuitive way to be used also for the interaction with smart environments. This paper investigates the usage of a unified gesture based interaction metaphor to control different devices in a smart environment (e.g. the living room) with an open system platform. The focus group is hereby not restricted to elderly people. Although Hassani et al [3] have shown a positive reaction of elderlies to in-air hand gestures, which could indicate a high acceptance on the market, this system can also be used by younger people who may already have some experience with modern interaction devices like Microsoft Kinect, which is also used in this work to capture hand gestures.

2 Related Work

Using hand gestures to interact with a system is not a new topic and is subject of a number of research papers. Currently, numerous researchers and companies use the Kinect as input device to capture gestural commands. Kim et al. [4] uses hand gestures to interact with elements projected on a wall. The gestures in this case replace the mouse of a traditional graphical UI and the displayed content seem to be fixed to a certain scenario. Caon et al. [2] investigates the use of multiple Kinects, but the interaction is restricted to turning devices on or off according to their current state (toggle status) by pointing at them. Other work use pointing gestures to interact with a robot and to invoke some specific functionality [1]. The relationship between hand gestures and filtered semantic services was not part of these papers.

An interesting approach is ReWiRe by Vanderhulst et al. [8] which describes the environment and its users, devices, tasks, and services on a semantic level and provides a semi-automatical way to cope with changes in the environmental setup. It could be beneficial to investigate if the semantic filtering approach presented in this paper could be realized as a rewiring strategy.

3 Gesture Based Interaction

This work is mainly concerned with gestures to control devices at home. This section investigates possible gestures that can be used according to the devices and functionalities of the environment.

3.1 Requirements for Gesture Based Interaction

To achieve a high acceptance of end users, the following requirements of the gestures have to be fulfilled (compare also to the guidelines provided by Nielsen et al. [6]):

1. *Unobtrusive*: since the gestures can occur at any time during the day in the living area, it is essential avoid any kind of marker or special remote control.

2. *Simple*: the gestures should be easy and fast to perform.
3. *Big interaction zone*: interaction can happen at any position in a room.
4. *Intuitive*: the gesture should be known from different use cases. If the same gesture can be used in different application contexts, it is easier to remember and is often assumed to be more intuitive.
5. *Context aware*: the interaction should take the context of the user into account to infer his/her intentions. For example, if the user is sitting on the couch and looking towards the TV which is turned on, a simple swipe gesture can be used for switching the channel. Otherwise, a more complex interaction might be needed to ensure that the gesture is really a command to a device in the environment.

3.2 Devices and Functionalities

Since the interaction takes place at home, the devices to consider are, amongst others: TV, hi-fi system, window, lamps, heater, and blinds. Most of these devices have the functionality to set a certain level, e.g. the brightness of a light source or the volume of the TV. Some of the functionalities can be treated as special case of this level-based interaction, i.e. turning the lamp off is equivalent to setting its level to zero. This special case is further analyzed in section section 4.4 on the filtering of semantic services. Thus, the focus of this work is on the level-based interaction. Additionally, simple commands are possible, e.g. to turn the hi-fi system on or off. If multiple devices are selected or a device offers multiple functionalities, and the system can not clearly identify what the user wants, the system has to ask the user for more information. This can be realized by a multiple choice question.

3.3 Selection of Gestures

In most cases the interaction is initialized by selecting the device that needs to be controlled. This device selection is typically realized by a pointing gesture. If the user is pointing at a device for a certain time (e.g. 1-2 seconds), it counts as being selected and can be controlled. During control and selection of the functionality, simple swipe gestures can be used to apply the chosen operation (swipe to the left) or cancel the interaction (swipe to the right). This is in accordance to the typical usage of horizontal swiping to accept and go to the next screen as it is known, for example, from touch-based mobile phones or tablet PCs. For level-based interaction, the *slider* interaction metaphor can be used (see Fig. 3 (right)). Since horizontal movement of the hand is used for cancel and accept, vertical movement of the hand is used for the slider to realize setting a level of the device.

If multiple devices are at the same position (like window and blind) or functionalities other than setting a level is provided by a device, then the system needs more information from the user to perform any action, so it has to ask in form of a multiple choice question. This question can be answered by the *very same interaction metaphor*: swiping horizontally to accept or cancel, and moving vertically to select one of the choices.

4 Semantic Services

In AAL scenarios an open system should be assumed in which at *any* time a new component or device can be attached to the platform. Currently, the trend goes in the direction of ontologies and semantic services with well-defined representations using the standards RDF[2], OWL[3], and OWL-S[4]. This section defines the requirements that the platform should fulfill that originate from the application scenario described in the previous section. According to these requirements the description of semantic services and the filtering for presenting it to the user is explained.

4.1 Requirements of the Platform

The following requirements must be fulfilled by the platform:

1. *Abstraction of hardware*: devices in the environment must be controllable by a unified interface independent from the underlying low-level protocols.
2. *Semantic interoperability*: as an open system it must be possible to interoperate on a high semantic level to allow for new components at any time and for semantic filtering before it is presented to the user (see section 4.4).
3. *Ontologies*: ontologies should be used as a standardized way to model the functionalities of devices in the environment.
4. *Service registry*: it must be possible to query all existing services of a device so that its functionalities can be presented to the user.
5. *Service profile*: all functionalities of a device must be formulated in way that is understandable by the system (e.g. using OWL-S).
6. *Matchmaking*: basic matchmaking functions are needed in a semantic platform to determine whether a service request matches a registered service profile. In this work, it is additionally used to filter the list of services.

4.2 Ontologies for Semantic Interoperability

Some parts of the ontologies are shown in Fig. 1. All functionalitiles are provided by sub classes of *Service*, i.e. all functionalities of devices are provided by subclasses of *DeviceService* which controls a set of *Device*s. For example, the *Lighting* service controls a set of *LightSource*s which have an integer value between 0 and 100 as brightness (given as percentage). The *Location* (like building, room, position in a room) and an orientation is available for every physical thing; it is described in more detail in a previous paper [5]. With this information, it is possible to query the interaction device (e.g. the Kinect) to determine the exact position, orientation and pointing direction of the user inside the room. The Kinect already provides the location information of the user in a metric

[2] http://www.w3.org/RDF/
[3] http://www.w3.org/OWL/
[4] http://www.w3.org/Submission/OWL-S/

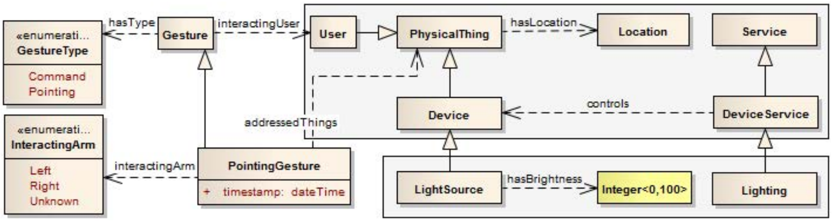


Fig. 1. Ontologies: physical world (top right), lighting (bottom right) and gestures (left)

system that can be used directly. By enumerating all devices in the room of interest and the pointing direction, the device that the user has selected can be calculated by testing for intersection of the pointing cone with the bounding box of the device (a cone is used to cope with increasing measurement errors with increasing distance). After the device is found, the service profiles that describe the functionalities of that device can be queried from a service registry.

4.3 Description of Semantic Services

Service profiles that describe the functionalities of a device must be formulated in way that is understandable by the system, e.g. by using OWL-S which is an OWL ontology to describe sematic services. However, some parts are not unambiguously specified, so in this work the extension of Tazari [7] is used which also shows some examples of service profiles to turn a light source on or off. Each service profile specifies the following concepts:

– *Service category*: a classification (here: sub classes of the class *Service*).
– *Input parameter*: the input parameter restricts the set of instances of the ontological model to which the result should be returned or applied. In case of the Lighting service this could, for example, be the URI of the light source that needs to be controlled.
– *Result*: a result is either a return value (which is not interesting in our application scenario) or an effect which somehow modifies the instances of the model. There are three type of effects which basically modify the underlying RDF graph: add, change, and remove. As this work is concerned with *controlling* devices, the *change* effect will be further investigated.

4.4 Service Filtering

After a selected device is determined and the service profiles have been retrieved, the functionalities could be offered to the user. However, some functionalities may be provided as special case by other functionalities. As mentioned before, turning a light source on equals setting its brightness to zero. Thus, the *dimming* function contains *turning on* and *turning off*. This can be used to reduce the number of functions that are offered to the user to decrease the mental load of the user and to allow for simple interaction since in some cases, the additional multiple choice question to select a service can be skipped.

To find out if one service can be realized by a different service, a matchmaking approach is used. Basically, the service profile specifies the set of all instances of the model that apply to the given change effect. In case of the switching function, we have only one instance to set the brightness to either 0 or 100. These two functions are obviously disjoint. However, the dimming service specifies the set of all possible brightness values from 0 to 100, inclusively. Thus, the matchmaking algorithm should provide two methods to determine whether (1) one set is a subset of another set and (2) one specific value is a member of a given set.

5 Realization

To demonstrate our approach, an implementation with the universAAL[5] platform was realized. This platform was chosen since it fulfills all the requirements. Position and pointing direction of the user is determined by the Kinect.

5.1 Architecture

The overall architecture is depicted in Fig. 2. All devices (including the Kinect) and their functionalities (if some functionalities are offered) are specified in a semantic way using ontologies, i.e. the location is available for every device. Each device may offer some services and provides some contextual events. Services are represented semantically and realize a functionality. For example, the lighting ontology can be realized by different low-level protocols (e.g. KNX); the calling application is not aware of the concrete realization. Instead, the service is called at an abstract level using OWL-S to control, for example, light sources, TVs, or blinds. The service provider component is responsible to register a valid service profile and call the low-level functions. Services are registered at a central registry inside the service broker. All communication is performed by the brokers. The application - in our case the user interface - can query the registered service

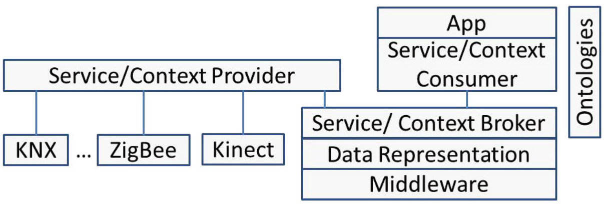


Fig. 2. Architecture

[5] http://www.universaal.org/

profiles from the service broker and can call the services through the service broker. The Data Representation building block provides the possibility to perform a basic matchmaking which enables the application to realize the filtering of services. The Kinect is in our implementation able to calculate the pointing direction (according to its location and orientation) and use the functionality provided by the platform to calculate all devices along this pointing direction. This information (see Fig. 1) is sent to the context broker and is consumed by the application.

5.2 Interaction

The service responsible for gesture recognition queries the information of the camera and waits for incoming skeleton data to perform the process of recognizing pointing gestures which are detected by analyzing the angle between the upper and lower arm. If this angle is beyond a certain threshold for a certain number of consecutive frames, the start of a pointing gesture is assumed. By taking the 3D point of the shoulder and the hand, a 3D cone relative to the enclosing room can be calculated. The location ontology supports the selection process by calculation all objects that intersect with this cone.

As soon as a device is selected, a graphical user interface is shown on the monitor next to the user (typically on the TV screen) to inform the user that s/he is going to control a certain device. In case of a multiple functionalities or multiple selected devices, a multiple choice question is presented to the user (see Fig. 3 (left)). Swiping horizontally either cancels or accepts the selected option. Although the selection process is implemented in case of multiple devices (e.g. window and blind that are at the same position in the environment), the filtering of functionalities is not yet integrated in this demonstrator. The level-based interaction is shown in Fig. 3 (right). Here, it is possible to control the current level of the device by moving the hand up or down. Again, the horizontal swipe cancels or accepts the setting. If the level is accepted, a service request is issued to the service broker to actually set the chosen value.

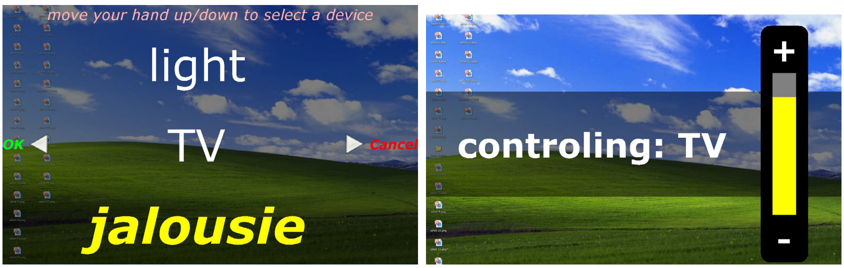


Fig. 3. Visual feedback: multiple choice question for disambiguation (left), controlling a device by moving the hand vertically to set a level, e.g. the volume of a TV (right)

6 Conclusion

This paper described the interaction with an intelligent environment to select and control devices with hand gestures. The Kinect is used to capture the position, pointing direction, and command gestures to identify a device of interest, select a certain functionality and control a *level*, e.g. the brightness of a lamp or the volume of a TV. All functionalities of the devices are available in form of semantic services. In case of multiple functionalities or multiple selected devices, a multiple choice question is presented to the user for disambiguation and determination of the intent of the user. To simplify the interaction, the list of functionalities is first semantically filtered to avoid functions that can be realized with other functions, e.g. turning a lamp on or off can be realized by a dimming function. In this case, the additional question for disambiguation can be avoided.

This concept was demonstrated by a realization that uses the semantic framework of the AAL platform universAAL. Future version will also use the UI framework to be independent of a concrete modality. Thus, the dialog can also be presented as audio instead of a graphical user interface, preventing the user from the need to be directed towards the TV to see the output of the system.

Acknowledgements. This work is partially financed by the European Commission under the FP7 IST Project *universAAL* (grand agreement FP7-247950).

References

1. den Bergh, M.V., Carton, D., Nijs, R.D., Mitsou, N., Landsiedel, C., Kuehnlenz, K., Wollherr, D., Gool, L.V., Buss, M.: Real-time 3d hand gesture interaction with a robot for understanding directions from humans. In: Proceedings of 20th IEEE International Symposium on Robot and Human Interactive Communication (2011)
2. Caon, M., Yue, Y.: Context-aware 3d gesture interaction based on multiple kinects. Applied Sciences (c), 7–12 (2011)
3. Hassani, A.: Touch versus in-air hand gestures: Evaluating the acceptance by seniors of human-robot interaction using microsoft kinect (December 2011)
4. Kim, H.J., Jeong, K.H., Kim, S.K., Han, T.D.: Ambient wall: Smart wall display interface which can be controlled by simple gesture for smart home. In: SIGGRAPH Asia 2011 Sketches, SA 2011, pp. 1:1–1:2. ACM, New York (2011)
5. Marinc, A., Stockloew, C., Tazari, S.: 3d interaction in aal environments based on ontologies. In: Wichert, R., Eberhardt, B. (eds.) Ambient Assisted Living. Advanced Technologies and Societal Change, pp. 289–302. Springer, Heidelberg (2012)
6. Nielsen, M., Störring, M., Moeslund, T.B., Granum, E.: A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In: Camurri, A., Volpe, G. (eds.) GW 2003. LNCS (LNAI), vol. 2915, pp. 409–420. Springer, Heidelberg (2004)
7. Tazari, S.: Using queries for semantic-based service utilization. In: Proceedings of SMR2 2009. CEUR Workshop Proceedings 525, p. 15 (2009)
8. Vanderhulst, G., Luyten, K., Coninx, K.: ReWiRe: Creating Interactive Pervasive Systems that cope with Changing Environments by Rewiring. In: Proceedings of the 4th IET International Conference on Intelligent Environments, IE 2008 (2008)

Understanding Complex Environments with the Feedforward Torch

Jo Vermeulen, Kris Luyten, and Karin Coninx

Hasselt University - tUL - IBBT,
Expertise Centre for Digital Media,
Wetenschapspark 2, B-3590 Diepenbeek, Belgium
{jo.vermeulen,kris.luyten,karin.coninx}@uhasselt.be

Abstract. In contrast with design flaws that occur in user interfaces, design flaws in physical spaces have a much higher cost and impact. Software is in fact fairly easy to change and update in contrast with legacy physical constructions where updating their physical appearance is often not an option. We present the Feedforward Torch, a mobile projection system that targets the augmentation of legacy hardware with feedforward information. Feedforward explains users what the results of their action will be, and can thus be seen as the opposite of feedback. A first user study suggests that providing feedforward in these environments could improve their usability.

Keywords: feedforward, intelligibility, mobile projection, legacy systems.

1 Introduction

Creating new usable ubicomp systems, especially systems that support walk-up-and-use scenario's, is covered in various facets in literature. While most literature discusses newly created systems and setups, we are interested in systems that are already present in our environment and are meant to be used over longer periods (e.g., 10 years and beyond). We start from our existing environments in which we reside on a daily basis. Our environment exposes many automated or computerized systems that were meant to be walk-up-and-use systems. These systems might not be context-aware or smart (i.e., are not pro-active nor use sensors for interaction), but their physical designs are tightly integrated in our environment and their usage is often part of our daily routines.

Fig. 1 shows three examples of legacy systems one can find in various public buildings. Fig. 1(a) shows a button that appears to trigger the power socket below for a predefined period of time. Nevertheless, using this button will also turn on the television and VCR in the same classroom. Fig. 1(b) presents three boxes that need to be controlled in case one detects fire. The left-most button's function is clear, but what do users have to do with the two others? Do you need a key? Is it safe to turn a key when it says "evacuate" below? What will happen in this case? Finally, Fig. 1(c) shows a terminal for recharging a contactless

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 312–319, 2012.

(a) Mysterious button

(b) Confusing fire alarm

(c) NFC recharging terminal

Fig. 1. Three examples of legacy systems that have design flaws that make them unsuitable for walk-up-and-use usage. All systems presented come from the same university building. People that frequent this environment tend to have difficulties using these systems because of unclear or missing feedforward information.

payment card using Near-Frequency Communication (NFC). Two cards need to be used: the contactless payment card and a debit card to transfer money from on to another. The sequence of steps and when to use each card is often a source of confusion. These are just three legacy systems that we found in the wild. While we were developing the Feedforward Torch, we collected many more examples of legacy systems that were designed for walk-up-and-use scenarios but cause problems for users as it is not immediately clear how to interact with them.

Due to the fact that the designs of these systems do not convey how they work, users have difficulties predicting the behaviour and the available features of the system. Moreover, it is often not clear to users how they can interact with these systems. For context-aware systems, Bellotti and Edwards [2] have proposed the concept of *intelligibility*. Intelligible systems have built-in support for helping users understand how they work. Intelligibility is an important feature for context-aware systems that take actions on the user's behalf, and rely on implicit input and complex inferencing. For non-smart systems, we tend to rely on "good design" to make sure users know how to interact with them. However, our environment is often comprised of a combination of legacy and new systems, due to budgetary considerations. While each of these systems might have been well-designed, their designers have often not considered these systems to be combined together. Since a combination of legacy and new systems tends to be seen as a temporary solution, it would often be too costly to rethink the entire physical design so that it better matches user's expectations. For this reason, we hypothesize that legacy systems (or combinations of legacy and newer systems) that cannot be altered, could also benefit from intelligibility to help users learn how to interact with them.

Our focus for augmenting the legacy systems is on *feedforward*, a specific type of intelligibility that tells users what will happen when they perform a certain action. Feedforward informs the user about what the result of an action will be, and can thus be seen as the opposite of feedback. Well-designed feedforward is an effective tool for bridging Norman's Gulf of Execution [4] – the gap between a user's goals for action and the means for executing those goals. Feedforward has been successfully applied in gesture-based interaction to help users learn, perform and remember gestures [1]. Additionally, Lim and Dey's "What if?"-questions [3] can be seen as a type of feedforward for context-aware systems. For physical interfaces, feedforward is often conveyed by the design and form-factor of the interface. However, if the design fails to convey feedforward information, this is very cumbersome and expensive to fix after deployment. Physically changing the interface design to include better feedforward would imply that every instance of the system needs to be fixed separately. For example, we found the fire alarm interface (shown in Fig. 1(b)) over 50 times in the same building.

In this paper, we present the *Feedforward Torch*, a combination of a mobile phone and mobile projector that provides feedforward information about different objects in their physical environment. Our solution augments the objects, more specifically legacy systems, during usage and does not require physical changes. We have currently built a prototype of the system and conducted a user study to investigate the suitability of this technique as a way to overcome design flaws of legacy systems.

2 Related Work

The possibility of augmenting physical environments using mobile projectors was first demonstrated by Raskar et al. with their iLamps [8] project. Earlier work [6] focused on steerable, ceiling-mounted projectors. Later, Raskar et al. extended these mobile projectors with RFID readers and photosensing capabilities to identify the physical objects that were being augmented [7]. In recent years, advances in hardware have enabled compact prototypes that can be embedded into smartphones, and different interaction possibilities have emerged [10].

The Feedforward Torch (see Sect. 3) is inspired by existing work on portable projectors. However, our contribution lies not in producing high-quality graphics on projected surfaces, or in interaction techniques. We rather explored how this setup can be used as a ubiquitous guidance system that helps users deal with legacy systems that suffer from design flaws.

Previously, Vermeulen et al. [11] investigated the use of steerable projectors to overlay an intelligent environment with real-time visualizations of actions occurring in this environment (e.g., lights that are turned on or off based on the presence of someone in the room). The Feedforward Torch serves a similar goal, but requires less infrastructure and allows users to control the object they require information about and when they need this information. Although we do not specifically focus on intelligent environments, we believe that showing feedforward through a mobile projector would also be useful for those environments.

3 The Feedforward Torch

The Feedforward Torch allows users to point at objects in their environment and reveal feedforward information about them, as if they were located under a spotlight. Users are shown under which conditions actions associated with the object will be executed by the system (e.g., a displacement in time or space), so that they can anticipate and adapt their behavior, if necessary. Animations are used to better convey the effect an action will have. The Feedforward Torch does not provide additional data for a physical object nor does it extend the features of a legacy system, its sole focus is on guiding the user to use the actual system.

The Feedforward Torch allows to project feedforward information on and around the system. Fig. 2 shows the Feedforward Torch prototype, consisting of a Samsung Galaxy S smart phone, a MicroVision SHOWWX$^+$ laser pico projector and a laser pointer to be able to point the device at physical objects. A custom casing was made in order to support one-handed interaction.

Fig. 2. The Feedforward Torch prototype

The Feedforward Torch provides an execution environment for various feedforward interfaces. The software can load configuration files describing (a) the environment, (b) the legacy systems within that environment, (c) available feedforward information for these systems and (d) properties that influence the type of presentation of the feedforward information. An example of the latter is whether the effect of an action happens after a delay and/or whether the effect will take place over a longer period of time. When the effect of an action will take place over a longer period of time the system will look for available animations that can communicate this to the user. A nice example of this can be found in Sect. 4, in Scenario 3 "the auditorium" where one has to lower the projection screen.

The combination of a mobile phone allows to offer both in- and out-of-context feedforward. For using the projection capabilities, the user needs to be co-located with the system and the system should be in the visual periphery of the user. But if the system is outside of the user's visual periphery, the screen of the mobile device can be used for presenting the feedforward information. The former implies the feedforward information is displayed “in-context”, while using the screen of the mobile device implies some parts of the system context is lost. Feedforward information on the screen of the mobile device is also important when using projection would not be suitable (e.g., outdoors).

4 User Study

4.1 Method

We conducted a small user study to assess (1) whether the feedforward torch allows users to better understand how to work with complex legacy systems and (2) whether visualizations and animations are preferred over textual descriptions.

Since we did not implement object recognition or 3D tracking, we used a Wizard-of-Oz approach to trigger the feedforward display. Fig. 3 shows the Wizard-of-Oz setup. The wizard is standing in the background to observe the participant and used a smartphone to control the Feedforward Torch. The wizard can use the smartphone UI to select the legacy system the participant is currently pointing at from a list of supported systems.

Fig. 3. The setup used for the Wizard-of-Oz study: the Feedforward Torch on the left and the control interface on the right

The Feedforward Torch was used by 7 participants (5 male, 2 female; 4 without and 3 with a technical background; ages ranged from 28 to 40, $\mu = 32.14$) in three different scenarios (Fig. 4):

- Scenario 1: “The television and the timer”. Participants were asked to turn on the TV in the room, and had to work around the timer that controlled the TV.
- Scenario 2: “The PingPing NFC terminal”. Participants were instructed to recharge their PingPing NFC card for the amount of 10 EUR. To do so, they had to use both their debit bank card as well as their PingPing card.

(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

Fig. 4. The three scenarios used for the study

- Scenario 3: "The auditorium". In this scenario, the objective was to prepare the auditorium for a presentation. This means the projector should be turned on, the projection screen should be lowered, and the lights should be dimmed.

After a short introduction of the Feedforward Torch, participants were given a specific goal they had to achieve in each of the three scenarios (e.g., turning on the TV). Each scenario took place in a different location. None of the participants were familiar with the different devices used in these scenarios.

Before participants started to explore how to complete the predefined goal, they were asked to describe to the observers how they thought the devices should be used for this purpose. Their assessment of the system was only based on its appearance and labels or signs already present in the physical space. Next, the participant used the Feedforward Torch to complete the assigned task.

When participants had performed the three tasks, we conducted semi-structured interviews in which we inquired them about the usefulness of the Feedforward Torch, and their preferences with respect to visualizations versus textual explanations. Moreover, they were asked in which situations mobile projection or the phone display was preferred.

4.2 Results

The Feedforward Torch helps users deal with complex situations All participants were able to complete the tasks using the Feedforward Torch. When asked about its usefulness, all participants mentioned they found the Feedforward Torch useful as a guide for complex situations. Several participants mentioned they would have been unable to complete the three scenarios without the Feedforward Torch or additional help from the experimenters. Two participants stated that the system would have come in handy when using the metro in a large city such as Paris or London: "*When I had to use the London Underground for the first time, it would have been useful to have a device like the Feedforward Torch to help me figure out how to use the ticketing machine. Now, I had to observe other passengers first before I knew how the system worked and what I had to do.*"

Visualizations were preferred over textual descriptions. Participants strongly preferred visualizations over textual explanations in the encountered scenarios, as they considered reading textual information to be more time-consuming. However, a number of users suggested providing detailed textual descriptions as an secondary source of information to complement the existing visualizations. As observed by Palmiter et al. [5], textual help may allow users to remember instructions more efficiently than demonstrations.

Animations are deemed useful in complex situations. Especially when the result of a certain action would happen over time or outside the user's periphery, participants appreciated the use of animations. During the user study, we used for example an animation of the projection screen coming down when the participants pressed the corresponding button.

Acceptance of Mobile Projection. The study revealed both advantages and disadvantages of mobile projection technology. Participants liked the fact that information was overlayed on the physical environment, so they did not have to switch between the phone display and the device they had to operate. One of the advantages of mobile projection that was mentioned during the semi-structured interviews was the fact that groups of people could explore the projection together. However, this could also cause privacy problems, in line with findings by Raskar et al. [8] and Holleis and Rukzio [9]. Another disadvantage was the difficulty of using mobile projection in low-lighting conditions. There was no clear preference for mobile projection, although we do expect hardware advancements to further improve the user experience. Based on these results, we feel that using an augmented reality approach for showing feedforward information is valuable. Although we currently use a mobile projector for this purpose, other technologies such as wearable devices (e.g., Google's Project Glass[1]) are also possible.

5 Discussion

We have presented the Feedforward Torch which overlays objects in a physical environment with feedforward information using a mobile projector. Based on a first user study, we feel the Feedforward Torch can help users to interact with complex devices in their environments and overcome design flaws in legacy systems. We feel this work opens up interesting avenues for further research.

First, our current prototype is not a fully working system, but is implemented using the Wizard-of-Oz technique because we were mainly interested in exploring whether users would find the Feedforward Torch helpful. A fully working implementation would need to be able to recognize different objects of interest in the environment. Existing systems have used a combination of a projector and camera to know what the user is pointing at [8,10], although the use of technologies such as QR codes or NFC tags could also prove to be useful.

Secondly, we believe it would be interesting to empower users to create feedforward information for the different objects in their environments themselves. This is similar to how people tend to augment complex devices with instructions

[1] https://plus.google.com/+projectglass/

written on labels or post-its. The success of websites such as instructables.com[2] could suggest that users might be willing to do this. The Feedforward Torch could then reveal these user-made feedforward elements when pointed at the corresponding object. Finally, a larger study would need to be performed in order to provide conclusive results on the suitability of the way users interact with the Feedforward Torch.

Acknowledgements. We warmly thank Gert Vos for building the prototype of the Feedforward Torch and conducting the user study described in this paper in the context of his Master's thesis.

References

1. Bau, O., Mackay, W.E.: OctoPocus: a dynamic guide for learning gesture-based command sets. In: Proc. UIST 2008, pp. 37–46. ACM (2008)
2. Bellotti, V., Edwards, W.K.: Intelligibility and accountability: human considerations in context-aware systems. Hum.-Comput. Interact. 16(2), 193–212 (2001)
3. Lim, B.Y., Dey, A.K.: Toolkit to support intelligibility in context-aware applications. In: Proc. Ubicomp 2010, pp. 13–22. ACM (2010)
4. Norman, D.A.: The Design of Everyday Things. Basic Books (2002)
5. Palmiter, S., Elkerton, J.: Animated demonstrations for learning procedural computer-based tasks. Hum.-Comput. Interact. 8(3), 193–216 (1993)
6. Pinhanez, C.: The Everywhere Displays Projector: A Device to Create Ubiquitous Graphical Interfaces. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 315–331. Springer, Heidelberg (2001)
7. Raskar, R., Beardsley, P., van Baar, J., Wang, Y., Dietz, P., Lee, J., Leigh, D., Willwacher, T.: Rfig lamps: interacting with a self-describing world via photosensing wireless tags and projectors. In: SIGGRAPH 2005 Courses. ACM (2005)
8. Raskar, R., van Baar, J., Beardsley, P., Willwacher, T., Rao, S., Forlines, C.: ilamps: geometrically aware and self-configuring projectors. In: Proc. SIGGRAPH 2003, pp. 809–818. ACM (2003)
9. Rukzio, E., Holleis, P.: Projector phone interactions: Design space and survey. In: PPD 2010 Workshop (2010)
10. Rukzio, E., Holleis, P., Gellersen, H.: Personal projectors for pervasive computing. IEEE Pervasive Computing 11, 30–37 (2012)
11. Vermeulen, J., Slenders, J., Luyten, K., Coninx, K.: I Bet You Look Good on the Wall: Making the Invisible Computer Visible. In: Tscheligi, M., de Ruyter, B., Markopoulus, P., Wichert, R., Mirlacher, T., Meschterjakov, A., Reitberger, W. (eds.) AmI 2009. LNCS, vol. 5859, pp. 196–205. Springer, Heidelberg (2009)

[2] http://www.instructables.com/

Open Objects for Ambient Intelligence

Paulo Ricca and Kostas Stathis

Royal Holloway, University of London, UK
{paulo.ricca,kostas.stathis}@cs.rhul.ac.uk
http://www.cs.rhul.ac.uk

Abstract. We present the Open Object, a framework for distributing capabilities over a system of inter-connected physical objects. We focus on allowing lightweight objects to not only share their capabilities with other objects but also to outsource capabilities when needed, in order to fulfil a user's goal. We exemplify our approach with a smart home scenario and a service-oriented implementation.

Keywords: Ubiquitous Computing, End-User Development, Events, Behaviours, Capabilities, Workflows, Service-Oriented Architecture.

1 Introduction

As technological advances produce smaller and cheaper electronics, we get closer to achieving the Ubiquitous Computing vision [16]. Although concepts such as the Internet of Things [8] attempt to implement this vision on many different kinds of objects, most of the approaches rely on mediated and, therefore, centralised models of interactions between physical objects. Mediated approaches, however, are not only less tolerant and averse to scaling [12], but also limit more open, dynamic, ad-hoc and network-less [11] interactions. Ideas such as [4] focus on coordination of resource-constrained devices by moving the software processes to more capable devices, which may result in under-exploitation of the smaller devices' capabilities. Although there is some research around decentralised service composition([3][1][2]), it is mainly focused on computationally heavy architectures, less suitable for embedded systems. [6] describes an approach and a series of standards for building ubiquitous applications over a network of self-managed cells (SMC's). While their work and this are related in some aspects, namely a similar system workflow, SMC's need a full implementation of the system architecture in order to work together, making it less useful for working seamlessly with very lightweight objects. There is, therefore, a lack of research in the current landscape, in what respects to end-user oriented service composition on lightweight and heterogenous decentralised environments, specifically those where constant connectivity/proximity between objects cannot be taken for granted.

The concept of Meta-Design [7] proposes the possibility of end-users participating in the process of designing the functionality of their objects. We aim at laying the foundations of a service-oriented framework targeting a further

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 320–327, 2012.

support of end-user development [10] over a system of interconnected Open Objects. Our main objective is to propose a service-oriented approach for the implementation of a decentralised environment of lightweight smart objects which collaborate by sharing and, at times, outsourcing capabilities to meet a user's goal. We are interested in exploring decentralisation on very constrained, ad-hoc or network-less interactions.

As an example, we choose a smart home environment scenario in which the user creates a new functionality involving an alarm clock and a bread machine. In the scenario, a user wishes the bread maker to be scheduled for the same time as his alarm clock is set. By combining the functionality of the two objects and embedding the user's intentions in them, the user avoids repeating the monotonous task of setting the alarm clock and the bread maker at the same time. We walk through the scenario as we define the foundations of the framework in the following sections. The scenario can easily scale up in terms of number and complexity of objects, however, we focus on simplicity to more easily and concisely demonstrate the features of the framework. Our main contribution is the design of a very lightweight, behaviour-based architecture, whose flexibility is illustrated via a prototype implementation of heterogenous interacting objects.

2 The Open Object

An *Open Object* is any physical object capable of capturing events from the environment in which it is situated, process them, and generate new events that allow it to interact with other objects and the user. The object is open because it can be augmented dynamically with new computational capabilities that allow the object to adapt its behaviour in order to become useful in situations that were not predicted at design time. Objects' capabilities are grouped together to form *behaviours* that the object can display in interactions with other objects, in order to serve a certain *purpose*. An Open Object is one that is able to share its capabilities with other objects but also to outsource capabilities when needed.

An Open Object environment consists of many interacting objects. We use four notions to support these interactions: **Events, Capabilities, Behaviours and Purposes**.

An **Event** is a description of a happening and it is said to belong to an Event Type [9], which classifies the happening. Event Types denote patterns and are useful in the way that they allow objects to respond to them in standard ways. For example, in our scenario, the two objects are able to produce a set of events: the alarm clock produces *Time Was Changed* and *Alarm Went Off*, and the bread maker produces *Bread Ready*, *New Schedule* and *No Ingredients*.

A **Capability** is the inherent sensing, actuation or processing ability of an object to perform an action. It may be internal - only accessible by the object - or external - accessible by other objects. A capability may be considered *atomic* (lower-level in nature) or *complex* (composed from atomic capabilities, orchestrated at a higher level). Capabilities can also be classified according to their origin. In this sense they can be *innate*, when they were hard-coded and embedded into the object at design-time, or *acquired*, when, during the life-time of the object, a new functionality was added to it. New functionalities can be added by, for example, the user when orchestrating the object's existing capabilities to form a new, complex one, or the object itself by downloading new firmware. For example, in our scenario, the alarm clock contains *Set/Get Time*, *Ring* and *Turn On/Off* capabilities, and the bread maker has *Set/Get Schedule* and *Turn On/Off* capabilities.

Behaviours and Purposes facilitate interactions by describing and generalising common procedures and produce an expectation on the object. The behaviour of an Open Object for a given purpose is the aggregation of all the capabilities and events that the object exposes to environment to serve that purpose. As an open object can serve different purposes, it can exhibit different behaviours.

Behaviours also allow each object to focus on fewer tasks, and interact more effectively. Objects can, therefore, *display* a set of behaviours to the outside environment, which in turn, *serve* designated purposes when interacting with other objects. In the scenario, the two objects interact through different behaviours, each serving a different purpose. One displays the *"alarm clock behaviour"* and the other the *"bread producer behaviour"*. Other objects, such as a Hi-Fi or a mobile phone may display the "alarm clock behaviour", as they may also have the right capabilities for that purpose (e.g. time keeping and a speaker). Just as well, a web-service connected to a bakery may also display the "bread producer behaviour" if it allows the clients to schedule target delivery times through electronic communications. These behaviours describe the necessary capabilities and events to serve its purpose, and imply a certain expectation for each of these.

We classify behaviours in three categories. **System Behaviours** manage the system itself and include tasks such as coordination and service registration. **Domain Independent Behaviours** offer generic support that the user may use to support his workflows. **Domain Dependent Behaviours** are behaviours whose domain the user wants to control (e.g. his electrical affordances).

3 An Ambient of Interacting Open Objects

Objects interact and collaborate to achieve the user's goal, sharing their capabilities and forming an intelligent ambient of smart objects. We view **Collaboration** as a set of objects working together to achieve a goal. In our scenario, the goal of the user is that the bread is ready when the alarm. In our framework goals of this kind are specified within a collaboration as **Workflow**. A workflow automates procedures according to a pre-defined set of rules, to achieve an overall goal [15]. In a collaboration, object's behaviours serve different purposes defined in the workflow. Workflows may also contain instructions on behaviour cardinality and on how to assign purposes to objects' behaviours. We have identified three different kinds of workflows. **Application Workflows** form the central behaviour of the system, from the user's point of view. **Capability Workflows** define the orchestration behaviour of complex capabilities. **System**

Workflows coordinate the system's activities and are implicitly defined as System Behaviours. Each object may play and implement one, several or none of the System Behaviours. This allows very lightweight objects to outsource the system components that coordinate the distribution system itself. The interactions between these behaviours are depicted in Figure 1 and described below.

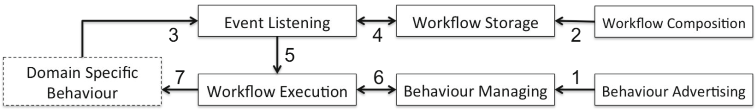


Fig. 1. System Workflow and System Behaviour Interactions

The **Behaviour Managing** assigns and keeps an up-to-date map between workflow purposes and objects' behaviours. The **Behaviour Advertising** advertises each object's behaviours. The **Workflow Composition**, is displayed by the object (i.e. a computer, a mobile phone or a web-service) that allows the user to design, edit and/or download existing workflows to configure the behaviour of objects, using a end-user development tool. The **Event Listening** triggers the execution of a workflow by listening to events (step 3), querying the Workflow Storage for suitable workflows (step 4) and, in the existence of one, requesting its execution from the Workflow Execution (step 5). Event Listeners may hold and share a registry (such as [12]) of past events to allow objects that arrive at the environment after an event has happened, to still react to it effectively. The **Workflow Storage** stores and provides workflows. Each workflow is mapped to the event(s) it related to. The **Workflow Execution** carries out the execution of workflows. The System Workflow regulates in the interactions in the **System Life-Cycle**, which can be decided into four steps: **Discovery** (step 1) is the process of registering behaviours in the Behaviour Manager; **Choreography** (step 2) is the process of creating and registering a workflow rule, carried out by the Workflow Composing and the Workflow Storage; **Purpose Assignment** (step 6) is performed by the Behaviour Manager. Behaviour Managing uses an internal logic when assigning purposes. It may, for example, keep a record of object performance or use a more complex reputation trust model such as [14], according to the metrics specified in the workflow, if existent; **Execution** is triggered by the Event Listening, which, after listening to an event (step 3), queries the Workflow Storage about appropriate workflows (step 4), and passes it on to the Workflow Execution (step 5). The workflow contains references to purposes, which are assigned to objects' behaviours by the Behaviour Manager (step 6). This process happens throughout the execution, and a purpose may be re-assigned due to, for example, an object becoming offline. When assigned, an object is requested to perform the appropriate action, defined by the workflow (step 7). Objects may produce events (step 3) completing the system life-cycle.

4 Open Object Development

Behaviours are a fundamental part of the Open Object framework, thus behaviour advertising and definition is a crucial part of the system workflow. When queried, an object's Behaviour Advertising reports on his behaviours and on where to get their definitions. In our implementation we defined a JSON[1] syntax for *Behaviour Reporting* and another for *Behaviour Definition*. Behaviour Reporting is a simple mapping from behaviour names to behaviour definition URI's. Keywords in capital letters are meant to be replaced by the respective identifiers and values (e.g. BEHAVIOUR_NAME is replaced by a behaviour identifier).

```
{behaviours:[{name: BEHAVIOUR_NAME, definition: DEF_URI}]}
```

The betaviour definition URI points to the virtual location of the behaviour definition defining a list of capabilities, events and characteristics, as well as information that assists the user during composition. Characteristics allow finer-grain over priorities on purpose assignment.

```
{ name: BEHAVIOUR_NAME, description: DESCRIPTION,
  capabilities:
   [ { name: CAP_NAME, description: DESCRIPTION,
     parameters:
      [ { name: PARAM_NAME,
          type: PARAM_TYPE,
          description: DESCRIPTION,
          optional: TRUE/FALSE } } ] ],
    events: [
       { name: EVENT_NAME, description: DESCRIPTION } ],
    characteristics: [
       { name: CHARACTERISTIC_NAME,
         description: DESCRIPTION,
         unit: UNIT_OF_MEASUREMENT } ] }
```

Workflows on an Open Object environment are defined as process trees [13] (we chose to use a tree structure instead of a graph as it is more lightweight implementation friendly, leaving complex graph-like behaviour to be implemented into dedicated capabilities). A generic workflow, shown in figure 2, begins with a start event, performs a series of processes and ends. In practice, each process is a single call to a behaviour's capability, executed by the respective assigned object.

When calling a capability, it may require and/or give the option of a set of parameters. A parameter can be left empty, if optional, be set to a constant hard-coded value or linked to the evaluation of a sub-process. Sub-workflows are created with trees of linked processes.

Differing from the well known Event-Condition-Action model, we consider conditions, loops, splits and other base elements of common workflow

[1] http://www.json.org

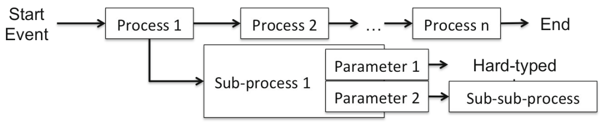


Fig. 2. A generic workflow process tree with an example sub-process with parameters

patterns [18] to be processes performed by capabilities. This Event-Process Tree allows a flexible yet powerful partial implementation of language structures on very constrained objects, allowing them to work seamlessly with more complex ones, without the need for bridges or proxies. When executing a workflow, an object may outsource other object's or a remote server's capabilities to evaluate branches of tree which use capabilities that it does not implement. Control over branch evaluation is left to the parent process to allows flexible behaviour, such as when branches are evaluate more than once or asynchronously.

Our JSON Workflow Syntax follows the following structure. The root of the workflow specifies an event and the branches translate into capability requests. A similar structure is used to define complex capabilities, using *"capability: BEHAVIOUR CAPABILITY"* instead of *"when: PURPOSE EVENT"*.

```
{ when: PURPOSE EVENT,
   do:
   [
      { request: PURPOSE CAPABILITY,
        ARGUMENT_NAME: ARGUMENT_VALUE OR REQUEST } ] }
```

5 Implementation Issues

Walking through the scenario, the user starts up the object configuration app on his phone and it enters a service discovery process where the objects tells the app which behaviours they are displaying. The user uses a visual tool on the app to write a rule that relates an event on the alarm clock to the alarm clocks and the bread makers capabilities, and exports it to the bread maker. The exported rule reads:

```
{ when: alarm_clock time_was_changed,
   do: {
      request: bread_producer schedule_time,
      time:
      { request: alarm_clock get_alarm_time } } }
```

When the user sets his alarm clock to 7:00 a.m. the alarm clock broadcasts a "Time Was Changed" event on the local network. The bread maker picks up the event, its Event Managing checks if there is a workflow for such event and requests its execution its from its Rule Execution component. While executing the rule, the Rule Execution finds a purpose called *bread_producer* and checks with the Behaviour Manager which object should serve this purpose. As it turns out, its himself who should do it, therefore it calls its own *schedule_time* capability, to be set as the time defined by the *time argument*, which is linked to another capability of the *alarm_clock* purpose, which the Behaviour Manager indicates to be played by the alarm clock object. The Rule Execution then calls the *get_alarm_time* capability of the alarm clock, which responds with the time the alarm was set, and the value is fed into the *time* argument. The *schedule_time* capability can now be complete, and a new time is schedule, for producing bread. In the morning, the user wakes up with the smell of freshly baked bread. On weekends, the alarm clock is set to go off a bit later, but thats no problem, the rule the user specified still applies and he gets bread right on time for his late weekend breakfast. The same workflow can be stored on the user's mobile phone so that, when he is on holidays and if the user wishes to, the wake up call service may serve the purpose of the alarm clock, and request breakfast in the room for the alarm time, if the hotel uses an Open Object framework for its services.

We have implemented the Open Object framework by following a Service-Oriented approach, considering behaviours and capabilities as Rest [17] resources. We developed a prototype similar to the scenario, in which we have two distinct objects: a laptop simulating the alarm clock and a Nanode[2] microcontroller (a network enabled Arduino[3] clone), simulating the bread machine, which interact through a series of system, domain independent and domain dependant behaviours. When the Arduino is online, whenever there is a *time_was_changed* event from the laptop, the laptop assigns the *bread_producer* purpose to the Nanode and requests its *schedule_time* capability with the time parameter set as the result of the *get_alarm_time* capability.

When the Arduino is manually turned off or, for some reason loses contact with the laptop, even if during the execution of the cycle, the *bread_producer* purpose is immediately passed on to the laptop (which also displays the respective behaviour) and it carries on with the task, as expected. The prototype demonstrates the feasibility of the Open Object framework in two widely different platforms. For prototyping reasons, The Behaviour Managing URI was hardcoded on both objects, to avoid the need of complex service discovery.

6 Concluding Remarks

We have described a a lightweight, behaviour-based framework aimed at end-user Open Object choreography on ambient intelligence environments, as well as an demonstrating implementation. There are two key innovations presented in this paper. Firstly, we showed an approach that allows an end-user to introduce new functionalities to Open Objects. Secondly, we introduced a decentralised approach for collaboration on ad-hoc and/or network-less environments, which allows dynamic context-aware interactions between users and objects.

Future work includes further, more objective evaluation of the present models, the research and evaluation of an appropriate user interface for a Workflow

[2] http://www.nanode.eu
[3] http://www.arduino.cc

Composer, better workflow regulation supported by policies, and user field trials with real-world Open Objects. We also want to explore the concept of hierarchies in behaviour-based systems, to allow for a more flexible composition of object's capabilities. Security and Privacy are important aspects of a system of this kind. Although these are out of the scope of this paper, they represent a challenge that needs to be addressed by future work.

References

1. Agarwal, V., Chaße, G., Dasgupta, K., Karnik, N., Kumar, A., Mittal, S., Srivastava, B.: Synthy: A system for end to end composition of web services. J. Web Semantics (2005)
2. Chafle, G., Chandra, S., Mann, V., Nanda, M.: Decentralized Orchestration of Composite Web Services. In: International World Wide Web Conference (2004)
3. Cottenier, T., Elrad, T.: Dynamic and decentralized service composition with aspect-sensitive services. In: Proc. of the 1st International Conference on Web Information Systems and Technologies (WEBIST), Miami, FL, USA (2005)
4. Dipsis, N., Stathis, K.: EVATAR – A Prototyping Middleware Embodying Virtual Agents to Autonomous Robots. In: Augusto, J.C., Corchado, J.M., Novais, P., Analide, C. (eds.) ISAmI 2010. AISC, vol. 72, pp. 167–175. Springer, Heidelberg (2010)
5. Dillon, T.S., Wu, C., Chang, E.: Reference Architectural Styles for Service-Oriented Computing. In: Li, K., Jesshope, C., Jin, H., Gaudiot, J.-L. (eds.) NPC 2007. LNCS, vol. 4672, pp. 543–555. Springer, Heidelberg (2007)
6. Filho, A.: Supporting Management Interaction and Composition of Self-Managed Cells. PhD Thesis. Imperial College, University of London (2009)
7. Fischer, G., Giaccardi, E.: Meta-design: A Framework for the Future of End-User Development. Human Computer Interaction Series (2006)
8. Gershenfeld, N., Krikorian, R., Cohen, D.: The internet of things. Scientific American, 291–294 (2004)
9. Galton, A., Augusto, J.C.: Two approaches to event definition. In: Int Conference on Database and Expert Systems Applications, pp. 547–556 (2002)
10. Lieberman, H., Patern, F., Wulf, V.: End-User Development, p. 492. Springer, Dordrecht (2006)
11. Miorandi, D.: Networkless networking. In: Proceedings of the 2nd International Conference on Autonomic Computing and Communication Systems (Autonomics 2008) (2008)
12. Ricca, P., Stathis, K., Peach, N.: A Lightweight Service Registry for Unstable Ad-Hoc Networks. In: Keyson, D.V., Maher, M.L., Streitz, N., Cheok, A., Augusto, J.C., Wichert, R., Englebienne, G., Aghajan, H., Kröse, B.J.A. (eds.) AmI 2011. LNCS, vol. 7040, pp. 136–140. Springer, Heidelberg (2011)
13. Vanhatalo, J., Völzer, H., Koehler, J.: The Refined Process Structure Tree. In: Dumas, M., Reichert, M., Shan, M.-C. (eds.) BPM 2008. LNCS, vol. 5240, pp. 100–115. Springer, Heidelberg (2008)
14. Wang, Y.: Trust and reputation model in peer-to-peer networks. In: Third International Conference on Peer-to-Peer Computing (2003)
15. Coalition, W.M.: Workflow Reference Model. Brussels (1994)
16. Weiser, M,: The computer for the 21st century. Sci. Amer. (1991)
17. Wilde, E.: Putting Things to REST. School of Information, UC Berkeley (2007)
18. van der Aalst, W.M.P., ter Hofstede, A.H.M., Kiepuszewski, B., Barros, A.P.: Workflow patterns. Distrib. Parallel Databases 14 (2003)

Towards Accessibility in Ambient Intelligence Environments

George Margetis[1], Margherita Antona[1], Stavroula Ntoa[1], and Constantine Stephanidis[1,2]

[1] Foundation for Research and Technology – Hellas (FORTH)
Institute of Computer Science
N. Plastira 100, Vassilika Vouton,
GR-700 13 Heraklion, Crete, Greece
{gmarget,antona,stant,cs}@ics.forth.gr
[2] University of Crete, Department of Computer Science

Abstract. This paper aims to set a landscape and elaborate a roadmap for accessible user interaction in AmI environments, including and beyond personal computational devices (PCs, mobiles, etc.), by identifying and addressing the new needs that emerge in the above context. This line of work is currently being pursued in the context of the AmI Research Programme of ICS-FORTH, and will be implemented in the new AmI Research Facility, which provides an ideal environment for developing the proposed solutions and a test-bed for validating them in a realistic simulation environment.

Keywords: ambient intelligence, accessibility, universal access, assistive technologies, personalized interaction, user model.

1 Introduction

In the years ahead, the potential of Ambient Intelligence (AmI) environments to address older and disabled people's everyday life needs is expected to have a radical impact on independent living and e-Inclusion. Many applications and services are already becoming available, for example in the domain of Ambient Assisted Living (AAL), which address a wide variety of issues critical for older and disabled people and are targeted to make possible and enjoyable a more independent, active and healthy life. The European strategy in ICT for Ageing Well of 2010 [1] identifies a number of ICT solutions addressing daily and independent living in areas such as social communication, daily shopping, travel, social life, public services, safety, reminders, telecare and telemedicine, personal health systems, and support for people with cognitive problems and their carers. The same report also points out the importance of user interaction for ICT solutions, and mentions user-friendly interfaces for all sorts of equipment in the home and outside, taking into account that many older people have impairments in vision, hearing, mobility or dexterity. Clearly, the benefits of AmI environments can only be fully achieved and accepted by their target end-users if such technologies can demonstrably be developed in such a way as to guarantee inclusive accessibility for a wide variety of functional limitations brought about by age or disabilities.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 328–337, 2012.

Accessibility in the context of individual applications and services has been defined as follows: for each task a user has to accomplish through an interactive system, and taking into account specific functional limitations and abilities, as well as other relevant contextual factors, there is a sequence of input and output actions which leads to successful task accomplishment [2]. However, the accessibility of AmI environments poses different problems and is more complex than currently available approaches to the accessibility of desktop or web applications and services, as AmI environments do not simply introduce a new technology, but an integrated set of technologies. Different levels of accessibility may be distinguished. A first level concerns accessibility of individual devices. Interactive devices need to be accessible to their owners according to their needs, but basic accessibility should also be provided for other users with potentially different needs. A second level concerns the accessibility of the environment as a whole, intended as equivalent access to content and functions for users with diverse characteristics, not necessarily through the same devices, but through a set of dynamic interaction options integrated in the environment.

It is likely that some of the built-in features of AmI environments, such as multimodality, will facilitate the provision of solutions that will be accessible by design ([3], [4]). For example, blind users will benefit from the wider availability of voice input and output. Different modalities can be used concurrently, so as to increase the quantity of information made available or present the same information in different contexts, or redundantly, to address different interaction channels, both to reinforce a particular piece of information or to cater for the different abilities of users. A novel aspect is that in AmI environments, the accessibility of the physical and of the virtual world need to be combined. For example, for blind, visually impaired and motor-impaired users, requirements related to interaction need to be combined with requirements related to physical navigation in the interactive environment.

Although several interaction technologies, such as, for example voice output, are already widely available, and other, such as, for example, eye-tracking, are reaching a maturity stage where they can be robustly exploited for accessibility purposes, a number of fundamental obstacles still hinders the provision of alternative and personalized accessibility solutions in AmI environments. These include:

- Limited knowledge of user requirements and of the appropriateness of different solutions for different combinations of user characteristics / functional limitations and environment characteristics / functions (for example, age-related factors).
- Lack of ready-to-use accessibility solutions supporting alternative interaction techniques for various combinations of user abilities / functional limitations. Even when optimal combinations of interaction devices and techniques are known, embedding them in AmI environments is far from easy, since most existing assistive technologies are limited in use to specific devices, and cannot be easily made compatible with complex environments including various devices.
- Lack of architectural frameworks, taking into account the need for practical accessibility solutions and supporting their integration and management.
- Lack of tools supporting various phases of the development lifecycle of accessible smart environments (e.g., requirements analysis, design and prototyping, evaluation, content creation).

As a result, developing truly accessible AmI environments is currently very expensive in terms of time, efforts, costs and required knowledge, and the results are often of limited flexibility and reusability in terms of accessibility solutions and addressed target user groups.

This paper aims to set a landscape and elaborate a roadmap towards Universally Accessible AmI environments, providing on one hand a user centered but also context-aware methodology for enabling accessibility in such environments and on the other hand modern tools and personalized assistive solutions that will constitute the building blocks for the development of independent living AmI environments addressing interaction needs of older and disabled persons.

This line of work is currently being pursued in the context of the AmI Research Programme of ICS-FORTH, and will be implemented in the new AmI Research Facility, which provides an ideal environment for developing the proposed solutions and a test-bed for validating them in a realistic simulation environment.

2 Related Work

Accessibility in the context of AmI environments is usually intended as inclusive mainstream product design (e.g., [5]), although a contextual definition is not yet available. AmI environments are expected to have profound consequences on the type, content and functionality of the emerging products and services, as well as on the way people will interact with them, bringing about multiple new requirements (e.g., [6], [7], [8]).

The issue of accessible interaction in AmI environments has been mainly explored so far through scenarios [9], while the impact of AmI environments on users with activity limitations has been studied in [10]. In [11] an overview is provided of EC-funded projects which include some concepts related to accessibility in AmI environments, although mainly focusing on services and applications set up in order to support people (including people with activity limitations). In [12], an AmI home care approach is presented aiming to address elderly and disabled persons' needs at home, mainly through smart users' monitoring in order for the proposed system to be able to notify on time their carers or doctors. An example of the AmI potential on helping people with particular needs in their everyday living is presented in [13]. Its major goal is the training of elderly people in order to handle modern interfaces for Assisted Living and evaluate the usability and suitability of these interfaces. A more systematic approach towards interactive personalization of ambient assisted living environments is presented in [14]. In more detail, a framework is presented providing interactive configuration of comprehensive AAL Environments at the level of authoring tools, focusing on the application of AAL at home.

Despite the aforementioned efforts towards supporting accessibility in AmI environments, a systematic and comprehensive approach to inclusive accessibility in AmI environments is still needed. This paper proposes some basic steps towards a radical new dimension of accessible user interaction in AmI environments for older and disabled inhabitants. The main idea behind the proposed roadmap is to make easy available, usable and "pluggable" in the intelligent environment mature and emerging assistive technology solutions, as well as alternative multimodal interaction techniques, applying the appropriate input and output modalities according to the user, the task at hand, and the current context of use.

3 The AmI Research Facility

ICS-FORTH is creating a state-of-the-art AmI Research Facility, targeted to support research, experimentation and multidisciplinary scientific collaboration. Such a facility is intended, amongst other things, to support the establishment and conduct of a line of research targeted towards the provision of accessibility in AmI technologies and environments. The Facility occupies a three-floor 3,000 square meters building (see **Fig. 1**), and comprises simulated AmI environments and support spaces, including as a house, a class-room, a collaborative workplace, a multifunctional doctor's office, and a multipurpose "exhibition" space.

Fig. 1. ICS-FORTH AmI Facility

The entire building has been designed to be accessible by people with disabilities, and follows DfA guidelines concerning stairs, elevators, ramps, corridors width, accessible bath-room facilities, multimodal signs and labels, etc. In particular, the house simulator will constitute a prototype accessible house for disabled and elderly people. Additionally, the building design takes into account issues of easy orientation and navigation in the physical environment. Designed to accommodate and simulate everyday life environments and to provide the necessary flexibility for testing news technologies, the AmI Facility will constitute a real test bed for accessible AmI spaces.

4 Indicative Scenario

The following is an indicative scenario which illustrates some basic requirements for accessible AmI environments.

Mary is a middle age blind teacher. She has just moved to her new AmI home where she lives alone. Today, she has a day off and she decides to spend most of her time listening to her favorite music. She moves to the smart living room using the “magic wand”, a plastic stick which can control the environment’s devices. Using the wand, she starts pointing to diverse directions. Every time she points to an artifact of the smart living room, audio feedback explains which is the artifact she pointed at, and what is its state.

After a few tries, she finds out the direction to the living room couch and starts moving towards it. When she approaches the couch, the smart living room informs her how close she is and if there is an obstacle in her way. It is important to note that the wand can provide similar information for every smart environment Mary may visit, thus making unfamiliar environments easier for her to explore.

Mary reaches the couch, sits down and then uses the “magic wand” again to find the direction of the smart audio system in order to listen to her favorite music. As soon as she finds it, she starts using the set of four buttons on the wand, with which she is able to control any device in the room in a seamless way. Using the “universal control” functionality of the “magic wand” she can interact with the display to choose the music category she wishes to listen to, as well as specific songs. The system constantly provides audio feedback for the available options and the interaction carried out. After selecting the list of songs she likes to listen to, she chooses the play function and sits back to the couch, relaxing and enjoying the music.

5 Roadmap Towards Accessible AmI Environments

In the context of the scenario outlined above, several challenges need to be addressed in order to elaborate a systematic approach to accessibility in AmI environments:

- Advancing knowledge of user requirements and of the appropriateness of different solutions for different combinations of user characteristics / functional limitations and environment characteristics / functions, and creating related ontological models
- Developing a reference architectural model that will address user needs for inclusive design in AmI environments, allowing for accessible multi – modal interaction
- Providing ready-to-use accessibility solutions supporting alternative interaction techniques for various combinations of user abilities / functional limitations
- Developing a design tool supporting the implementation of accessible AmI environments
- Developing AmI accessible applications in three fundamental everyday life domain, namely home, work and self-care.
- Evaluating the developed assistive solutions tools and applications in order to assess their accessibility, usability and added value for the target users.

The above challenges are further analyzed in the next subsections.

5.1 Ontology-Based User and Interaction Modeling

An extensible context ontology [15] for user and interaction techniques modeling in Ambient Intelligence environments is necessary in order to effectively address the needs for context-aware personalization in AmI environments. To this end, building on existing knowledge and best practices, a generic ontology framework for personalized assistive solution modeling needs to support the interconnection of heterogeneous domain ontologies, integrate single services using ontological layering and provide contents-aware update and maintenance mechanisms. This framework is intended to provide a simplified and highly abstract model of ontology which is independent of a specific ontology representation language and operates with ontologies on a conceptual rather than syntactic level, thus supporting compatibility and interoperability with ontologies described in diverse language forms, and ensuring significant advantages in simplicity of software development.

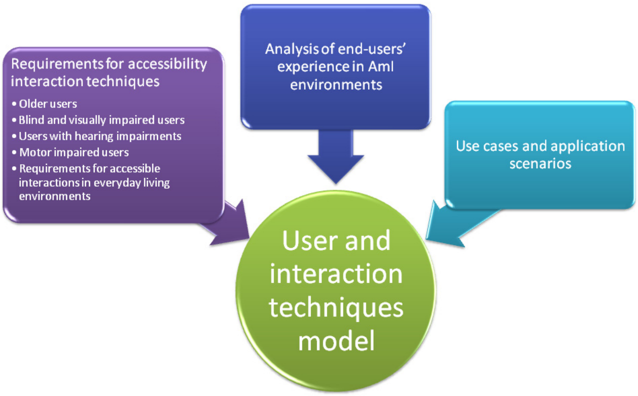


Fig. 2. User and interaction techniques model's implementation steps

Fig. 2 illustrates the steps to be followed in order to implement the ontology based framework. In more details, requirements elicitation for accessibility interaction techniques, analysis of end-users' experience on AmI environments and use cases and application development will be conducted. The consolidate results of these activities will constitute the input for the design and implementation of the aforementioned user and interaction techniques model.

5.2 Ready-to-Use Accessibility Solutions

In the context of accessible AmI environments, the assistive technologies that are currently used to make interactive applications and services accessible to disabled users will continue to constitute the basis for users' interaction, but they will also be extended and enriched in terms of functionality and use in order to provide multimodal and personalized accessible interaction beyond conventional ICT environments.

Fig. 3 illustrates the main categories of assistive technologies and interaction techniques that constitute the state of the art today and offer significant potential of providing accessibility means in AmI environments.

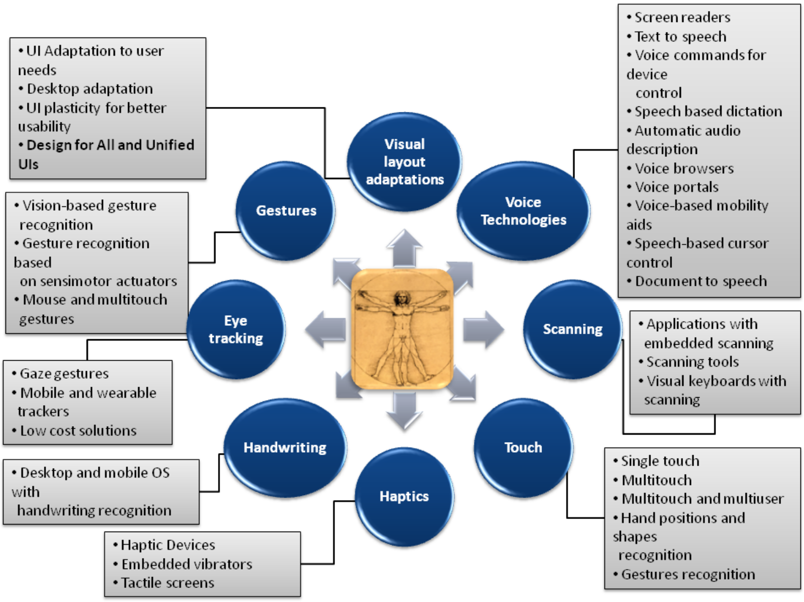


Fig. 3. Assistive technology categories

Assistive solutions and accessibility technologies have so far supported the augmentation of the capabilities of the individuals and the adaptation of single artifacts for accessibility. In the context of AmI, it is necessary to build on current state of the art in the field and further research new interaction methods as they emerge, taking advantage of built-in features in these living spaces, such as multimodality. A number of personalized cross – domain AmI assistive solutions that will address the accessibility and usability needs of older and disabled users can be designed and developed, building upon state-of-the-art solutions in user interface visual adaptation, voice-based interaction, scanning-based interaction, touch interaction, haptic feedback, handwriting, and eye tracking technologies (see **Fig. 3**). Different modalities can be provided based on user's preferences and needs according to the ontology model.

Personalized assistive solutions for AmI environments are necessary to provide alternative input, output and information rendering modalities capable of addressing each individual's needs (see **Fig. 4**). An innovative aspect relates to the interaction and co-operation between different / changing applications and devices used in each environment. Furthermore, such an approach will provide the building blocks for the harmonized interoperability of assistive technologies and conventional computational devices (PCs, smart phones, etc.) with innovative high end technology artifacts in the context of AmI accessible environments. The developed assistive solutions will be available through an open architecture that will allow their interoperation over different application domains, enabling use through a combination of personal and ambient devices.

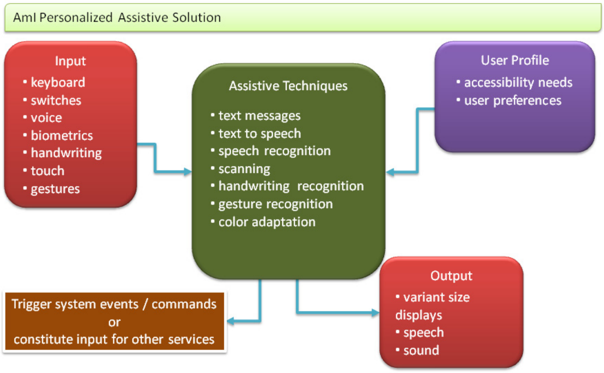


Fig. 4. AmI Personalized Assistive Solution concept

5.3 Accessible AmI Applications

The provision of accessibility solutions addressing the interaction needs of older users and people with activity limitations will foster the adoption of effective approaches for the design of AmI accessible environments' applications. Innovative multi-modal application concepts, using personalized AmI assistive solutions, will be investigated in three domains encompassing fundamental activities of daily life, namely home, work and self-care. Through the requirements analysis and the assistive solutions and infrastructure that will be developed, new AmI accessible interactive applications will be proposed addressing everyday needs of older people and people with activity limitations.

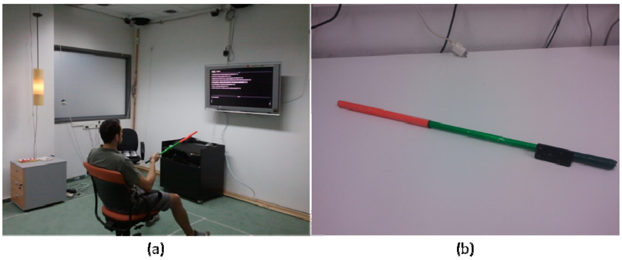


Fig. 5. AmI home navigation and control for blind people

The adopted approach will move beyond AmI – home / office / self-care technologies to integrating assistive solutions into things of everyday objects.

One of the applications under development is the implementation of the "magic wand" [16] and "universal control" described in the scenario of section 4. As depicted in **Fig. 5** (a), a laboratory AmI living room has already been set-up able to monitor and give auditory feedback about the location and state of the room's objects to blind inhabitants pointing with the plastic stick of **Fig. 5** (b). The remote mounted on the stick enables to control the room's devices and applications in a seamless way.

5.4 Deployment and Pilot Evaluation

The evaluation of AmI technologies and environments will need to go beyond traditional usability evaluation in a number of dimensions, concerning assessment methods and tools as well as metrics. AmI technologies and systems challenge traditional usability evaluation methods, because the context of use can be difficult to recreate in a laboratory setting. This suggests that the evaluation of user's experience with AmI technologies should take place in real world contexts. However, evaluation in real settings also presents difficulties, as there are limited possibilities of continuously monitoring users and their activities. In this respect, the AmI Facility will offer an ideal experimental environment, combining user experience in context with the availability of the necessary technical infrastructure for studying the users' behavior over extended periods of time. The current deployment and evaluation plans include accessible AmI applications in the home, work, learning and health-care domains.

6 Conclusions

This paper has discussed a landscape and roadmap to enable accessible user interaction in AmI environments for people with disabilities and older inhabitants in a systematic way. To achieve these objectives a number of implementation steps are proposed including: the development of a reference architectural model, which will address user needs for inclusive design in AmI environments and the implementation of AmI personalized assistive solutions, which will enable accessible user interaction in AmI environments. Moreover, design tools will be implemented in order to enable developers of AmI environments to effectively design accessible everyday life applications, while a number of test AmI accessible applications will be designed and developed aiming to enable the users to interact in the AmI environment using various conventional devices and AmI artifacts available in the environment, according to their needs and requirements. The adopted approach goes in the direction of catering for accessibility at design time in a systematic fashion, as due to their complexity AmI environments cannot be made accessible 'a posteriori' after development.

Finally, evaluations of the developed assistive solutions tools and applications will be carried out in interconnected simulation environments, in order to assess their accessibility and added value for the target users.

References

1. Overview of the European strategy in ICT for Ageing Well, `http://ec.europa.eu/information_society/activities/einclusion/docs/ageing/overview.pdf`
2. Savidis, A., Stephanidis, C.: Unified User Interface Design: Designing Universally Accessible Interactions. Int. J. Interact. Comput. 16(2), 243–270 (2004)
3. Carbonell, N.: Ambient multimodality: towards advancing computer accessibility and assisted living. Univ. Access. Inf. Soc. 5, 96–104 (2006)

4. Richter, K., Hellenschmidt, M.: Position paper: Interacting with the ambience: Multimodal interaction and ambient intelligence. In: W3C Workshop on Multi-modal Interaction, Sophia Antipolis, France, July 19/20 (2004)
5. Kemppainen, E., Abascal, J., Allen, B., Delaitre, S., Giovannini, C., Soede, M.: Ethical and legislative issues with regard to Ambient Intelligence. In: Roe, P.R.W. (ed.) Impact and Wider Potential of Information and Communication Technologies, COST, Brussels, pp. 188–205 (2007)
6. Han, J.Y.: Multi-touch interaction wall. In: ACM SIGGRAPH 2006 Emerging technologies (SIGGRAPH 2006). ACM, New York (2006)
7. Coroama, V., Bohn, J., Mattern, F.: Living in a Smart Environment – Implications for the Coming Ubiquitous Information Society. In: Proceedings of the International Conference on Systems, Man and Cybernetics 2004 (IEEE SMC 2004), vol. 6, pp. 5633–5638 (2004)
8. Edwards, W.K., Grinter, R.E.: At Home with Ubiquitous Computing: Seven Challenges. In: Abowd, G.D., Brumitt, B., Shafer, S. (eds.) UbiComp 2001. LNCS, vol. 2201, pp. 256–272. Springer, Heidelberg (2001)
9. Antona, M., Burzagli, L., Emiliani, P.-L., Stephanidis, C.: The ISTAG scenarios: a case study. In: Roe, P.R.W. (ed.) Towards an Inclusive Future: Impact and Wider Potential of Information and Communication Technologies, Ambient Intelligence and implications for people with disabilities, ch.4, sec. 4.1, pp. 158–187. COST219ter, Brussels (2007)
10. Emiliani, P.-L., Burzagli, L., Billi, M., Gabbanini, F., Palchetti, E.: Report on the impact of technological developments on eAccessibility. DfA@eInclusion Deliverable D2.1. (2009), `http://www.dfaei.org/deliverables/D2.1.pdf`
11. Emiliani, P.-L., Aalykke, S., Antona, M., Burzagli, L., Gabbanini, F., Klironomos, I.: Document on necessary research activities related to DfA. DfA@eInclusion Deliverable D2.6 (2009), `http://www.dfaei.org/deliverables/D2.6.pdf`
12. Nehmer, J., Becker, M., Karshmer, A., Lamm, R.: Living assistance systems: an ambient intelligence approach. In: Proceedings of the 28th International Conference on Software Engineering (ICSE 2006), pp. 43–50. ACM, New York (2006)
13. Kleinberger, T., Becker, M., Ras, E., Holzinger, A., Müller, P.: Ambient Intelligence in Assisted Living: Enable Elderly People to Handle Future Interfaces. In: Stephanidis, C. (ed.) UAHCI 2007 (Part II). LNCS, vol. 4555, pp. 103–112. Springer, Heidelberg (2007)
14. Marinc, A., Stocklöw, C., Braun, A., Limberger, C., Hofmann, C., Kuijper, A.: Interactive Personalization of Ambient Assisted Living Environments. In: Smith, M.J., Salvendy, G. (eds.) HCII 2011, Part I. LNCS, vol. 6771, pp. 567–576. Springer, Heidelberg (2011)
15. Preuveneers, D., Van den Bergh, J., Wagelaar, D., Georges, A., Rigole, P., Clerckx, T., Berbers, Y., Coninx, K., Jonckers, V., De Bosschere, K.: Towards an Extensible Context Ontology for Ambient Intelligence. In: Markopoulos, P., Eggen, B., Aarts, E., Crowley, J.L. (eds.) EUSAI 2004. LNCS, vol. 3295, pp. 148–159. Springer, Heidelberg (2004)
16. Zabulis, X., Koutlemanis, P., Baltzakis, H., Grammenos, D.: Multiview 3D Pose Estimation of a Wand for Human-Computer Interaction. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Wang, S., Kyungnam, K., Benes, B., Moreland, K., Borst, C., DiVerdi, S., Yi-Jen, C., Ming, J. (eds.) ISVC 2011, Part II. LNCS, vol. 6939, pp. 104–115. Springer, Heidelberg (2011)

INCOME – Multi-scale Context Management for the Internet of Things

Jean-Paul Arcangeli[1], Amel Bouzeghoub[2], Valérie Camps[1], Marie-Françoise Canut[1], Sophie Chabridon[2], Denis Conan[2], Thierry Desprats[1], Romain Laborde[1], Emmanuel Lavinal[1], Sébastien Leriche[2], Hervé Maurel[3], André Péninou[1], Chantal Taconet[2], and Pascale Zaraté[1]

[1] Université de Toulouse, IRIT UMR 5505, France
[2] Télécom SudParis, UMR CNRS 5157 SAMOVAR, France
[3] ARTAL Technologies, Toulouse, France
Jean-Paul.Arcangeli@irit.fr

Abstract. Nowadays, context management solutions in ambient networks are well-known. However, with the IoT paradigm, ambient information is not anymore the only source of context. Context management solutions able to address multiple network scales ranging from ambient networks to the Internet of Things (IoT) are required. We present the INCOME project whose goal is to provide generic software and middleware components to ease the design and development of mass market context-aware applications built above the Internet of Things. By revisiting ambient intelligence (AmI) context management solutions for extending them to the IoT, INCOME allows to bridge the gap between these two very active research domains. In this landscape paper, we identify how INCOME plans to advance the state of the art and we briefly describe its scientific program which consists of three main tasks: (*i*) multi-scale context management, (*ii*) management of extrafunctional concerns (quality of context and privacy), and (*iii*) autonomous deployment of context management entities.

1 Motivations

The INCOME project, supported by the French National Research Agency from February 2012 to October 2015[1], intends to provide software framework solutions for multi-scale context management in the Internet of Things. INCOME innovates by bridging the gap between the AmI and the IoT research domains. Applications consume high-level context data, obtained after processing, fusing and filtering a large number of low-level context information. Nowadays, context management solutions in ambient networks are well-known [12,17,23,29,32]. However, the IoT paradigm opens the way for a continuous increase of the number of connectable items requiring new solutions able to cope with this change of scale. INCOME proposes to design context management solutions for addressing

[1] http://anr-income.fr

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 338–347, 2012.

not only ambient networks but also the IoT and clouds in a multi-scale framework able to operate at different scales and to deal with the passage from a scale to another one.

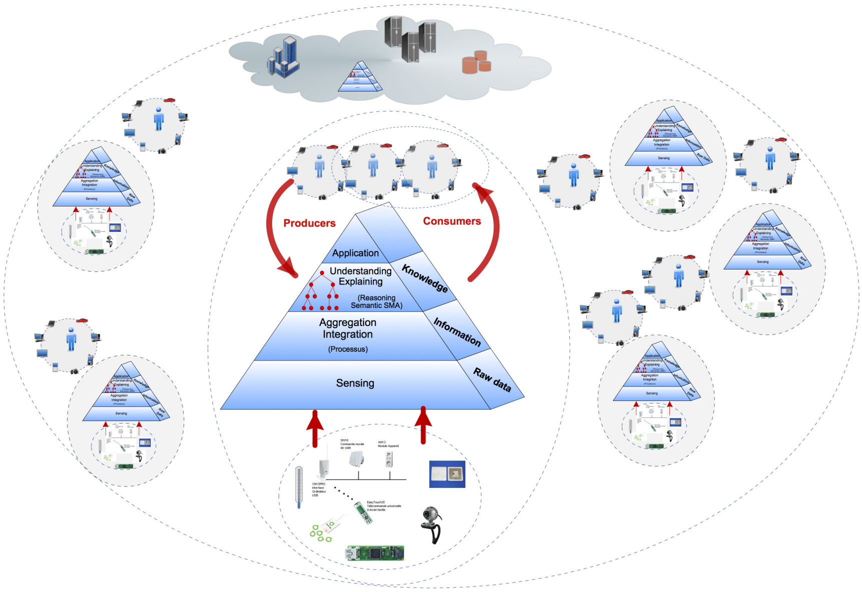


Fig. 1. Overview of the INCOME project

On Figure 1, we give an overview of our vision of multi-scale context management. In the INCOME project, the multi-scale qualifier applies to several context management aspects: the production, the processing and the consumption of context information. Concerning the production, context information originates not only from the ambient environment, but also from the IoT and from clouds (eg, context inferred from knowledge bases). The processing of context information can be distributed on resource-constrained mobile devices, on servers of the ambient network or on servers deployed in a cloud. Regarding the consumption, context information can be consumed both by mobile applications deployed on personal devices and by applications deployed on fixed nodes in a cloud.

The multi-scale factor has several consequences in terms of context management. Therefore, INCOME ambitions to answer to three scientific challenges: (i) multi-scale context management, (ii) management of the two extrafunctional concerns of quality of context and privacy, becoming crucial in the IoT, and (iii) autonomous deployment of context management entities. The proposed solutions will be validated in at least two application domains such as context-aware applications for mobile users in a multi-location university campus and in a large skiing domain.

This paper is composed of two main sections. In Section 2, we present how INCOME will advance the state of the art on context management. In Section 3, we state the objectives and the workplan of INCOME and conclude the paper in Section 4.

2 Advancing the State of the Art

In this section, we discuss a survey of the scientific domains addressed by the INCOME project, namely context management, quality of context, privacy protection and deployment technologies. In each domain, we indicate in which research directions we envision to advance the current state of the art.

2.1 Context Management

As INCOME is concerned with multi-scale context management, we focus in this section on two subjects: the distribution of the data flows between the different scales and the heterogeneity of the context management approaches.

Distributed Context Management. The distribution of context information can take place according to various paradigms. Historically, the first works come from the domain of system supervision with the monitoring of distributed systems or computation grids. The architectural style of these systems is based on message passing with asynchronous communication [8].

The second paradigm corresponds to the software bus, initially relying on remote procedure calls. Examples in this category are CoBrA [10], built using intelligent agents following the FIPA model of the OMG, and the context service of Gaia [32] which is a CORBA event service distributing events occurring in "active spaces". The software bus paradigm is not sufficiently scalable and lacks flexibility to support the inherent dynamicity of multi-scale systems.

The last architectural style encountered in context management frameworks follows the Publish/Subscribe paradigm, mainly in peer-to-peer systems. In [39], each piece of context information is managed by an overlay network, but with no possibility to compose them. In the IST MADAM and MUSIC projects [27,35], sensor peers transmit raw data obtained from physical sensors; disseminator peers have more processing resources and play the role of context processors, distributors or consumers; consumer peers are only context users with scarce resource. Peer-to-peer systems manage efficiently the volatility of peers and are relevant solutions for large scale systems. However, they are not designed to address the different degrees of multi-scale management.

In INCOME, we favor event-based solutions [2] and the Complex Event Processing paradigm [18] which appear to be easily scalable, very flexible and powerful enough to express constraints on the quality of context information and on the protection of user's privacy.

Detection of Situations of Interest. The processing of context information has for objective the identification of situations of interest. With the high number of sources of context information, the dynamicity of the user's behaviour and the increasing amount of context information to be analysed to identify one relevant situation, the self-adaptation of the detection system appears to be a promising solution. Self-adaptation allows a system to adjust its own behaviour as a response to its perception of its environment and of itself, and to modify its internal organization without any external control [11,41]. Only very few multi-agent architectures [45] with self-adaptation capabilities have been proposed to tackle complex open systems. Among them, the AMAS architecture [7,21] allows in a first step to modify the function of the agents, and if it is not sufficient, to change the interaction links among agents, and even the composition of the system by adding or removing agents. The INCOME project will address this openness aspect related to the volatility of agents as one research direction.

Besides, situation detection requires also the interpretation of context information at various levels of abstraction: from physical sensors to the application, while taking into account the semantics of context information. Various approaches have been suggested to implement context management. A toolkit approach provides tools and guidelines for the development and the execution of context-aware applications and libraries of reusable components. Examples are Context Toolkit [17], Contextor [13] and COSMOS [12]. In these works, process units are organised in graphs or trees, where the leaves collect raw data from context sources and the other nodes compute context information of higher levels of abstraction. Ontology-based approaches like CoBrA [10], SOCAM (conon ontology) [23], GAIA [32] and MUSE [4] allow to automatically deduce by inference high-level implicit information (eg, the activity of a user) from low-level context data (such as location, temperature or noise level). They also promote interoperability as they provide the definition of common concepts of context information. Finally, they define a generic representational model of context information that any system can use.

The innovation of INCOME is to study how imperative approaches, ontological (or deductive) approaches and adaptive multi-agent systems can jointly contribute to make situation detection more dynamic and better fitted to open and multi-scale environments.

2.2 Quality of Context

The concept of Quality of context (QoC) as a first-class concept for context-aware services has first been identified by [5] defining it as "any information describing the quality of information that is used as context", and considering that it is independent of the computing process (eg, quality of service) or of the hardware equipment (eg, quality of device). The notion of worth has then been added to introduce the point of view of the targeted applications [3,30]. QoC can be represented by meta-data, such as up-to-dateness, accuracy, completeness, associated to context information. A first set of these parameters is directly collected from context sources, depending on the information available at the sources, and additional parameters can be computed at the acquisition step or even later during the inference process by the context management service [1].

Context data are indeed known to be inherently uncertain due to the imperfection of physical sensors and the real world itself [6,25]. As context data are by nature dynamic and very heterogeneous, they also tend to be *incorrect* and not exactly reflecting the real state of the modeled entity. They can be *inconsistent* as there is a risk of having contradictory information from different context sources. They finally can be *incomplete* when some aspects of the context are missing [26]. Therefore, taking into account the knowledge of the quality of context information appears to be essential to reach an effective and efficient context management and furthermore context-awareness.

Multi-level or hierarchical approaches for QoC management have been proposed [31,37] to provide an aggregated view essential to manipulate a high amount of context data. Some efforts to deal with adaptive QoC management, as in [36], are also pursued to dynamically tackle both the level(s) of QoC expected by the applications and the limitations constraints that are imposed by the underlying execution resources. However, they are not sufficient for multi-scale context management and must be complemented with mechanisms allowing to reason on a subpart of the context information space, as we propose in INCOME.

2.3 Trust and Privacy

Trust can be defined as "*the extent to which one party is willing to depend on something or somebody in a given situation with a feeling of relative security, even though negative consequences are possible*" [28,34]. This implies that an entity (human or not) can trust a third party only if it can have access to the information required to take a decision. The most widely used trust indicator is the source of the information. The reasoning is then very simple: either the information source is trustworthy and the information is accepted, or it is denied. Reputation [22,28] is another trust indicator that represents a social indicator based on the number of recommendations, their date, their volatility, etc. Some approaches propose to quantify the process that produced the information. For instance, [38] relies on the level of the authentication (LoA) technology to assess the confidence one can have on users' identity. However, this is not a completely reliable indicator as the practical usages of the technology are as important as the technology itself. For instance, [44] quantifies the quality of X.509 electronical certificates with respect to the procedures for the management of the certificates. INCOME studies the integration of trust indicators as part of the quality of context.

Concerning access control strategies, some works are investigating which ones are best fitted to context-aware environments. Role-based access control (RBAC) models appear to be promising in that they allow to define environment roles that can be activated or not [14]. [15] proposed the Contextual Attribute Based Access Control (CABAC) model with the definition of context-aware access policies. However, only a few works [19,43] take into account the quality of context in the access control decisions.

Also, some recent works are starting to add the notion of QoC in the implementation of privacy policies [42], but the visible contradiction that exists between the quality of the information expected by a context consumer and the protection constraint for privacy is still an open issue [9] and appeals for new solutions that are investigated in INCOME.

2.4 Autonomic Deployment

Deploying in an open environment at a large scale implies to handle a large number of nodes, with heterogeneous network links and a multitude of software versions. The management of all these aspects requires well adapted tools that allow to control and automate the deployment process. Our vision of autonomic deployment is characterized by the absence of human intervention and the capability to solve automatically problems caused by the instability and the openness of the environment while respecting a set of quality of service and security constraints. Existing solutions answer only partly to this vision. SoftwareDock [24] rely on mobile agents for the management of a completely decentralised deployment process. However, it considers a closed environment with a fixed architecture. Agents enable the movement of the components from one node to another, but they do not really take part to the deployment activity. FDF/Deployware (Fractal Deployment Framework) [20] is a generic deployment framework but it does not address instable and open systems. Domain Specific Languages have been proposed for expressing deployment constraints in open environments. The Deladas (Declarative Language for Describing Autonomic Systems) [16] language is associated to a constraint solver to generate a concrete configuration from an initial description and to activate the deployment. In case of a node failure or of a conflict during the deployment phase, the deployment manager restarts the constraint resolution phase to obtain a new deployment plan. The relevance of such an approach for multi-scale systems is investigated in INCOME. The most recent works on software deployment consider some quality of service criteria. [33] introduced a *framework* with a formalism and tools to specify the deployment plan that is the most appropriate for a distributed system relatively to several quality of service constraints that can be contradictory. The contribution of INCOME is to evaluate the approach based on autonomous agents for the supervision and the dynamic adaptation of the deployment process in order to tolerate disconnections and QoS variations at all scales.

3 Scientific Program of INCOME

The three functionalities of a context manager, namely the collection, the processing and the consumption of context information are all impacted by the multi-scale dimension as well as by the extrafunctional concerns that are the quality of context and the protection of the information (eg, privacy protection). Therefore, INCOME addresses challenges on three different aspects, each aspect being the subject of a different workpackage (WP).

WP 1 - Functional aspect. Context managers have to distribute both the processing and the data flows. Secondly, they must scale in terms of volume of data, number of sources and number of consumers. They also have to face the heterogeneity of context data. Moreover, a muti-scale context manager is intrinsically an ubiquitous system: users are mobile, the things considered in the IoT are embedded in the physical environment, and can also be mobile. Finally, a context manager must have self-adaptation capabilities to be able to take into account new objects / participants / observation contracts, new network topology, etc.

In such a complex environment, the detection of situations of interest to consumer applications involves several levels of processing, requiring a hybrid context management. The hybrid approach that we promote in the INCOME project combines imperative (treatments described by imperative expressions), deductive ontology-based (treatments described by logical rules), and multi-agent based context managers. Each of these context management approaches has its own advantages: efficient reactive time (imperative), processing of unforeseen situations (deductive) or intrinsic dynamic adaptation (multi-agent), enabling to provide the appropriate solution at the appropriate level.

On this functional aspect, INCOME aims to define an architectural modeling of a generic framework able to support *(i)* the distribution of context management processing, *(ii)* the heterogeneity, multiplicity, dispersion and instability of context sources and the management of information flows, *(iii)* the integration of multiple types of managers (imperative, deductive and multi-agent based).

WP 2 - Extrafunctional aspect. An important challenge addressed by the INCOME project is to consider two extrafunctional aspects which are fundamental for multi-scale context management, namely the quality of the context information (QoC) and the protection of privacy. At the scale of the IoT, where context information providers are numerous and unknown, it is highly required to associate the quality of the context information together with the context information itself. This allows context managers to take into account the correctness or the uncertainty of the information manipulated by context-aware applications. Moreover, privacy protection is an essential element for guaranteeing ethical properties to the next generation of context-aware applications. Strategies for the protection of personal context information must be embodied into context management in order to be able to protect sensible data.

Therefore, INCOME will define models and tools allowing *(i)* to reason on context data which are potentially uncertain or incomplete, while integrating (resp. removing) dynamically new data that became available (resp. unavailable) according to the scale considered and to the discovery of new context sources, *(ii)* to adapt on the fly the acquisition modalities of context information considering various constraints such as the variation of the expectations of consumers or the optimisation of execution and interaction time, *(iii)* to control adaptively context information dissemination in order to protect the user's privacy.

WP 3 - Operational aspect. INCOME targets the infrastructure level for mass market context-aware applications. Those applications have to be deployed at a wide scale both in terms of the number of deployment locations and the number of users. For this kind of applications, available context data vary according to both geographical and temporal dimensions. In these conditions, autonomous deployment strategies for context management entities are essential. These strategies allow the infrastructure to automatically support the instability and openness of the environment.

In this last workpackage, INCOME will propose a dedicated middleware solution for *(i)* the autonomous deployment of context management software components distributed on heterogeneous physical devices in the multi-scale environment, *(ii)* the adaptive redeployment of these components in reaction to constraint variations (topology changes, variations of the expected quality of context, security policies, ...), and *(iii)* their execution.

4 Conclusion

With the IoT paradigm, context-aware applications not only have to deal with context data collected from ambient networks but also with remote context sources located at multiple scales. The originality of the INCOME project is to design a multi-scale framework able to operate at different scales and to deal with the passage from a scale to another one. As AmI represents the first scale level, INCOME revisits existing context management solutions for AmI and will extend them to the IoT. As identified by the *Privacy, Trust and Interaction in the Internet of Things* workshop of the AmI-2011 conference [40] in relation with the uTRUSTit FP7 project (http://www.utrustit.eu), privacy is a central concern in the IoT. One of the main contributions of INCOME is to study privacy together with the quality of the context information as these two issues are intimately related. The results of INCOME will benefit to multi-scale context management on three aspects (i) distributed context management with the integration of imperative, deductive and multi-agent based managers (ii) dynamic adaptation of QoC and privacy requirements from the consumer and the producer points of view respectively (iii) autonomous deployment and reconfiguration of context management software components in the multi-scale environment.

Acknowledgments. This work is part of the French National Research Agency (ANR) project INCOME[2] (ANR-11-INFR-009, 2012-2015). The authors thank all the members of the project that contributed directly or indirectly to this paper.

References

1. Abid, Z., Chabridon, S., Conan, D.: A Framework for Quality of Context Management. In: Rothermel, K., Fritsch, D., Blochinger, W., Dürr, F. (eds.) QuaCon 2009. LNCS, vol. 5786, pp. 120–131. Springer, Heidelberg (2009)
2. Baldoni, R., Querzoni, L., Tarkoma, S., Virgillito, A.: A Framework for Quality of Context Management. In: Middleware for Network Eccentric and Mobile Applications, pp. 219–244. Springer (2009)

[2] http://anr-income.fr

3. Bisdikian, C.: On sensor sampling and quality of information: A starting point. In: 5th IQ2S PerCom Workshop, pp. 279–284 (March 2007)
4. Bouzeghoub, A., Do Ngoc, K.: A situation based metadata for describing pervasive learning objects. In: Proc. mLearn 2008, Ironbridge, UK (October 2008)
5. Buchholz, T., Kupper, A., Schiffers, M.: Quality of context information: What it is and why we need it. In: 10th HPOVUA Workshop, Switzerland (July 2003)
6. Bettini, C., et al.: A survey of context modelling and reasoning techniques. Pervasive and Mobile Computing (2009)
7. Capera, D., George, J.-P., Gleizes, M.-P., Glize, P.: The amas theory for complex problem solving based on self-organizing cooperative agents. In: 12th Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (2003)
8. Cecchet, E., Elmeleegy, H., Layada, O., Quma, V.: Implementing Probes for J2EE Cluster Monitoring. Studia Informatica (2005)
9. Chakraborty, S., Choi, H., Srivastava, M.B.: Demystifying Privacy In Sensory Data: A QoI based approach. In: 3d IQ2S PerCom Workshop (March 2011)
10. Chen, H.: An Intelligent Broker Architecture for Pervasive Context-Aware Systems. PhD thesis, University of Maryland, Baltimore, USA (January 2003)
11. Cheng, B.H.C., de Lemos, R., Giese, H., Inverardi, P., Magee, J. (eds.): Software Engineering for Self-Adaptive Systems. LNCS, vol. 5525. Springer, Heidelberg (2009)
12. Conan, D., Rouvoy, R., Seinturier, L.: Scalable Processing of Context Information with COSMOS. In: Indulska, J., Raymond, K. (eds.) DAIS 2007. LNCS, vol. 4531, pp. 210–224. Springer, Heidelberg (2007)
13. Coutaz, J., Rey, G.: Foundations for a Theory of Contextors. In: 4th Int. Conf. on Computer-Aided Design of User Interfaces, France. Kluwer (May 2002)
14. Covington, M.J., Fogla, P., Zhan, Z., Ahamad, M.: A context-aware security architecture for emerging applications. In: IEEE Conf. on Computer Security Applications, Los Alamitos, CA, USA (2002)
15. Covington, M.J., Sastry, M.R.: A Contextual Attribute-Based Access Control Model. In: Meersman, R., Tari, Z., Herrero, P. (eds.) OTM 2006 Workshops. LNCS, vol. 4278, pp. 1996–2006. Springer, Heidelberg (2006)
16. Dearle, A., et al.: A framework for constraint-based deployment and autonomic management of distributed applications. In: ICAC. IEEE Computer Society (2004)
17. Dey, A.K., Abowd, G.D., Salber, D.: A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. Special Issue on Context-Aware Computing, HCI Journal 16(2-4) (2001)
18. Etzion, O., Niblett, P.: Event Processing in Action. Manning Pub. Co. (2010)
19. Filho, J.B., Martin, H.: A generalized context-based access control model for pervasive environments. In: 2nd SIGSPATIAL ACM Workshop on Security and Privacy in GIS and LBS, SPRINGL 2009. ACM, New York (2009)
20. Flissi, A., Dubus, J., Dolet, N., Merle, P.: Deploying on the grid with deployware. In: CCGRID, pp. 177–184 (2008)
21. Gleizes, M.-P., Camps, V., Georgé, J.-P., Capera, D.: Engineering Systems Which Generate Emergent Functionalities. In: Weyns, D., Brueckner, S.A., Demazeau, Y. (eds.) EEMMAS 2007. LNCS (LNAI), vol. 5049, pp. 58–75. Springer, Heidelberg (2008)
22. Gray, E.L.: A trust-based reputation management system. PhD thesis, University of Dublin (2006)
23. Gu, T., Wang, X.H., Pung, H.K., Zhang, D.Q.: An Ontology-based Context Model in Intelligent Environments. In: CNDS 2004, San Diego, CA, USA (2004)
24. Hall, R., Heimbigner, D., Wolf, A.: A cooperative approach to support software deployment using the software dock. In: 21st ACM ICSE 1999 (1999)

25. Henricksen, K., Indulska, J.: Modelling and using imperfect context information. In: 1st CoMoRea PerCom 2004 Workshop. IEEE Computer Society (March 2004)
26. Henricksen, K., Indulska, J., Rakotonirainy, A.: Modeling Context Information in Pervasive Computing Systems. In: Mattern, F., Naghshineh, M. (eds.) PERVASIVE 2002. LNCS, vol. 2414, pp. 167–180. Springer, Heidelberg (2002)
27. Hu, X., Ding, Y., Paspallis, N., Papadopoulos, G.A., Bratskas, P., Barone, P., Mamelli, A., Vanrompay, Y., Berbers, Y.: A Peer-to-Peer Based Infrastructure for Context Distribution in Mobile and Ubiquitous Environments. In: Meersman, R., Tari, Z. (eds.) OTM-WS 2007, Part I. LNCS, vol. 4805, pp. 236–239. Springer, Heidelberg (2007)
28. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decis. Support Syst. 43(2) (March 2005)
29. Julien, C., Roman, G.-C.: EgoSpaces: Facilitating Rapid Development of Context-Aware Mobile Applications. IEEE Trans. on S.E. 32(5) (May 2006)
30. Krause, M., Hochstatter, I.: Challenges in Modelling and Using Quality of Context (QoC). In: Magedanz, T., Karmouch, A., Pierre, S., Venieris, I.S. (eds.) MATA 2005. LNCS, vol. 3744, pp. 324–333. Springer, Heidelberg (2005)
31. Lange, R., et al.: Making the World Wide Space Happen: New Challenges for the Nexus Context Platform. In: 7th IEEE PerCom, Galveston, TX, USA (March 2009)
32. Román, M., et al.: Gaia: A Middleware Infrastructure to Enable Active Spaces. IEEE Pervasive Computing 1(4), 74–83 (2002)
33. Malek, S., Medvidovic, N., Mikic-Rakic, M.: An extensible framework for improving a distributed software system's deployment architecture. IEEE Trans. on Software Engineering (99) (2012)
34. McKnight, H., Carter, M., Clay, P.: Trust in technology: development of a set of constructs and measures. In: DIGITS (2009)
35. Mikalsen, M., et al.: Distributed context management in mobility and adaptation enabling middleware. In: Proc. ACM SAC 2006, France (2006)
36. Moui, A., Desprats, T., Lavinal, E., Sibilla, M.: Management information model for monitoring adaptation enforcement. In: Proc. ADAPTIVE Conf. (July 2012)
37. Mühlhäuser, M., Hartmann, M.: Interacting with Context. In: Rothermel, K., Fritsch, D., Blochinger, W., Dürr, F. (eds.) QuaCon 2009. LNCS, vol. 5786, pp. 1–14. Springer, Heidelberg (2009)
38. Nenadic, A., Zhang, N., Yao, L., Morrow, T.: Levels of authentication assurance: an investigation. In: 3d IAS Symposium (2007)
39. Nguyen, T.D., Rouvrais, S.: Towards a Peer-to-peer Middleware for Context Provisioning in Spontaneous Networks. In: 5th MiNEMA Workshop, Magdeburg, Germany (September 2007)
40. Wichert, R., Van Laerhoven, K., Gelissen, J. (eds.): AmI 2011. CCIS, vol. 277. Springer, Heidelberg (2012)
41. Di Marzo Serugendo, G., Gleizes, M.-P., Karageorgos, A.: Self-organization in multi-agent systems. Knowledge Eng. Review 20(2), 165–189 (2005)
42. Sheikh, K., Wegdam, M., van Sinderen, M.J.: Quality-of-context and its use for protecting privacy in context aware systems. J. of Software 3(3) (March 2008)
43. Toninelli, A., Corradi, A., Montanari, R.: A Quality of Context-Aware Approach to Access Control in Pervasive Environments. In: Bonnin, J.-M., Giannelli, C., Magedanz, T. (eds.) Mobilware 2009. LNICST, vol. 7, pp. 236–251. Springer, Heidelberg (2009)
44. Wazan, S.A., Laborde, R., Barrere, F., Benzekri, A.: A Formal Model of Trust for Calculating the Quality of X.509 Certificates. In: Security and Communication Networks (2010)
45. Wooldridge, M.: Introduction to MultiAgent Systems. John Wiley and S. (2002)

New Forms of Work Assistance by Ambient Intelligence

Overview of the Focal Research Topic of BAuA

Armin Windel and Matthias Hartwig

Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Department Products and Work Systems, Scientific Management, Dortmund, Germany
{windel.armin,hartwig.matthias}@baua.bund.de

Abstract. Research and development in the area of AmI are currently migrating from basic technologies to application-readiness. In the consumer domain a field of research has now become established bearing the description Ambient Assisted Living. In contrast, for the area of work, practice-oriented and work science-relevant questions in the field of AmI are only considered in isolated cases and unsystematically. To capture the intersection of the world of work and the concept of AmI, the term adaptive work assistance systems (AWAS) is coined in this paper. Furthermore, the different research project bundles at BAuA are described briefly, each focusing on opportunities and risks of new I&C technologies in the working environment and their impact on work systems. Most of the projects are handled as BAuA's own research and supported by external research contracted extramurally.

Keywords: Ambient Intelligence, application in working life, adaptive work assistance systems, AWAS, work science, research programme.

1 AmI as a New Focal Research Topic of BAuA

In its current research and development programme (starting 2010) BAuA has announced as one of its priority research topics Ambient Intelligence in the working environment. It is thus tackling the need for application-oriented research highlighted in the future report "Work in the Future – Structures and Trends of Industrial Work" of the Office of Technology Assessment at the German Bundestag (TAB) with respect to the two key technologies of "bio- and nanotechnology" and "ambient intelligence". This was developed further with a view to occupational safety and health and occupational medicine. The focus of the research at BAuA will be the technology assessment with respect to opportunities and risks of new I&C technologies in the working environment and their impact on work systems [1]. This contribution aims at giving an overview of BAuA activities (internal projects and funding) regarding AmI.

Research funding to date in relation to the subject of ambient intelligence has concentrated, especially in the sixth framework programme of the EU (2002 to 2006), primarily on developing basic technologies (e.g. hardware: sensors and actuators). In

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 348–355, 2012.

the recent past questions of software and the impact of invisible, ubiquitous technologies have played an increasingly important role (see the EU's seventh framework programme [2]). In the consumer domain a field of research has now become established bearing the description AAL (Ambient Assisted Living). Central to this is the development of I&C technologies and services for the everyday living of elderly people. In contrast, for the area of work it can be said that practice-oriented, work science-relevant questions in the field of AmI are only considered at present in isolated cases and unsystematically.

One major reason for this is certainly that there is still a manageable number of realistic application scenarios militating against establishment of the term ambient intelligence in policy and research funding. But BAuA sees here a special opportunity to grasp developments in the field of ambient intelligence as work assistance systems. It also sees the chance to be able to influence the impact on man-machine interactions, working sequences and processes and the efficiency of the work system as a whole, not correctively but prior to any widespread introduction into the world of work.

2 Definition: AWAS - Application-Oriented AmI in Working Life

Ambient Intelligence extends the living and working environment by networking sensors, actuators and computer processors so that well-being, health and efficiency in one's working and personal life are supported. Therefore, concrete information and communication technologies (e.g. wearable IT - smart clothes) or work science-relevant fields (e.g. augmented reality) have to be seen as relevant subject areas for Ambient Intelligence. For the BAuA's practice-oriented research there are features of AmI which can be highlighted:

- The information provided by telemetry which is intended in particular to give support in working life and
- The provision of information which depends on the state of the user and/or the working environment (including the specific task).

Both aspects of the definition stress the assistance function of Ambient Intelligence. The BAuA thus emphasizes the anthropocentric approach of technological development [3] where products and environments change adaptively and (largely) autonomously to fit the user's conditions, needs and goals. Based on these considerations, the development of a definition was primarily guided by the following objectives:

- Explicit coverage of the intersection of work systems and AmI
- General comprehensibility
- Conceptual wideness for dialogue with researchers, developers and users alike
- Definitional clarity in terms of a set of criteria
- Consistency with ongoing and planned projects within the BAuA research area AmI
- Clarity and accessibility in terms of a one-sentence definition

To capture the intersection of the world of work and the concept of AmI the term **adaptive work assistance systems (AWAS)** was chosen. In addition, a structured set of criteria has been developed in order to fulfil the partly conflicting demands of handling, completeness, clarity and comprehensibility. In the selection of categories and the formulation of the different facets scientific definitions were considered as well as the perspective of potential users. For the topic of adaptive work assistance systems, the BAuA introduces the following definition, specifying the application of AmI in the world of work. The first paragraph grasps the core concept of AWAS, while the four additional criteria explain different facets more in detail.

Core Concept. The term "adaptive work assistance systems (AWAS)" subsumes methods, concepts and (electronic) systems/products, which assist persons at work in a context-sensitive and autonomous way.

Key Features and Core Functions. The technologies are user-centred: The system uses sensors to detect one or more facets of the situation (especially features or conditions of the user, the task at hand or the environment) and adapts to it autonomously. Thereby, the system supports the user either by information or direct intervention.

System Architecture. In order to acquire and utilize situation-awareness, AWAS typically is based on decentralized networks.

User Groups. The usage is not limited to certain sectors or branches and can therefore include persons in the whole world of work.

Extent and Objectives of Support. AWAS can enhance safety, health, well-being and / or performance of the working people.

Based on this concept the tern AWAS can be understood as the application of AmI in the world of work like the term AAL refers to the application of AmI in private life. From this point of view, AWAS aims at connecting AmI to work sciences.

3 BAuA's Focal Topics Related to AmI, an Overview

Figure 1 gives an overview of the current project bundles or projects related to the focal topic of "ambient intelligence". It also illustrates that BAuA distinguishes between AmI-based products and the application of AmI systems in the working environment, at present primarily environmental control. Higher level questions concern the work-psychological and medical impact of AmI-based technologies, the safety of such systems and ethical matters relating to their application.

All project bundles subsumed in the focal topic of AmI pursue the common objective of ascertaining and evaluating opportunities and risks of AmI-based work assistance systems in terms of work science and occupational medicine. This should be taken to cover effects both on people in the work process (health and performance) and on the work system (safety, efficiency). In this respect all projects or project bundles make a contribution to the technology assessment of AmI-based systems.

Below the individual project bundles are characterised briefly. One part of the projects are handled as BAuA's own research, the other is conducted by external research contracted extramurally.

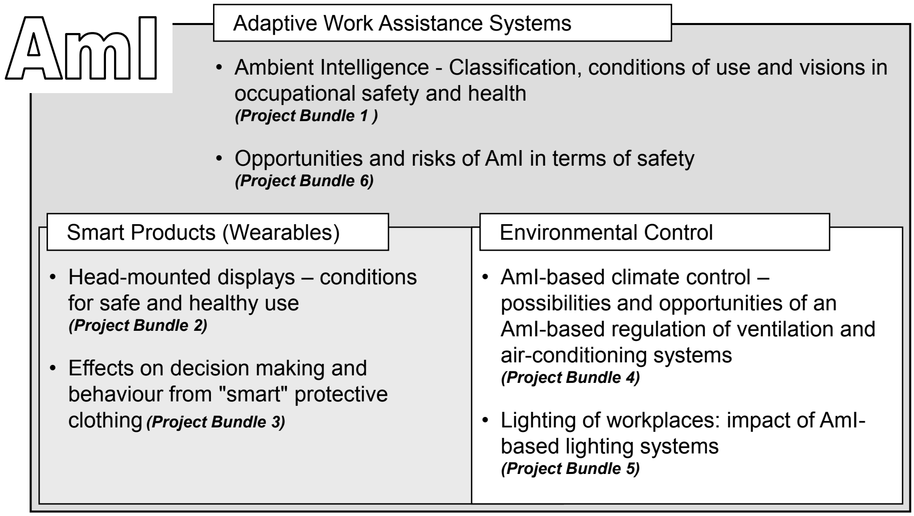


Fig. 1. Overview of the project bundles of BAuA relating to the focal topic of AmI

3.1 Project Bundle 1 "Classification and Conditions of Use as Work Assistance Systems"

AmI-based technologies leave a lot of scope for the design of the man-machine interface. This can range from simple support for the user through to an almost complete automation of functions or practical steps. The question arising here is what form and what degree of assistance is appropriate in the work system in order to enhance both the system's efficiency and user's health. A primary objective of the project bundle is a description of features and criteria of AmI-based systems as distinct from conventional technologies and from automation.

The second goal of the project bundle is to structure the field of AmI in the work of world and facilitate the networking of active institutions. To this end, the BAuA is currently preparing a dynamic online-map „Adaptive work assistance systems - Ambient Intelligence in the world of work“. Active parties can nominate their projects and institutions to appear on the map, while interested parties can search the database for compatible projects and possible cooperation-partners. The map is scheduled to be online at January 2013 and will be accessible via www.ami-map.de

Thirdly, AWAS offer new possibilities regarding the perspective of persuasive technology, defined by Fogg as “technology designed to intentionally change attitude and/or behaviour of users” [4]. By being able to adapt to the current situation and the individual user, work systems can potentially select the most effective strategy and timing to facilitate safety and health at work significantly [5]) . Simultaneously, the BAuA sees the necessity to explore and define ethical limits for this area to rule out unacceptable manipulation.

3.2 Project Bundle 2 "Opportunities and Risks of AmI in Terms of Safety"

Integrated, AmI-based safety solutions represent a new form of safety technology. In the production process they facilitate the close collaboration of man and machine without separating protective devices, such as protective fences, and they therefore create both a high degree of freedom of movement for the operator and a great flexibility in the process sequences. But it is not clear whether such AmI-based solutions actually improve the safety standard or only shift it to another level. It is conceivable that technical protective measures using AmI technology replace measures of inherently safe design, which would be a cause for concern in terms of safety – especially taking into account the foreseeable use (manipulation) of such measures.

The prime objective of the project bundle is therefore to assess the risk of AmI-based safety technology/safety measures and to compare the AmI solutions with inherently safe design and classic protective measures with respect to the safety level and including the foreseeable behaviour of the users (manipulation of protective devices). But AmI-based safety solutions also lead to increasing complexity in (programmable) electronic systems. With this the requirements to which such safety technology is subject are different from those for conventional safety technology. A further objective of the project bundle is therefore to determine these specific safety requirements and, in a subsequent stage, to incorporate product development (e.g. by standardisation). The project bundle is thus intended to redirect the view of the product developers away from the purely technocentric approach and towards an anthropocentric, holistic one.

3.3 Project Bundle 3 "Head-Mounted Displays – Conditions for Safe and Strain-Optimized Use"

Augmented Reality (AR) describes the enrichment of the real world with virtual objects. AR is ambient intelligence if the correct information is made available at the right time without the user having to search actively for it. For the presentation of this information handhelds such as smartphones and tablet PCs are the most highly developed systems at present. But they have to be held in the hand and are therefore not practical for many working areas. Head-mounted displays (HMDs) solve this problem, but to date they have not been adequately studied with respect to their ergonomic aspects. In an occupational context the combination of AR with handhelds or HMDs offers many possibilities for increasing safety and productivity by reducing the requirement imposed on workers' memory capacity or by supporting decision-making in time-critical situations. Technicians in industrial repair can, for example, obtain information at the right time on what actions are required and what tools are to be used.

The aim of project bundle 3 is therefore to gather knowledge on the potentials and also on the adverse effects that may arise for the user and to generate from this recommendations for occupational use. The central question here is that concerning the tasks for which HMDs should be made available to the user as assistance. For this purpose in a sub-project suitable methods of task analysis are being identified, or they are being modified and applied to various industrial activities. This also yields information on the optimum presentation of information. In other sub-projects the physical effects (on the eye and the musculoskeletal system) and the psychological effects (including matters of acceptance) of the long-term use of HMDs under practical conditions are examined. In an experimental laboratory study the suitability of selected types of task and display are being examined (including a comparison with handhelds). Finally the knowledge gained from the sub-projects is incorporated in guidelines for the support of companies in decision-making regarding the use of AR technologies in combination with HMDs and/or handhelds.

3.4 Project Bundle 4 "Decision-Making Relevance and Behavioural Effects of Information on "Intelligent" Protective Clothing"

Rescue personnel, such as those in the fire brigade, are exposed to a large number of hazards during their assignments. Heat, polluted atmospheres, mechanical impacts and possibly even warfare agents make it difficult or impossible to rescue people. For some years, under the main heading of "smart personal protective equipment” (PPE) or "smart clothes", functional textiles have been developed which interact actively with the wearer and/or the environment and hence count as ambient intelligence technologies [6]. The aim is to reduce the number of incorrect decisions by providing an improved information base and to increase the use possibilities and the safety of task forces in extreme situations.

The project bundle "Decision-making relevance and behavioural effects of information on "intelligent" protective clothing" links up with third-party funded (Federal Ministry of Education and Research in Germany) research at BAuA. It studies resulting changes in the organisation of rescue personnel. In a number of sub-projects it is intended to analyse how smart clothes actively support people's action in the actual complexity of the assignments, and also what hazards may arise from them. Of special importance here are questions of the interpretation of information, the responsibility for decision-making and acceptance. In a sub-project, for instance, a mathematical model is being developed which interprets the vital parameter of the rescue person (e.g. heartbeat frequency, oxygen saturation of the blood) and ambient parameters (heat, pollutants etc.) and integrates them suitable for further decision-making. Smart PPE may also give rise to new risk or load factors, for example if the protective clothing suggests safety thus prompting the rescue person to behave in a careless fashion (risk compensation). A further sub-project deals with legal issues regarding data protection and personality rights

3.5 Project Bundle 5 "Possibilities and Opportunities of an AmI-Based Regulation of Ventilation and Air-Conditioning Systems"

When it comes to energy and physiological matters, the establishment of a comfortable, healthily conducive indoor climate, i.e. one which encourages performance, is one of the complex problems in the area of technical services in buildings. This is also due to the matter of inter-individual preferences. While there is extensive knowledge available on the theoretical relations between influencing factors, conventional ventilation and air-conditioning systems do not allow for individually oriented control mechanisms. This opens up the chances for an AmI-based building automation system: a microcomputer integrated in the building or working room uses appropriate sensors (e.g. temperature, humidity) to determine the actual situation and automatically regulates the state of the indoor climate by transmitting control signals to the building's existing technical systems as a function of the presence of people, where relevant also of their individual needs.

The aim of the project bundle is to study the possibilities and limits of climate regulation using conventional systems as compared to AmI systems and to formulate approaches to improving the regulation of ventilation and air-conditioning systems with the inclusion of AmI systems. This is done by way of example, taking the phenomenon of "dry air" in offices: for this purpose experiments are conducted in order to develop, test and implement a concept for laboratory and field study which encompasses the relations between room climate exposure variables (relative humidity, air movement, air temperature etc.) and physiological parameters (e.g. skin moistness, blinking, tear film break-up time). The knowledge acquired is to be applied in the programming of the AmI modules.

3.6 Project Bundle 6 "Lighting of Workplaces: Technology Assessment of AmI-Based Lighting Systems"

Facility management in the field of lighting is conducted today almost exclusively via time-controlled building technical services. But now AmI-based systems are also under discussion, systems in which light colour and brightness, for example, as well as office IT (illumination of Visual Display Units -VDUs) are controlled by means of set points. These AmI systems for workplace lighting are based on so-called dynamic light control. The major parameters which the AmI-based control influences are light sockets, light spectra and light distribution. In this way ambient parameters and work equipment exert a major influence on the workers' circadian rhythms (sleeping/waking rhythm) or physiological parameters. The effects on workers' safety, health and efficiency have not been adequately studied scientifically to date.

The aim of the project is to determine the chances and risks arising from an AmI-based control of the lighting of workplaces. An analysis is conducted of the effects of such control systems on selected parameters of the lighting (changes in light intensity, modulation of colour temperature and dynamic light control) and their physiological and psychological consequences (with respect to health, well-being and efficiency). This is conducted in laboratory research. Based on this, a synopsis is drawn up of the technology assessment for AmI-based lighting systems with regard to safety-relevant, health-relevant and efficiency-related questions.

4 Conclusions

In its current research and development programme BAuA has identified the emergence of new technologies as a primary challenge for occupational safety and health. These developments offer new perspectives for well known issues while at the same time invalidating established solutions. Both aspects require systematic research and active networking across Europe. BAuA aims at bringing forward this goal by connecting ongoing AmI-research to application in the world of work, conducting own research and funding.

References

1. Nakashima, H., Aghaja, H., Auguso, J.C. (eds.): Handbook of Ambient Intelligence and Smart Environments. Springer, New York (2009)
2. European Comission: Research and Innovation, `http://ec.europa.eu/research/home.cfm`
3. Aarts, E.H.L., Encarnação, J.L.: True Visions: The Emergence of Ambient Intelligence. Springer, Heidelberg (2006)
4. Kaptein, M., Duplinsky, S., Markopoulos, P.: Means based adaptive persuasive systems. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 335–344. ACM, New York (2011)
5. Fogg, B.J.: Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufmann, San Francisco (2002)
6. ISO Technical Report No. ISO/PDTR 00248435: Textiles and textile products – Smart textiles – Definitions, application and standardization needs. International Organization for Standardization ISO, Geneva, Switzerland

Living Labs as Educational Tool for Ambient Intelligence

Ben Kröse, Mettina Veenstra, Saskia Robben, and Marije Kanis

Amsterdam University of Applied Sciences, Duivendrechtsekade 36-38, 1096 AH, Amsterdam, The Netherlands
{B.J.A.Krose,M.Veenstra,S.M.B.Robben,M.Kanis}@hva.nl

Abstract. The way that innovation is currently done requires a new research methodology that enables co-creation and frequent, iterative evaluation in real-world settings. This paper describes the employment of the living lab methodology that corresponds to this need. Particularly, this paper presents the way that the Amsterdam University of Applies Sciences (HvA) incorporates living labs in its educational program with a particular focus on ambient intelligence. A number of examples are given to illustrate its place in the university's curriculum. Drawing on from this, problems and solutions are highlighted in a 'lessons learned' section.

Keywords: Ambient Intelligence, Living Labs, Education, participatory design.

1 Introduction

Current societal changes are important external drivers for new innovations in information technology. For example, in The Netherlands 'topsectoren' (innovation spearheads) are driven by grand societal challenges, such as global warming, ageing societies or security.

Also the *way* that innovation is done, is changing. Traditionally, new products used to be put on the market after a linear sequence of research and development, that was done in research labs and companies. Faster development and production techniques have brought on a new type of design methodology that has an iterative nature: products and services are developed in close collaboration with stakeholders (e.g. end users, distributers and financers) that are continuously tested and redesigned from early on. Furthermore, design is no longer solely focused on the product, but more directed towards the service and the user experience. On top of this, the design cycle should be open: stakeholders should work together.

The new way of innovating requires a new research methodology that enables co-creation, early testing and iterative evaluation in real-world settings. Such a methodology is the 'Living lab' methodology.

In this paper, we will focus on the importance of the living lab methodology for the education of new professionals. We will start describing the living lab concept, the context of our university and show the way of working by describing a number of projects. We will give some guidelines for implementation.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 356–363, 2012.

2 Living Lab Methodology

The concept of living labs was originally coined by Bill Mitchell at MIT, as an approach that represents a user-centric methodology for sensing, prototyping, validating and refining complex solutions in evolving real life contexts. The first Living Labs that were created from the initial idea happened in the area of smart/future homes. In those settings, real people (visitors) were observed in their usage of emerging technologies in the setting of a real home. In many of the implementations, people stayed in these homes for several days or weeks. New types of ICT, such as video based communication [6] and sensor monitoring [1] were studied. Later on, this concept was extended, in the sense that there was a strong emphasis on co-design, and participation of end-users. In [7] a number of cases have been described in the European context. However, according to Følstad [4], the current body of Living Lab research literature indicates a lack of common understanding of how Living Labs can be used for ICT innovation and development. Moreover, there appear to be only few cases describing how living labs are used in educational settings (such as done at the Amsterdam University of Applied Sciences). We consider a living lab as:

- an environment where partners are jointly involved in Research & Design
- a setting that exposes users to novel (ICT) solutions
- a (semi) realistic context
- suitable for longterm experiments
- a place for the evaluation of new developments as well as for the discovery of new possibilities

It needs:

- A physical set-up (such as a building or part of the city)
- An (ICT) infrastructure
- Access to end users
- Willingness to collaborate
- Evaluation tools (for doing research ‘in the wild’)

Two research groups at the Amsterdam University of Applied Sciences (‘Digital Life’ and ‘Interactive Public Spaces’) have adopted the living lab methodology in their educational programs, resulting in a living lab on public screens and a health living lab.

3 Embedding Living Labs in the Educational Program

The Amsterdam University of Applied Sciences (HvA, Hogeschool van Amsterdam), is focused on higher professional education and consists of seven schools. The school of Design and Communication runs a wide range of creative departments ranging from AMFI (Amsterdam Fashion Institute) to ITCS (Information Technology and Computer Science). These departments offer courses with creativity and innovation as its core elements.

CREATE-IT Applied Research is the research centre of the school of Design and Communication. At this centre, lecturers, students and researchers conduct practical and applied research on behalf of the creative and ICT industry, in collaboration with universities and other research institutes. The results of the research that is being conducted is aimed to highly benefit the professional field and are also used to help maintain the educational program on the cutting edge.

The research conducted can be called pragmatic, as the research relates to real-life cases and problems in the business segment. The general goals of the research are mainly directed towards the improvement and innovation of the professional practice. The research produced is aimed to be of high quality, and the research structure and methodology are aimed to fulfill all necessary criteria with regard to diligence, reliability, validity, impartiality and independence.

CREATE-IT's Applied Research have jointly set up a course on 'Ambient Interaction' in which students spend six months in research teams. These teams work on applied research projects, together with a (creative) company or institute that has particular real-life research questions. Also a number of students working on their bachelor's thesis were involved in these projects.

For these students we defined projects that were to be carried out in one of our living labs. The expected benefits of the approach were:

- The students collaborate with all stakeholders (end users, technology providers, roll out partners);
- The students are confronted with the entire cycle of user analysis, design, prototyping and user studies;
- Because of the existing infrastructure (ICT, end users) the students can start rapidly.

To offer more insights in our approach, a number of cases are given in the next two sections, in which we describe how students are involved in research in living labs. Then, in section 6 we present our lessons learned.

4 Public Spaces

In a dynamic context such as a public space, the situation and needs of the public constantly evolve and change over time. In a living lab, system designers and researchers are able to take this changing context into consideration so that systems can evolve simultaneously. BiebBeep is an example of a project that directly derived from using the living lab methodology and has been developed in the public space of a library in Almere, The Netherlands. This library as a living lab environment was used as explorative and educational space for students and has inspired continuous projects in this setting.

4.1 BiebBeep

To meet the demands of current and next-generation users, libraries –such as the one in Almere– nowadays face the challenge of innovating their physical and virtual services and utilizing digital media to provide state-of-the-art, so called Library 2.0 services. Casey [3] describes this Library 2.0 concept as a modernized form of library service whereby the focus lies on user-centered change and participation in the creation of content and community. BiebBeep, a large interactive touchscreen is an interactive screen that has been developed with the aim to augment the information and social function in the public space of a library. Two students from computer science and electrical engineering developed this together with Create-IT and the Novay research centre. BiebBeep displays user-generated and context-relevant content, such as information about local events and book trailers. The system's distinctive feature is that people can add information to the screen themselves, such as tweets and Flickr photos, so that the library and its visitors can inform and connect with each other. For more than a year, the BiebBeep system has been iterated and studied in the library to best meet the demands for its actual use in present and future. Other projects that stemmed from this research have focused on giving further shape to the Library 2.0 of the future, by experimental projects exploring how the library experience can be made more attractive, personal and sociable.

Fig. 1. Interactive touchscreen in the public library of Almere

5 Health-Lab

Health-Lab, a program in the metropolitan region of Amsterdam, focuses on innovative solutions for enabling people to live longer independently. In Health-lab, people from care institutions, research centers, educational institutes and companies work closely together with the end-users to co-create (technical) solutions. The HvA is participating partner in the Health-lab program and uses the program to stimulate innovation in education and research. Within this program, we have set up two ‘living lab’ locations, where real users can test applications in their daily life so to help designers and developers improve their products.

The first living lab, which was set in motion in 2006, was nursing home Naarderheem, The first projects conducted in this environment were focused on an intramural setting in which patients were monitored during the night. Soon after, a second ‘living lab’ came about, involving apartments that were built close to the nursing home and eventually equipped as ambient assisted living environments. Two typical projects in which students were involved are described in the next sections.

5.1 Monitoring ADL in Assisted Living Apartments

A number of projects conducted in our Living Labs evolve around the monitoring of daily activities in ambient assisted living environments. Engaging the elderly participants and other stakeholders through co-design plays a crucial role in these projects, but also the technical side of employing sensors for examining (deviations in) activity patterns in the home is studied in depth, e.g. [9]. Students were involved in both the technical- as well as the user- oriented studies.

One of the first user-centred studies involved the monitoring of daily activities from the perspective of medical specialists, in which interviews with medical experts and professional caregivers were carried out by two students to make an inventory of their needs for telemonitoring [2]. A second project, Senior-Create IT [5], which was done in collaboration with researchers and four students from the program ‘Ambient Interaction’, focused on the needs and attitudes of elderly people with regards to monitoring their daily activities with sensors in the home. Following on from this, 25 students from the university’s program on Communication and Multimedia Design were given the assignment to develop five iPad applications in groups of five that displays the ambient data in relevant and meaningful way to senior users. In the design of these products, elderly were also actively engaged in the design of these products, in order to contribute to the success and acceptance of such technology. Another student from a Technical Informatics course got then involved in linking the developed iPad applications with the actual systems in the Living Lab environment in Naarderheem, while graduation students in occupational therapy continued with more in-depth studies concerning elderly’s attitutes with regards to telemonitoring and also evaluated the developed iPad applications with the elderly users.

Fig. 2. In the Senior-Create IT project, the students used a mock-up for co-design

5.2 Interactive Wall in Psychogeriatric Ward of Nursing Home

A group of three students from the program ‘Ambient Interaction’ was asked to work on an assignment in nursing home Naarderheem, where approximately 60 elderly with severe dementia live in a psychogeriatric ward. People suffering from dementia often have a problem with way finding and are being restless. The students were asked to make an interactive installation for reducing the amount of wandering of the inhabitants of the ward. The interactive wall aims at making these people feel more at home in the nursing homes by guiding them with a motion triggered audio path and showing them images and short movie tracks from their hometown on large windows. The students started with observational studies in the nursing home and a literature study. Following on from this, they created some prototypes that were discussed with the caregivers. The head of the nursing home allocated a budget for building an interactive wall that was built by the students. After finalization of the wall, the students carried out an observational study, of which the results show that the interactive wall succeeds in attracting people and thus reducing the wandering behaviour. Remarks of the elderly as well as the family and caretakers support this conclusion.

6 Lessons Learned

During the five years that we adopted the Living Lab methodology and worked with students, we experienced a number of findings that relate to either the methodology or working with students. These are collated in the following lessons:

Engage the Stakeholders. The involvement of different stakeholders is of particular importance in a Living Lab to secure the development of usable and useful products and services, as also pointed out by [8]. In the BiebBeep project the involvement of for example the library staff was seen as very valuable, for example in stimulating interesting content. With regard to the Health-lab projects, all stakeholders (elderly, staff, management) were involved to enable the successful realization of the projects. Regular meetings with all stakeholders are needed to discuss progress and give feedback. However, in other projects not described here the role of the stakeholders was less positive. Some tended to modify the targets during the project, in some cases the hard- or software was not available in time and some stakeholders we just very slow in responding, with as a result too much delay in the project.

Take Care of Sufficient Technical Support. In many projects, the applications or services need to be tested for more than six months. During such a long period, many technical issues may arise, such as batteries that get empty, hard disks that crash or cleaners that pull out the plug of the router.. It is advised to have a watchdog function installed on the systems and still consider the maintenance, which is likely to cost a lot of effort.

Improve the Student's Skillset. In all of the projects, the students had to carry out user studies. We found that students differed very much in their communication skills, and often lacked basic research skills. Therefore, students are likely to need an intensive training and guidance before and during participating.

Embedding in the School Calendar. At the HvA, the student projects last 20 weeks and start either in September or February. This means that the infrastructure and all stakeholders need to be ready to start the projects at that particular time. Furthermore, 20 weeks is often too short for students to participate in the entire cycle of problem > research question >prototype building >user evaluation.

Evaluation of the Student Work. Teachers needed different evaluation methods as they found it difficult to evaluate the (inter-disciplinary) group achievements of students in the research projects, as it differs from the regular projects at HvA computer science, that are usually focussed on product building.

Students Learn to do Research Hands on. The positive aspects of getting students to do research in the wild is that students are actively made aware of the various real-world aspects of doing research. In the Health-lab, for example, the ethical part of doing research on monitoring systems was highlighted.

Use Multidisciplinary Groups. In the Health-lab projects, the students involved came from from different disciplines, such as occupational therapy and communication and media design. Such interdisciplinary approach needs some extra support (from the supervisors) in the beginning, but gave more thoughtful results from a broader perspective in the end.

7 Conclusion

This paper presented a myriad of projects to exemplify and explain how living labs can be used and employed in education settings, such as done at the Amsterdam University of Applied Sciences. The lessons learned that are presented in this paper increase the understanding of how Living Labs can be used for educational purposes, but also for ICT innovation and development.

Acknowledgments. The research reported in this paper has been supported by the Foundation Innovation Alliance (SIA) with funding from the Dutch Ministry of Education, Culture and Science (OCW), as part of the 'Smart Systems for Smart Services' and 'SpaceSee' projects, and via the 'Pieken in de Delta' program by the Ministry of Economic Affairs as part of the 'Health-lab' project.

References

[1] Abowd, G.D., Bobick, A.F., Essa, I.A., Mynatt, E.D., Rogers, W.A.: The aware home: A living laboratory for technologies for successful aging. In: Proceedings of the AAAI 2002 Workshop Automation as Caregiver, pp. 1–7 (2002)

[2] Alizadeh, S., Bakkes, S., Kanis, M., Rijken, M., Kröse, B.: Telemonitoring for assisted living residences: The medical specialists' view. In: Jordanova, M., Lievens, F. (eds.), In Proceedings of the Med-e-Tel 2011; The International eHealth, Telemedicine and Health ICT Forum for Educational, Networking and Business, pp. 75–78 (2011)

[3] Casey, M.E., Savastinuk, L.C.: Library 2.0-service for the next-generation library. Library Journal 131(14), 40–43 (2006)

[4] Følstad, A.: Living labs for innovation and development of information and communication technology: a literature review. The Electronic Journal for Virtual Organizations and Networks 10, 99–131 (2008)

[5] Kanis, M., Alizadeh, S., Groen, J., Khalili, M., Robben, S., Bakkes, S., Kröse, B.: Ambient Monitoring from an Elderly-Centred Design Perspective: What, Who and How. In: Keyson, D.V., Maher, M.L., Streitz, N., Cheok, A., Augusto, J.C., Wichert, R., Englebienne, G., Aghajan, H., Kröse, B.J.A. (eds.) AmI 2011. LNCS, vol. 7040, pp. 330–334. Springer, Heidelberg (2011)

[6] Markopoulos, P.: Towards a living lab research facility and a ubiquitous computing research programme. In: The CHI2001 workshop Distributed and Disappearing UI's in Ubiquitous Computing. Citeseer (2001)

[7] Schumacher, J., Niitamo, V.P., Vorarlberg, F.: European living labs: A new approach for human centric regional innovation. wvb, Wiss. Verl. (2008)

[8] Svensson, J., Ihlström Eriksson, C., Ebbesson, E., Åkesson, M.: Methods and techniques for user contribution: Challenges from a living lab perspective (2009)

[9] van Kasteren, T.L.M., Englebienne, G., Kröse, B.J.A.: Activity recognition using semi-Markov models on real world smart home datasets. J. Ambient Intell. Smart Environ. 2(3), 311–325 (2010)

Intel Collaborative Research Institute - Sustainable Connected Cities

Johannes Schöning[1,2,4], Yvonne Rogers[1,2], Jon Bird[1], Licia Capra[1,2], Julie A. McCann[2,3], David Prendergast[2], and Charles Sheridan[2]

[1] University College London, Gower Street, London, WC1E 6BT, UK
[2] Intel Collaborative Research Institute - Sustainable Connected Cities, London, UK
[3] Imperial College London, Huxley Bldg, London, SW7 2AZ, UK
[4] Hasselt University, Wetenschapspark 2, 3590 Diepenbeek, Belgium

Abstract. Cities are places where people, meet, exchange, work, live and interact. They bring people with different interests, experiences and knowledge close together. They are the centres of culture, economic development and social change. They offer many opportunities to innovate with technologies, from the infrastructures that underlie the sewers to computing in the cloud. One of the overarching goals of Intel's Collaborative Research Institute on Sustainable Connected Cities is to integrate the technological, economic and social needs of cities in ways that are sustainable and human-centred. Our objective is to inform, develop and evaluate services that enhance the quality of living in the city.

1 Motivation

There are many visions of the future focusing on smart cities, future cities and ambient intelligence. An underlying theme is that cities become increasingly embedded with invisible (and some still visible) technologies, sensitive and responsive to peoples needs that deliver advanced functions, services and experiences [1], following the vision of Mark Weise [2]. Intel's Collaborative Research Institute on Sustainable Connected Cities [1] is concerned with enhancing and changing how people live, interact and engage with cities. Our goal is to enhance city sustainability and improve citizen well being.

Our perspective in the Sustainable Connected Cities Institute is to be human-centred. We have wide-ranging expertise and background in user experience, interaction design, ethnography, together with research in the built environment, commerce, engineering, anthropology, the arts, and social psychology. We also work as inter-disciplinary teams that can make a real change to enrich and extend city dwellers lives. This fits in with Bells vision for 2012 [3] of computers not just acting on our behalf and anticipating what we want but also enabling people to be more creative, using state of the art computer technologies and toolkits.

[1] http://www.connected-cities.eu/

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 364–372, 2012.

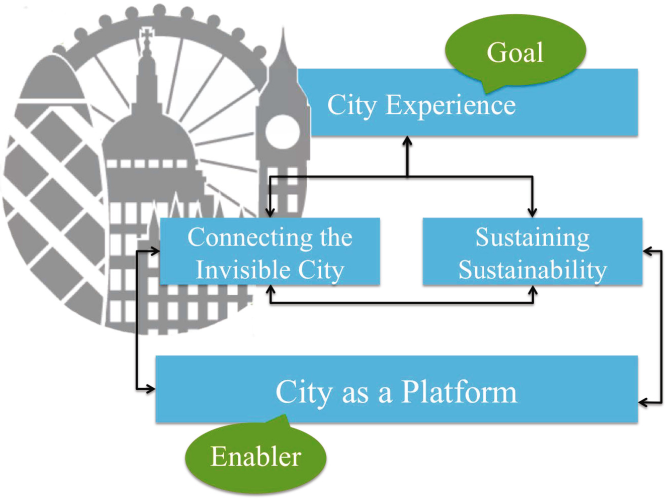


Fig. 1. Intel's Collaborative Research Institute on Sustainable Connected Cities thematic research agenda

There are major challenges that future cities will face not least a huge increase in the populations living in urban areas. According to a United Nation report [4] every second the global urban population grows by 2 people. Therefore the urban population is expected to increase from 3.6 billion people in 2011 to 6.3 billion in 2050. In 2020 more than 700 cities will exist with populations of +1million; today we have just 500 cities with populations of +1million. The exploding urban population growth creates unprecedented challenges, among which provision for water and sanitation are the most pressing and painfully felt when lacking [5]. Cities cannot be sustainable without ensuring reliable access to safe drinking water and adequate sanitation.

A focus will be on the techniques and solutions to tackle these future challenges. We will develop and exploit pervasive and sensing technologies, analytics and new interfaces, putting humans at the centre of technological developments. Our approach is to address four main themes (see figure 1)

– City Experience: How do we enhance the City Experience and communicate services?
– City as a Platform: How do we create the digital platform of the city from sensor/edge to cloud?
– Sustaining Sustainability: How to sustain behavioural change?
– Connecting the Invisible City: How do we visualize the Human-Environment Interface?

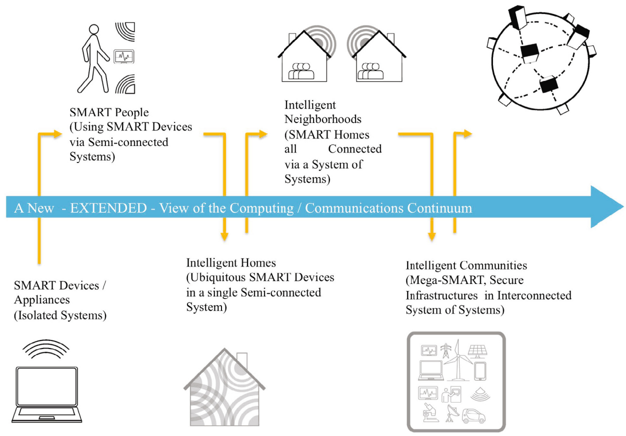


Fig. 2. Path to Connected Cities

2 Connected Cities

We consider how cities will become instrumented environments that serve the needs of their citizens (see figure 2). Inspired by the Internet of Things (IoT), a number of projects are beginning to develop connected products that can be embedded in the environment and people's homes. In addition, our working, learning and recreation places are becoming more equipped with technologies and sensors. Secure infrastructures are essential for such interconnected systems of systems. Besides the question of how to support such an elastic city platform from sensor/edge to cloud (see [6]), it is necessary to research how citizens interface with such an interconnected system of technologies and services.

Another growing concern, given limited resources, is how cities can be more sustainable and encourage their citizens to use utilities (water, electricity, gas, etc.) in more resourceful ways.

3 Sustaining Sustainability

Helping citizens adapt their behaviour in order make cities more sustainable involves increasing their awareness of how they live and then encourage changing habits, at an individual, family, local community and city level. This requires adopting a broad unit (ranging from the individual to communities at large) of analysis: considering the needs of the individual city dweller, families, whole neighbourhoods, councils, and communities at large.

There are a number of behavioural change techniques that have been identified in social psychology and behavioural science. One of the most well known approaches used in behavioural therapy is the Transtheoretical Model. This predicts a persons success or failure in achieving a proposed behaviour (e.g., dieting) in terms of their preparedness to change as described by Prochaska [7]. Despite its popularity, however, recent reviews of a large number of empirical studies have shown little evidence for the effectiveness of stage-based interventions as a basis for behavioural change (e.g., Aveyard [8]).

Other promising techniques that have been suggested for use in behavioural change include peer pressure, positive reinforcement, and social norms [9]. Peer pressure has been used to help people change their behaviour for a variety of health-related areas, such as alcoholism, obesity and smoking. Computer applications that have tried to capitalise on this approach have displayed data, such as the amount of exercise completed on a mobile device, comparing how members of a group are doing relative to each other (e.g [10]). An innovative prototype, that used both positive reinforcement and peer pressure, was the WaterBot system, designed to help householders reduce their usage of water in their homes [11]. There is much evidence to suggest that people are wasteful with water, often leaving the tap running continuously for long periods of time while cleaning their teeth or washing. The research team thought that the use of monitoring technology could help persuade householders to change their behaviour and be more conservative in their water usage. To this end, they used the theory of positive reinforcement to inform their design, which states that activities are likely to be repeated if some kind of reward is given occasionally and randomly (similar to the reward system used in slot machines). A sensor-based system was developed where positive auditory messages and chimes were sounded when the tap was turned off. Here, the idea was to encourage peer pressure and for the members of the household to talk to each other about their water usage.

Since this study, the global concern about climate change has led a number of researchers to design and evaluate various energy sensing devices that display real-time feedback. A goal is to find ways of helping people reduce their energy consumption (and is part of a larger research agenda called sustainable HCI, e.g., Mankoff [12]). A focus is on persuading people to change their everyday habits with respect to environmental concerns, such as reducing their own carbon footprint. The effect of social norms on people's energy consumption has also begun to be studied. For example, Schultz et al. [13] looked at how people behaved when provided with information that compared their electricity usage to the neighbourhood average. They found that providing the average had a big impact on households own energy consumption. As hoped, households above the average tended to decrease their consumption but those using less electricity than average tended to increase their consumption. The study suggested that such boomerang effects could be counteracted by providing households with further salient information that was emotive. When provided with an emoticon along with the numerical information about their energy usage, households using greater than average decreased their consumption more if they were given a sad icon and, significantly, households using less energy than average continued to do so if they received a smiley icon.

In the last few years, there have been many research and council led initiatives in cities across the world to reduce energy consumption, from reducing street lighting to implementing smart metering. Some have shown impressive results. However, many of these have proven to be short-lived. Moreover, there is a tendency to return to old habits once the local champion, publicity intervention, etc. have been taken away. A key question, which we will address at the centre, is: How can sustainable behaviour in its various forms be sustained over a long period of time, preferably indefinitely? What mix of policies and technologies can be used to best effect? Which behaviours are most amenable? How do communities take on the sustainable challenge themselves and understand what it takes?

We intend to develop a science of behavioural change that is predictive and generalizable to different contexts; longitudinal empirical studies will be carried out to investigate long-term effects. We further argue that the efficacy of the techniques and methods used will be affected by how ethical they are. The aim of this theme is to investigate how behaviour can be changed effectively, is socially acceptable and will persist over a variety of contexts and settings. The overarching goal is to engage citizens proactively with new kinds of technologically augmented information in different aspect of their lives and cities. Moreover, we intend to involve them directly in identifying problem behaviours they care about in city life, generating prototype designs and actively participating in the evaluation studies.

The research intends to push the frontiers of the science of behavioural change by systematically addressing many of the assumptions and unknowns in this exciting new field, using a three-pronged approach:

- designing and implementing a range of new pervasive technologies that can facilitate behaviour change by operationalizing theories from behavioural economics and social psychology
- assessing how new kinds of information and multimodal real-time feedback are best delivered by pervasive technologies and which are the most effective techniques for different contexts and behaviours
- ascertaining whether and how salient information can lead people to change their behaviour in both the short-term and the long-term.

A key objective is to show how different combinations of technologies, behaviour techniques and salient information can systematically facilitate behaviour change, with a focus on those behaviours that either have not been considered before or have been resistant to change using other methods. A further goal is to design technologies that are affordable and customisable so that they can be adopted by individuals and communities who have a problem they wish to address - for example, they may wish to reduce vandalism in their neighbourhood, encourage more volunteering or increase local shopping.

Fig. 3. Tidy Street: Measuring and Publicly Displaying Domestic Electricity Consumption (Photo taken by Nora O'Murchu)

3.1 Tidy Street: An Example of Sustaining Sustainability

Our own research has shown that publicly displaying households electricity consumption can have a significant impact on their energy usage [14]. Participants reported an increased awareness of their electricity usage together with reduced their electricity usage by 15%. A display sprayed with chalk on the street based on social norms, that represent both the communitys average compared to the city at large, can be seen in figure 3.

4 Connecting the Invisible City: Visualizing the Human-Environment Interface

The Connecting the Invisible City theme focuses on how technology can help recognize, leverage, and support the out-of-sight, hidden or forgotten resources of urban environments from volunteers to subterranean water systems and other underlying city infrastructures. In future cities a lot of data streams and information will be embedded and stored within their infrastructure. Besides determining new ways of how to store, save and update all this information within complex infrastructures, new ways of thinking about and analysing information will need to be developed. A fundamental question is what novel multimodal interfaces and interactions are required to encourage participation of citizens, business and government. Over the last few years, advances in graphical interfaces (e.g., the iPhone UI), speech recognition (e.g. Siri), gesture and handwriting recognition (e.g., Kinect), together with the arrival of the mobile broadband, smartphones, sensor technologies, and an assortment of other new technologies providing large and small interactive displays, have changed the face of human-computer interaction [15].

A challenge is to develop displays, services and apps that can visualize the invisible information flows in future cities and help people to make informed decisions during their daily routines. We will investigate and develop a framework for Human Environment Interfaces (HEI) that lets individuals and groups engage with the information available in the city. But how do we visualize the HEI? What resources are visible/invisible? What actors are invisible/invisible? How and where should city information be represented? Possible visualisations include aggregation of quantified self and community data, via ambient displays, mobile devices and public signage. Our research will focus on the following topics:

– The development of novel interaction techniques which will afford interaction, to help participants to discover services and data around them.
– The development of services or interfaces that turn data into information and help people to make better informed decisions.
– The development of technologies to encourage sustainable behaviour through ambient and invisible interfaces which capture information relating to citizens behaviour.
– The development of interaction techniques that connect people to their cities..

The topics are wide-ranging from making invisible data visible, turning data into useful information, supporting sustainable behaviour and helping citizens to experience their cities in new ways. We will focus on providing information and experiences instead of pure data and facts. The recent range of technological developments described above has encouraged different ways of thinking about interaction design. Researchers and developers have combined the physical and digital in novel ways, resulting in mixed realities, augmented realities, tangible interfaces, and wearable computing. Overall, it is important to design multi-modal techniques (addressing the visual, hearing, and haptic senses), which provide the right degree of abstraction for each citizen in each context.

4.1 PhotoMap: An Example for New Ways of Interaction within Cities

In the following we try to describe, based on our own research, new ways of interacting with a city platform. Other examples (looking at different perspectives on this problem) are described in [16,17]. In many mid-to large-sized cities public maps are ubiquitous. These public maps help to facilitate orientation and provide special information to tourists but also to locals who just want to look up an unfamiliar place while on the go. These maps offer many advantages compared to mobile maps from services like Google Maps Mobile or Nokia Maps. They often show local landmarks and sights that are not shown on standard digital maps. Often these YOU ARE HERE (YAH) maps are adapted to a special use case, e.g. a zoo map or a hiking map of a certain area and tailored to a specific usage context. Being designed for a fashioned purpose these maps are often aesthetically well designed and their usage is therefore more pleasant.

But how do we bring this information to the city platform to encourage sustainable behaviour? The Photomap application [18] is a novel technique and application that uses images of YAH maps taken with a GPS-enhanced smart phone as background maps for on-the-y navigation tasks. We have shown that pedestrians can georeference the taken images with sufcient accuracy to support navigation tasks and feed data to the platform.

5 Summary

The aim of the new institute is to create and realize a compelling vision of a sustainable future made possible by adaptive technologies that optimize resource efficiency, enable new services and support the quality of life of urban inhabitants. Our vision is to enhance city sustainability and improve citizen well-being. For our research the city of London will be the main testbed. It will be important to research what technologies are needed that can best identify, model, and promote understanding of the impacts of urban heat and climate change on the people and infrastructure. The project will provide new understandings and insights into how cities can be better instrumented and citizens better informed about the resources and utilities they use. Of course, London will be not the only testbed in Europe. Various cities exists that will face important challenges within the next few years. In addition the Urbanization will heavily effect other parts of the world, e.g. Africa and Asia, and will of course need attention on a global scale. We hope to collaborate with others to innovate in and with communities, within Europe and within other parts of the world, using sustainable affordable technologies that can transform cities.

References

1. Aarts, E.: Ambient intelligence: A multimedia perspective. IEEE Multimedia 11(1), 12–19 (2004)
2. Weiser, M.: The computer for the 21st century. Scientific American 265(3), 94–104 (1991)
3. Bell, G., Dourish, P.: Yesterdays tomorrows: notes on ubiquitous computings dominant vision. Personal and Ubiquitous Computing 11(2), 133–143 (2007)
4. Department of Economic: The Millennium Development Goals Report 2010. United Nations Publications (2010)
5. United Nations Development Program. Human Development Report Office: Human development report: Background papers. Human Development Report Office, United Nations Development Programme (2001)
6. McCann, J.A.: Engineering beauty with the beasties on wireless sensor networks. Perada Magazine (2008)
7. Prochaska, J.O., Norcross, J.C.: Systems of psychotherapy: A transtheoretical analysis. Brooks/Cole Pub. Co. (2009)

8. Aveyard, P., Lawrence, T., Cheng, K., Griffin, C., Croghan, E., Johnson, C.: A randomized controlled trial of smoking cessation for pregnant women to test the effect of a transtheoretical model-based intervention on movement in stage and interaction with baseline stage. British journal of health psychology 11(2), 263–278 (2006)
9. Consolvo, S., McDonald, D., Landay, J.: Theory-driven design strategies for technologies that support behavior change in everyday life. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 405–414. ACM (2009)
10. Toscos, T., Faber, A., An, S., Gandhi, M.: Chick clique: persuasive technology to motivate teenage girls to exercise. In: CHI 2006 Extended Abstracts on Human factors in Computing Systems, pp. 1873–1878. ACM (2006)
11. Arroyo, E., Bonanni, L., Selker, T.: Waterbot: exploring feedback and persuasive techniques at the sink. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 631–639. ACM (2005)
12. Mankoff, J., Kravets, R., Blevis, E.: Some computer science issues in creating a sustainable world. Computer 41(8), 102–105 (2008)
13. Schultz, P., Nolan, J., Cialdini, R., Goldstein, N., Griskevicius, V.: The constructive, destructive, and reconstructive power of social norms. Psychological Science 18(5), 429–434 (2007)
14. Bird, J., Rogers, Y.: The pulse of tidy street: Measuring and publicly displaying domestic electricity consumption. In: Workshop on Energy Awareness and Conservation through Pervasive Applications, Pervasive 2010 (2010)
15. Preece, J., Rogers, Y., Sharp, H.: Interaction design: beyond human-computer interaction. Wiley (2002)
16. Böhmer, M., Hecht, B., Schöning, J., Krüger, A., Bauer, G.: Falling asleep with angry birds, facebook and kindle: a large scale study on mobile application usage. In: Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 47–56. ACM (2011)
17. Schöning, J., Hecht, B., Starosielski, N.: Evaluating automatically generated location-based stories for tourists. In: CHI 2008 Extended Abstracts on Human Factors in Computing Systems, pp. 2937–2942. ACM (2008)
18. Schöning, J., Krüger, A., Cheverst, K., Rohs, M., Löchtefeld, M., Taher, F.: Photomap: using spontaneously taken images of public maps for pedestrian navigation tasks on mobile devices. In: Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, p. 14. ACM (2009)

IE Sim – A Flexible Tool for the Simulation of Data Generated within Intelligent Environments

Jonathan Synnott, Liming Chen, Chris Nugent, and George Moore

The Computer Science Research Institute and the School of Computing and Mathematics, University of Ulster, Jordanstown, Northern Ireland, UK
synnott-j2@email.ulster.ac.uk,
{l.chen,cd.nugent,g.moore}@ulster.ac.uk

Abstract. Availability of datasets generated by Intelligent Environments for the testing of new approaches may be limited due to constraints including time, space, and money. The use of simulated Intelligent Environments offers a method of creating datasets with maximum control and minimal costs and constraints. Such datasets facilitate the testing of novel approaches to areas such as activity recognition and ambient assisted living. IE Sim is a flexible feature-rich approach that supports graphical interactive construction and simulation of virtual Intelligent Environments This paper discusses the key features of IE Sim and discusses the results of a software evaluation performed by 21 international researchers with an interest in such datasets. Results from the evaluation rated IE Sim highly in terms of ease of use and usefulness in research and identified key requirements for future developments.

Keywords: Data Simulation, Virtual Environment Creation, Virtual Sensors, Intelligent Environments, Environment Prototyping.

1 Introduction

Intelligent Environments (IEs) are becoming increasingly prevalent in the area of healthcare management [1]. Researchers require access to datasets containing rich sensor data in order to develop and test new approaches in fields that utilize this data, such as activity recognition and ambient assisted living. Access to such datasets are limited due to the various barriers associated with the creation of physical implementations. The creation of such environments is resource intensive in terms of time and money, requiring careful long-term planning and being subject to resource constraints such as space, budget, and availability of technology. These constraints may also limit environment expansion in terms of space increases or the addition of new sensor technology. As a result such environments may lack the scalability to reflect that of real world applications [2]. Additionally, physical IEs require the use of human subjects in order to generate rich data. This provides further limitations, as such subjects may be difficult to recruit and it would not be feasible for such subjects to test all possible scenarios within the environment in a naturalistic manner [3].

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 373–378, 2012.

Data simulation approaches can alleviate these constraints in several ways. Appropriate software can facilitate the rapid and low-cost creation and expansion of virtual environments with no spatial constraints. Additionally, virtual sensors can be designed and incorporated into such environments, facilitating the inclusion of unlimited numbers of sensors which may be otherwise unavailable due to expense or current technical capabilities. Finally, the use of virtual inhabitants within these environments can facilitate the repeat testing of large numbers of scenarios which may not be possible with real participants due to ethical, safety, or practicality concerns. The detailed testing that such environments facilitate ultimately promotes the creation of more robust approaches [3].

A number of approaches have been considered in the research domain of IE simulation. These approaches include parameterized activity model approaches as used by PerSim [2] that facilitate the simulation of activity data based on pre-defined parameters. These approaches are capable of generating data representing a range of typical activity performances over extended periods of time, however, the accuracy of the data generated is dependent on the quality of the pre-defined activity models and the method of inserting anomalous event occurrences is unintuitive. IE Sim provides an interactive graphical approach to virtual environment creation and simulation, promoting full control over simulated inhabitant activities. This provides the ability to record specific activities within a virtual environment, allowing users to make subtle changes to activity performance based on specific test cases. IE Sim has been developed to offer improvements over existing research [2] [4] by providing a flexible and customizable simulation environment in which users can construct environments and create new objects or sensor types using intuitive interfaces without the need for third party software or the use of development languages.

Section 2 provides an overview of IE Sim's key features and methods of implementation, Section 3 provides results from a user evaluation, and Section 4 provides concluding remarks with suggestions for future work.

2 IE Sim and Implementation

IE Sim has been designed to extend existing research within the domain of IE data simulation by facilitating rapid creation, customization and interaction with virtual IEs with minimal constraints. This Section discusses the key features of IE Sim and the methods used to implement them. Fig. 1 provides an overview of the main component of the IE Sim User Interface (UI).

2.1 Virtual Environment Creation

Virtual environments in IE Sim are presented using a 2D floor-plan layout, chosen for its simplicity, clarity and ease of customization. The method of environment construction makes use of an object toolbox which allows users to click and drag new objects into the environment. A properties panel (Fig. 1-B) allows users to adjust the properties of any objects placed within the environment, facilitating fine tuning of object

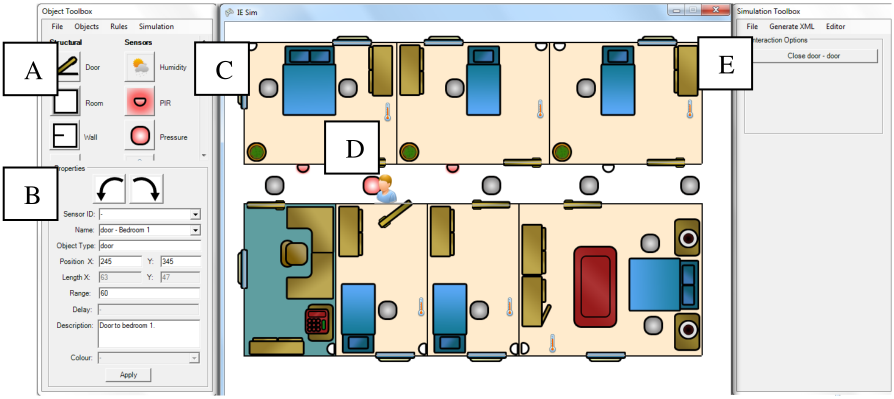


Fig. 1. An overview of the IE Sim environment creation UI. A - The object toolbox; B- The object properties panel; C - A virtual environment representing a care home; D- A user-controlled avatar; E- The object activation menu.

size, placement, orientation, association with physical sensors, and data generation characteristics. This method of environment creation was originally introduced in a previous study that facilitated the construction of static virtual IEs for real-time and longitudinal visualization of activities within physical IEs [5]. IE Sim significantly expands this concept to facilitate the creation of virtual IEs that are interactive rather than static, with the addition of virtual objects and sensors that are fully interactive and generate virtual data, catering for the creation of environments that may represent a real physical implementation of an environment, a conceptual environment, or a hybrid consisting of both physical and simulated objects. Fig. 1-C provides an example of a virtual IE created in IE Sim, representing a care home. These environments can be saved as XML files for future use.

2.2 Generation of Simulated Sensor Data

IE Sim supports the generation of simulated data through direct user interaction with virtual environments. Once a virtual environment has been created, users can interact with the environment through the use of a virtual avatar (Fig. 1-D). The avatar can be moved through the environment using the arrows keys on the keyboard. Virtual sensor data is generated based on the avatar's movements within the environment and avatar interactions with objects within the environment. Interaction with sensors may be passive (e.g. with PIR sensors or pressure sensors) or active (e.g. when opening doors or cupboards). Active interaction with objects is facilitated through the use of an object activation menu that lists all actions that are available based on the avatar's current location (Fig. 1 - E). Objects that are currently being interacted with indicate their status through a red glow. Once the user has completed a scenario, virtual sensor data based on the user's interactions can be output in the HomeML format [6] by selecting the 'Generate XML' button on the Simulation Toolbox.

2.3 Creation of Custom Objects and Sensors

Environment objects within IE Sim are objects with an associated appearance, states, and interaction options. Sensors are defined as objects that output data upon interaction. The object toolbox (Fig.1-A) provides a range of ready-made virtual object and sensors to be deployed within virtual IEs. Users may also create custom objects and sensors that meet specific requirements, facilitating the specification of object type, appearance, and interaction characteristics. This functionality promotes applicability of the tool to a wide range of environment types, and ensures extensibility of the tool, supporting adjustment of sensor specification as requirements change or as technology with improves capabilities is developed. The creation of custom objects is supported through the object editor. This is a form based UI that allows users to specify the properties of new objects. The main properties include: *Passable, ActivationType, ActivationDelay, Range, ImageSources,* and *DataValues*.

The *Passable* property is used during collision detection, and indicates whether the avatar is able to pass through the object. This property may be set to *True* for objects such as pressure sensors, and set to *False* for objects such as beds and tables.

The object editor supports the creation of objects with four different interaction methods, specified by the *ActivationType* property. *Active* objects are those that must be actively engaged with using the object activation menu (Fig. 1-E). As the avatar navigates throughout an environment, the object activation menu updates to show available interaction options, based on the avatar's distance from active objects. Users can interact with active objects when the avatar's distance is less than the *Range* property of that object. When the user selects the object in the object activation menu, the object will switch between states, changing in appearance and outputting the values specified in the *State1 Val* or *State2 Val* fields. *Auto* objects are sensors that trigger and generate data at a constant rate (defined by the *ActivationDelay* property), independent of the user's actions. *Proximity* objects are sensors that trigger automatically once the avatar meets certain criteria, including being in the same room as the object, and being within the distance specified by the *Range* property. Once the avatar meets these criteria, these sensors will trigger at the rate specified by the *Activation Delay* property. Objects with an activation type of *None* are static objects that are not interactive, have a single state and do not generate data. These objects are typically used for adding additional complexity to environment layouts, such as walls, desks, or kitchen workbenches that may impact on the avatar's movement path throughout the environment.

The *Image Source* fields may be used to set the images (in *.png* format) that represent the objects within the virtual IE. *Active* and *Proximity* objects have two states (i.e. triggered or not triggered; open or closed), and therefore two images must be specified in order to depict the change in state.

Custom objects are stored within an XML file and are automatically imported into the application upon loading.

3 Evaluation

Evaluation of IE Sim was performed through a usability questionnaire. The questionnaire was designed to assess the utility and the ease of use of the key components of the software. Participants in the study consisted of researchers whose work related to the use of data generated by intelligent environments. In total, 21 participants completed the evaluation. These participants were recruited at an international conference, and during research visits to the Smart Environment Research Group [7]. These participants came from research institutes based in the America, Asia, and Europe. All participants were provided with a live demonstration of the software, and were given the opportunity to engage with the software before completing the questionnaire.

Participants were asked to rate each key feature of the software using a five point scale in terms of ease of use (0 = Very difficult to use, 5 = Very easy to use) and usefulness in their research (0 = Not useful, 5 = Very Useful). Fig. 2 presents the mean results from the questionnaire.

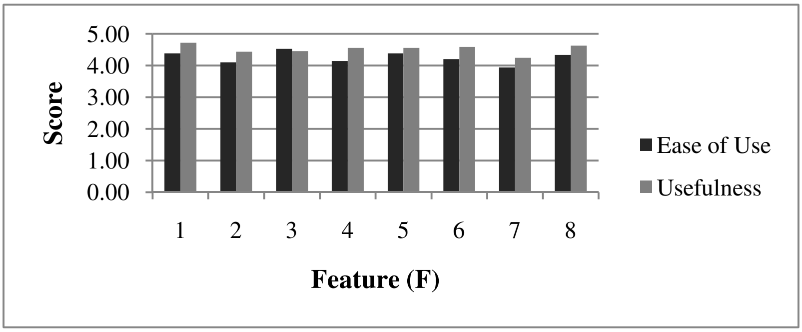


Fig. 2. A bar chart illustrating the mean feature scores collected during the IE Sim software evaluation. 21 participants rated the features (F) of the software in terms of ease of use (0 = Very difficult to use, 5 = Very easy to use) and usefulness within research (0 = Not useful, 5 = Very Useful). F1 = Environment layout creation and population with objects/sensors, F2 = Sensor property adjustment, F3 = Visualization of environments using a 2D floor-plan approach, F4 = Generation of simulated datasets, F5 = Navigation of the avatar throughout the environment for passive and active sensor interaction, F6 = Interaction with objects using the object activation menu, F7 = Creation of custom sensors and F8 = The complete software package.

Participants responded positively to all of the key features, rating each feature highly in terms of ease of use and usefulness in research. No major usability issues were identified, and qualitative feedback from participants highlighted a number of areas for expansion and future work. In addition to rating the key features of the software, participants were asked a series of true/false questions in relation to other aspects of the software. 90.48% of the participants responded that the software would be of use to them in their research. 71.43% of the participants indicated that the software would provide an advantage over the methods they currently utilize for data visualization and simulation. Finally, 85.71% of the participants indicated that they would be interested in the ability to record multiple avatars interacting with the virtual environments simultaneously to simulate sensor data representing multiple occupancy.

4 Conclusion and Future Work

This manuscript has introduced IE Sim, a tool for the simulation of data generated within IEs. The tool facilitates the creation of virtual environments that can be populated with objects and sensors and interacted with using a virtual avatar. Flexibility and extensibility of the software is supported through the ability to create custom object and sensor types for use within the virtual environments. An evaluation of the software was completed by 21 international researchers working within the area of IEs. Results from the evaluation provided positive scores for all key features of IE Sim in terms of ease of use and usefulness in research. 90.48% of participants indicated that IE Sim would be of use to them in research, and 71.43% indicated that IE Sim would provide advantages over existing methods of visualization and simulation. The evaluation also revealed a number of future work areas that were of interest to the research community. In particular, support for the recording of data representing multiple occupancy was of interest to 85.71% of the participants who took part in the evaluation. Other areas for future work included support for PIR occlusion and sensor data noise. IE Sim development is currently ongoing, and these areas, amongst others, will be addressed in future versions. Future evaluation will involve use of the tool in ambient assisted living and activity recognition projects.

Acknowledgements. This work was funded by the Department of Employment and Learning, Northern Ireland and the Healthcare Informatics Society of Ireland Research Bursary.

References

1. Chan, M., Campo, E., Estève, D., Fourniols, J.Y.: Smart Homes - Current Features and Future Perspectives. Maturitas 64(2), 90–97 (2009)
2. Armac, I., Retkowitz, D.: Simulation of Smart Environments. In: IEEE International Conference on Pervasive Services, Istanbul, pp. 257–266 (2007)
3. Helal, S., Lee, J.W., Hossain, S., Kim, E., Hagras, H., Cook, D.: Persim - Simulator for Human Activities in Pervasive Spaces. In: 7th International Conference on Intelligent Environments, Nottingham, pp. 192–199 (2011)
4. Lei, Z., Yue, S., Yu, C., Yuanchun, S.: SHSim: An OSGI-Based Smart Home Simulator. In: 3rd IEEE International Conference on Ubi-Media Computing, Jinhua, pp. 87–90 (2010)
5. Synnott, J., Chen, L., Nugent, C., Moore, G.: Flexible and Customizable Visualization of Data Generated within Intelligent Environments. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego (to be published, 2012)
6. HomeML, `http://www.home-ml.org/Browser/index.php`
7. Smart Environments Research Group, `http://scm.ulster.ac.uk/~scmresearch/SERG/index.html`

Intention Recognition with Clustering

Fariba Sadri, Weikun Wang, and Afroditi Xafi

Imperial College London
{fs,ww2210,ax10}@doc.ic.ac.uk

Abstract. Intention recognition has significant applications in ambient intelligence, assisted living and care of the elderly, amongst others. In this paper we explore an approach to intention recognition based on clustering. To this end we show how to map the intention recognition problem into a clustering problem. To our knowledge the use of clustering techniques for intention recognition is novel, and this paper suggests it is promising.

Keywords: Intention recognition, clustering, Fuzzy C-means.

1 Introduction

Intention recognition (IR) is the problem of recognising the intentions (or synonymously here, the goals) of an agent by (incrementally) observing its actions. Many applications of intention recognition have been explored, including ambient intelligence, elder care (e.g. [9]), and prediction of military maneuvers (e.g. [8]).

Ambient intelligence (AMI) environments must be capable of anticipating the needs, desires and behaviour of their inhabitants [1] in order to provide suitable support to the inhabitants. Intention recognition can make a significant contribution to AMI systems by enabling and enriching their anticipatory capabilities. Various techniques have been used for intention recognition. The most common of these are logic-based (e.g. [4, 11]), case-based (e.g. [3]) and probabilistic approaches (e.g. [6, 9]).

In this paper we explore the use of clustering techniques for intention recognition. Clustering is the task of classifying objects into groups in such a way that the objects in each group are more "similar" to one another than to objects outside the group. Clustering is more commonly applied to pattern recognition, image analysis, information retrieval, and bioinformatics. To our knowledge the application of clustering to intention recognition is novel.

In order to map intention recognition to a clustering problem, we have to overcome several difficulties. For example clustering is usually applied to elements modeled in Euclidean spaces, and we have to find a way of crafting the intention recognition problem as a clustering problem. Intuitively the fundamental functionality of an IR system is to classify observed actions into intentions (and plans to achieve intentions). Thus actions "related" to one another, according to some suitable criteria, have to be grouped within clusters identifying potential intentions.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 379–384, 2012.

Thus we must map plans and actions to a format suitable for clustering, and we must do so in a robust fashion that can deal with noisy and partial data. To this end we need to devise a measure of “relatedness” or “similarity” between actions, and we need to devise a way of interpreting the result of the clustering, to associate an intention with each cluster, and a ranking with each intention indicating its likelihood, given some observed actions.

In this work we show how we can overcome these difficulties to build a bridge between the two fields of intention recognition and machine learning via clustering.

2 Background

2.1 Intention Recognition (IR)

The input to an IR system usually consists of a sequence of observed actions (actions executed by an agent whose intention is being determined), and either a plan library, providing plans for intentions, or an action theory describing the semantics of actions in terms of their pre- and post-conditions. The task of the IR system then is to determine the most likely intention(s) the observed agent is trying to achieve.

Several approaches to IR have been proposed. For example, Demolombe and Fernandez [4] uses logic-based specifications of macro-actions, Sadri [11] and Hong [7] map reasoning about intentions with logic-based theories of causality into problems of graph generation and path finding, and Geib and Goldman [6] use probabilistic techniques. A survey of the logic-based approaches can be found in [10, 12].

2.2 Clustering Techniques

Clustering is an unsupervised learning technique and involves steps shown in figure 1. Clustering algorithms may be *exclusive*, classifying objects into non-overlapping clusters, or *fuzzy* allowing overlapping clusters, where each object belongs to each cluster to a certain degree. For our work we have chosen the fuzzy C-means algorithm (FCM) [5]. Fuzziness is needed because an action may be part of plans for achieving more than one intention. For example *getting milk from the fridge* may be an action in a plan for *making tea* and a plan for *making porridge*.

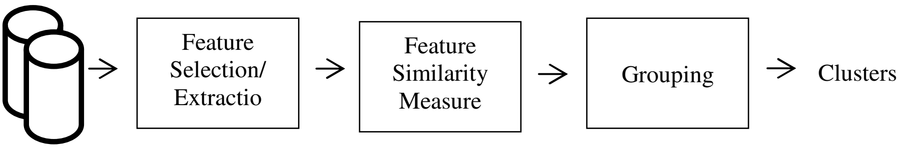


Fig. 1. Clustering procedure

The similarity measure used for clustering is dependent on the domain of the data and the feature extraction applied. For instance, when data entries are represented as points in a Euclidean space, each dimension represents a feature that has descriptive power, and the Euclidean distance can be used to compare proximity of two points.

3 The Intention Recognition Task

We assume we have a plan library. Example 3.1 shows a simple plan library which will be used as our running example. The library consists of plans for three intentions. In general an intention may have any number of associated plans.

Example 3.1.
Intention I1: *Make Tea* Plan 1: 1, 2, 3, 4, 5
Intention I2: *Make Cocoa* Plan 2: 1, 2, 6, 7, 5
Intention I3: *Make Breakfast* Plan 3: 1, 8, 9, 10, 11
where the numbers correspond to actions as follows:

1	2	3	4	5	6	7	8	9	10	11
get milk	get cup	put tea-bag in cup	pour boiled water in cup	add milk to cup	boil milk	put cocoa in cup	get bowl	put cereal in bowl	pour milk in bowl	add sugar to bowl

The agent may have multiple intentions, may be interleaving action executions in pursuit of these intentions, and may make mistakes, or the sensor data may be faulty. We observe the actions the agent executes. Observations can be *partial*, in that we may not observe every action that the agent executes. They may also be *noisy*, in that, for example due to sensor faults, we may observe actions incorrectly, or the agent may execute an action by mistake or towards an intention that he later abandons.

Example 3.2. Given the plans above, sequence S1 is a partial sequence of observations, S2 is noisy, and S3 is an interleaved partial sequence that goes towards achieving both intentions 1 and 3 (quite a likely sequence when one is preparing breakfast!).

S1= 1; 6 S2= 1; 2; 12; 3 S3= 1; 2; 8; 3; 9.

Given a set of intention I={I_1, I_2, …, I_n}, a library L of plans for these intentions, a sequence of observed actions A= A_1; A_2; …; A_k, the intention recognition task is to identify a subset I’ of I, of the *most likely* intentions associated with A. As the sequence of observed actions grows the set of most likely intentions may change.

4 The Intention Recognition Task as a Clustering Problem

To apply clustering to intention recognition, first we use the information in the plan library to cluster actions that occur in plans. To achieve this we invent an appropriate similarity metric for actions. The metric is used to provide a pairwise similarity matrix. To this matrix we apply the standard Laplacian Eigenmap technique [2], which will allow us to visualise the resulting clusters and identify their prototypes (centroids). Thus we will obtain a membership matrix giving the likelihood of each intention given an observed action. Finally with each incoming observed action we use this membership matrix to compute the accumulated likelihood of each intention.

4.1 Similarity Calculation for Actions

Normal similarity metrics, such as Euclidean or Mahalanobis distances, are not suitable for intention recognition, since we do not have a coordinate system for actions. We propose a new similarity measure *W(i, j)* between two actions *i* and *j*, as follows:

$$\mathrm{W(i,j)} = \begin{cases} \mathrm{freq(i,j)}\dfrac{|P(i) \cap P(j)|}{|P(i) \cup P(j)|}, & i \neq j \\ 0 & , i = j \end{cases}$$

P(i) denotes the set of plans that include action *i*, and *freq(i, j)* denotes the maximum number of times the two actions *i* and *j* occur together in any plan. The term *freq(i, j)* acts as a weight, so that if a pair of actions occurs many times in a plan, their relationship (similarity) will be stronger. The term $|P(i) \cap P(j)| / |P(i) \cup P(j)|$ has the effect that a pair of actions is similar if they co-occur in a large number of plans, but not if either of them appears in many plans (if an action is present in many plans, it is considered to be an *untypical* action). An analogy could be the prominence of words such as “a” and “the” in the English language, and their lack of usefulness when it comes to identifying topics of a document.

4.2 Clustering Using Fuzzy C-Means (FCM)

The application of clustering to intention recognition produces clusters corresponding to the intentions in the plan library, one cluster for each intention. Figure 2 is based on the plan library of example 3.1 and uses the FCM algorithm. The top left and right clusters correspond to intentions 1 and 3, respectively, and the bottom cluster corresponds to intention 2. The cluster prototypes are denoted by hollow circles. The fuzziness allows action 1 to be all clusters, and actions 2 and 5 in two clusters.

The clustering algorithm also provides a membership matrix showing the probability of the membership of each action in each cluster (shown for action 11 in figure 2). Given a sequence of actions, we simply sum up the membership values of these actions for each intention. The intentions with the highest scores are the most likely intentions. Figure 3 shows how the scores of the intentions change as more actions are observed.

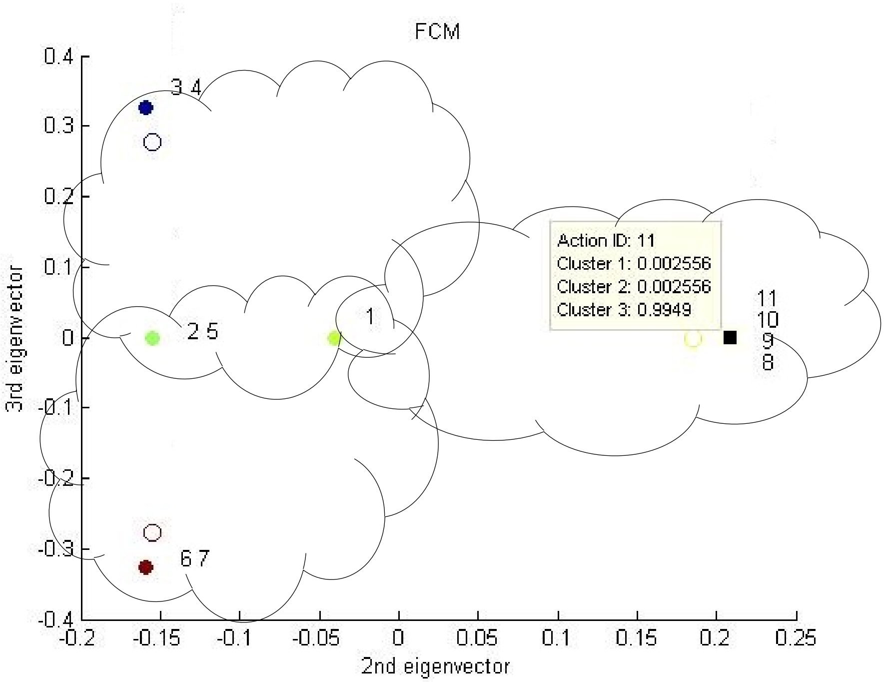


Fig. 2. Laplacian Eigenmap visualization using two eigenvectors

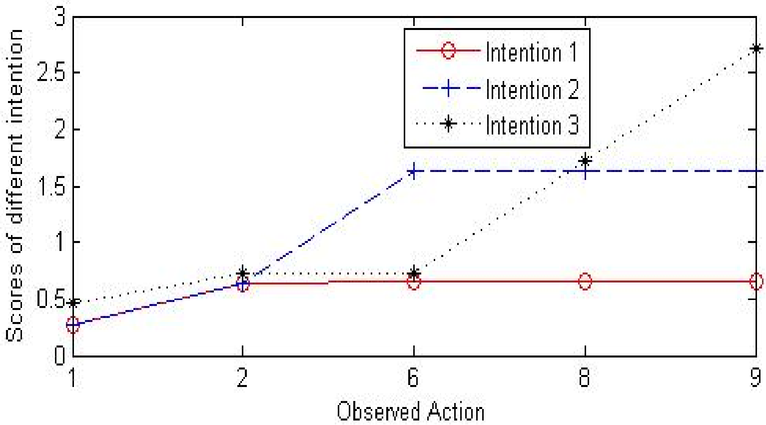


Fig. 3. Incremental intention recognition

5 Empirical Results and Conclusion

We have run several experiments. Table 1 summarises some of them. The OA_i represent action observations. OA7 is the simplest (1,2,3,4,5), complete and not interleaved. OA1 and OA2 are interleaved observations (OA1, for example, is 1, 2, 6, 3, 1, 2, 4, 7, 5, 5). OA3 and OA4 are partial (OA4, for example, is 9, 10, 11). OA5 and OA6 are partial and interleaved (OA5, for example, is 3, 4, 6, 7, 5, 5). OA8 and OA9 are partial, interleaved and noisy. In all cases the results are the expected intentions.

The technique seems promising, but more tests are needed with scalable and more realistic data sets. Future work also includes exploring other ways of computing similarity between actions, for example to take into account the order of actions.

Table 1. Some Experimental Results: Scores for each intention for 9 observation sets

Intention	OA1	OA2	OA3	OA4	OA5	OA6	OA7	OA8	OA9
I1	4.011	3.256	2.604	0.007	2.729	0.662	2.980	3.372	2.995
I2	4.012	1.307	0.655	0.007	2.729	2.612	1.031	3.373	2.021
I3	1.975	5.436	0.739	2.984	0.541	3.724	0.987	2.253	3.983
Likeliest	I1&I2	I1&I3	I1	I3	I1&I2	I2&I3	I1	I1&I2	I1&I3

References

1. Aarts, E.: Ambient intelligence: a multimedia perspective. IEEE Intelligent Systems 19(1), 12–19 (2004)
2. Belkin, M., Niyogi, P.: Laplacian Eigenmapsfor Dimensionality Reductionand Data Representation. Neural Computation 15, 1373–1396 (2002)
3. Cox, M.T., Kerkez, B.: Case-Based Plan Recognition with Novel Input. International Journal of Control and Intelligent Systems 34(2), 96–104 (2006)
4. Demolombe, R., Fernandez, A.M.O.: Intention Recognition in the Situation Calculus and Probability Theory Frameworks. In: Toni, F., Torroni, P. (eds.) CLIMA 2005. LNCS (LNAI), vol. 3900, pp. 358–372. Springer, Heidelberg (2006)
5. Dunn, J.: A fuzzy relative of the ISODATA process and its use in detecting compact well separated clusters. J. Cybern. 3(3), 32–57 (1974)
6. Geib, C.W., Goldman, R.P.: Partial observability and probabilistic plan/goal recognition. In: Proceedings of the 2005 International Workshop on Modeling Others from Observations (MOO 2005) (July 2005)
7. Hong, J.: 2001 Goal recognition through goal graph analysis. Journal of Artificial Intelligence Research 15, 1–30 (2001)
8. Mao, W., Gratch, J.: A utility-based approach to intention recognition. In: AAMAS Workshop on Agent Tracking: Modelling Other Agents from Observations (2004)
9. Pereira, L.M., Anh, H.T.: Elder care via intention recognition and evolution prospection. In: Abreu, S., Seipel, D. (eds.) Proc. 18th International Conference on Applications of Declarative Programming and Knowledge Management, Évora, Portugal (November 2009)
10. Sadri, F.: Logic-based Approaches to Intention Recognition. In: Chong, N.-Y., Mastrogiovanni, F. (eds.) Handbook of Research on Ambient Intelligence: Trends and Perspectives, pp. 346–375. IGI Global (2011)
11. Sadri, F.: Intention Recognition with Event Calculus Graphs and Weight of Evidence. In: Filipe, J. (ed.) Procs. of the 3rd International Conference on Agents and Artificial Intelligence. Springer (2011)
12. Sadri, F.: Intention Recognition in Agents for Ambient Intelligence: Logic-Based Approaches. In: Bosse (ed.) Agent-Based Approaches to Ambient Intelligence. IOS Press (to appear, 2012)

Behavior Modeling and Recognition Methods to Facilitate Transitions between Application-Specific Personalized Assistance Systems

Arun Ramakrishnan[1], Zubair Bhatti[1], Davy Preuveneers[1], Yolande Berbers[1], Aliaksei Andrushevich[2], Rolf Kistler[2], and Alexander Klapproth[2]

[1] IBBT-DistriNet, KU Leuven, Belgium
{firstname.lastname}@cs.kuleuven.be
[2] iHomeLab, Lucerne University of Applied Sciences and Arts, Switzerland
{firstname.lastname}@hslu.ch

Abstract. Activity recognition mandates complex sensor fusion processing. Many contributions in the literature focus on improving the recognition accuracy of a limited set of activities or the efficiency of the algorithms. However, there is little work on how to dynamically adapt the activity recognition techniques when evolving from one situation to the next. We present tool support to model transitions between activities, and a modular distributed framework of human activity recognition components with support for analyzing resource and recognition tradeoffs for different deployments and configurations.

Keywords: Behavior and activity recognition, smart home and health.

1 Introduction

The recognition of human activities based on various environment/user aware sensors and associated context information underpins the user-centric paradigm of Ambient Intelligence (AmI). A prerequisite to developing truly supportive and personalized systems is being able to recognize and anticipate typical human behavior and intent in a variety of different contexts [2]. However, the unpredictability of human behavior, the unanticipated circumstances of execution and a growing heterogeneity of future operational environments impose significant development challenges.

In this work, we focus on personalized assistance systems, and more particularly on transitioning between different activities. The recognition techniques might rely on the same sensors or algorithms (on different sensors), but both may need to be fine-tuned at runtime to the new context. Therefore, we need mechanisms to adapt recognition techniques from one scenario to the next. We address this challenge with (1) hybrid behavior modeling that links typical activities with different recognition techniques and (2) an adaptable and modular distributed framework that optimizes the processing and communication of the

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 385–390, 2012.

sensor data in a way that respects the resource constraints of the software and hardware components involved.

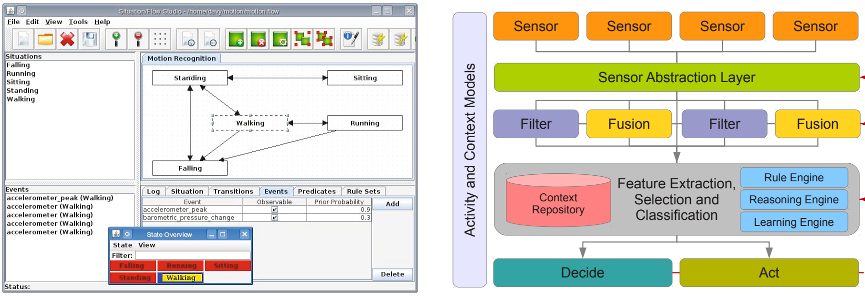
2 Related Work

A detailed review on the state-of-the-art on activity recognition [2,5] is beyond the scope of this paper. Ativity recognition under changing scenarios has been investigated in works on transfer learning [10] that investigate reusing the knowledge of learned activities in a new target space. However, our approach differs in that it specifically focuses on transitions between the activities themselves. As our experiments focus on resource trade-offs, we mainly highlight works that looked at device and system energy efficiency as a key concern. A first approach [14] looks at selecting the proper sensors. Accurate user activity prediction needs continuous sampling and the authors propose a method to select an optimal set of sensors at run-time. A similar approach was suggested in [6], arguing that certain sensors are more power consuming than others, with the authors favoring dynamically switching on certain high-cost sensors. Another approach is to adapt the sampling frequency. [12] demonstrates a non-linear relation between sampling frequency and energy consumption especially when frequency domain features (e.g. entropy) are being calculated. It also suggests activity dependent optimal sampling frequencies for mobile based accelerometer sensing, and adapting the window size as the authors found a linear relation between window size and energy consumption on mobile based activity recognition. One can also adapt the features being extracted and selected from the sensor data stream. [1] extensively studied the influence of selected features on classification accuracy and recall for wearable sensors (using 5 accelerometers) and concluded that sometimes fewer features can be more efficient without compensating classification accuracy. Communication efficiency can also significantly impact power consumption, as accelerometer based activity recognition requires high precision data and hence high sample frequencies (up to 100Hz). Techniques to optimize the communication and processing of large amounts of data for wearable and other wireless sensors was the research objective in [11,4]. Our work focuses on analyzing the trade-offs between all these concerns.

3 Smart Home and Health Motivating Use Cases

In our activity recognition experiments we compare two FP7 BUTLER [1] use cases taken from the smart home and smart health domain. A first activity we aim to recognize is taken from Ambient Assisted Living scenarios, i.e. *detecting a fall* [7,13,3]. The first technique only uses a triaxial accelerometer running at 50Hz and looking for patterns of interest through continuous feature extraction and selection using a high pass FIR filter to take out the gravity component as a fall is characterized by a dynamic acceleration in a small time window (usually less than a second). Our second approach combines the accelerometer with a barometric pressure sensor. A sudden acceleration triggers a sampling of the

[1] http://www.iot-butler.eu



(a) Situation Studio (b) Human activity recognition architecture

Fig. 1. Activity transition modeling tool (a) and architectural overview (b)

pressure sensor to detect the current altitude. After a predefined or automatically calculated time delay, another sample is taken. If a significant difference in height is detected, we assume the person has fallen and was not able to get up. This technique does not require complex feature extraction from the accelerometer and pressure sensors.

We use the same accelerometer in a personalized diabetes assistant [8] to *track the exercise level of physical activities* and offer decision support on medicine intake based on past occurrences. We analyze the sensor data stream for activities like walking and running to estimate the calories consumed. This requires feature extraction and selection in both the time and the frequency domain. Other differences with the previous use case are that data should be stored for future reference and comparison, the sampling rate is lower, and the sliding window for signal analysis is several seconds long for better accuracy.

4 A Multiple Applications Approach in Behavior Modeling and Activity Recognition

Each of the feature extraction and analysis building blocks are developed as separate components that can be deployed on a sensor mote or mobile. In general, the overall distributed architecture is depicted in Figure 1(b). The set of sensors, filters, aggregators, classification and learning components can be deployed, composed and configured dynamically at runtime, depending on the activities to be recognized and corresponding resource trade-offs for wireless communication, computation and memory.

4.1 Modeling and Use of Contextual Domain Knowledge

Many works focus on a limited set of activities and validate the accuracy of their approach with the implicit assumption that the activity of interest is taking place. One hardly finds numbers about false positives or false negatives. In our approach, we instead use the cognitive loop in our architecture (see Figure 1(b)) to infer the most probable activities given the current time and location and

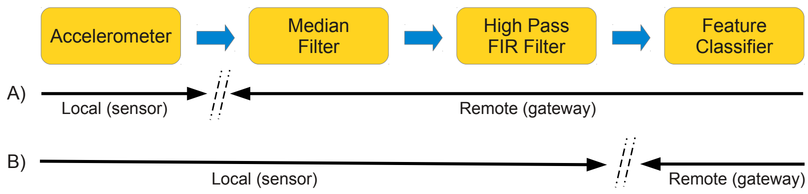


Fig. 2. Two deployment scenarios for feature extraction and selection components

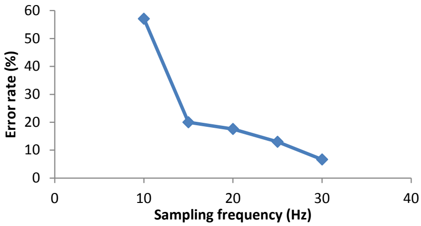
initiate the corresponding activity recognition techniques. We explicitly consider situations where techniques could lead to false positives, etc. For example, the fall detection with the barometric pressure might detect a false positive when going down the stairs, because with each step the accelerometer triggers the pressure sensor and the latter detects a lower altitude. However, one can also fall down the stairs. All of these interrelationships between different kinds of contexts and activities and corresponding recognition techniques are modeled with our Situation Studio [9]. This tool (see Figure 1(a)) borrows concepts from workflow modeling languages, and represents situations that evolve from one to the next through constrained sequential and parallel transitions. For each of them, we identity the contextual boundaries, the likelihood of activities of interest, the relevant contextual events, and the recognition schemes available.

4.2 Trade-Offs with Explicit and Implicit Interaction

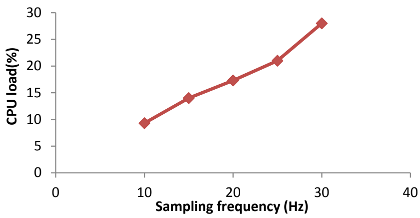
Recognizing activities of daily living can be based on data acquired through explicit or implicit interaction with the user. The decision on which approach to pursue is based on the classification and recognition accuracy of the corresponding technique, and on the resource constraints of the feature extraction and selection components for an optimal deployment. For example, sampling at 100Hz on a triaxial accelerometer and transmitting the raw data to a gateway base station for further processing will incur minimal computational overhead, but will be very expensive from a communication point of view (about 10MB per hour). By carrying out some of the data processing and analysis on the sensor, the amount of communication will be reduced. See the two deployment scenarios in Figure 2. Obviously, from a power consumption perspective there are various trade-offs to be investigated for an optimal deployment.

5 Experimental Evaluation

We implemented two use cases: (1) fall detection and (2) step counting, and designed the activity recognition building blocks using a modular component based approach to simplify their distributed deployment. Each of these components has been profiled on various platforms. In our experiments, we used the SunSPOT sensor and an HTC Android mobile phone for profiling. We analyzed for each component on each platform the trade-offs of sampling frequency against:



(a) Error trade-off



(b) Performance trade-off

Fig. 3. Sampling rate vs. recognition (a) and performance (b) on the SunSPOT

1. Recognition accuracy
2. Computational complexity
3. Communication overhead and latency
4. Power consumption
5. Size and memory consumption

Similar analyses were carried out for trade-offs against the size of the sliding window, etc, but due to space considerations, we only provide in Figure 3 the results of the first two trade-off analyses for the step-counting components (features and feature classifiers) all running on the SunSPOT sensor. The figure shows both trade-offs of interest, i.e. (1) *recognition rate*, to compare different algorithms and configurations (e.g. size of sliding window and use of certain filters), and (2) *performance impact*, to decide which components to deploy and on which platform (computation vs. communication trade-off). Other trade-offs – also not shown here – investigate scenarios with all the processing done on a gateway and intermediate deployments to compare the network overhead and power consumption vs. the sampling frequency. These kind of trade-offs help us to find Pareto optimal deployments and configurations for activity recognition.

6 Discussion and Future Work

In our work, we are not necessarily aiming to improve the recognition rate for certain kinds of behavior and activities with complex algorithms. Rather, we are interested in finding the trade-offs between different human activity recognition components for feature selection, extraction and classification and (1) their recognition rate and (2) their resource impact for distributed deployments. We briefly discussed our *Situation Studio* tool support to model activity transitions and contextual background allowing us link that with possible recognition techniques. The techniques are implemented as modular software building blocks which can be dynamically configured, composed and deployed on our component based middleware platform that runs on sensors, smartphones and backend systems. The effects of deploying these components are profiled on each of these platforms, which helps us to find trade-offs for a distributed deployment of these component considering both recognition accuracy as well as the performance impact. As future work, we will investigate metrics for analyzing the influence of contextual background knowledge, the non-intrusiveness with explicit vs. implicit interaction.

References

1. Atallah, L., Lo, B., King, R., Yang, G.-Z.: Sensor positioning for activity recognition using wearable accelerometers. IEEE Transactions on Biomedical Circuits and Systems 5(4), 320–329 (2011)
2. Aztiria, A., Izaguirre, A., Augusto, J.: Learning Patterns in Ambient Intelligence Environments: A Survey. Artificial Intelligence Review 34(1), 35–51 (2010)
3. Kangas, M., Konttila, A., Lindgren, P., Winblad, I., Jamsa, T.: Comparison of low-complexity fall detection algorithms for body attached accelerometers. Gait & Posture 28(2), 285–291 (2008)
4. Keally, M., Zhou, G., Xing, G., Wu, J., Pyles, A.: Pbn: towards practical activity recognition using smartphone-based body sensor networks. In: Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems, SenSys 2011, pp. 246–259. ACM, New York (2011)
5. Kim, E., Helal, S., Cook, D.: Human activity recognition and pattern discovery. IEEE Pervasive Computing 9(1), 48–53 (2010)
6. Miyaki, T., Czerny, J., Gordon, D., Beigl, M.: Energy-efficient activity recognition using prediction. In: 2012 16th International Symposium on Wearable Computers, pp. 29–36 (2012)
7. Noury, N., Fleury, A., Rumeau, P., Bourke, A.K., Laighin, G.O., Rialle, V., Lundy, J.E.: Fall detection - principles and methods. In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2007, pp. 1663–1666 (August 2007)
8. Preuveneers, D., Berbers, Y.: Mobile phones assisting with health self-care: a diabetes case study. In: Mobile HCI, pp. 177–186 (2008)
9. Preuveneers, D., Landmark, A.D., Wienhofen, L.W.M.: Probabilistic event processing for situational awareness. In: 12th International Conference on Innovative Internet Community Systems (I2CS 2012). LNI, vol. P-204, pp. 96–107 (June 2012)
10. Rashidi, P., Cook, D.: Activity knowledge transfer in smart environments. Pervasive Mob. Comput. 7(3), 331–343 (2011)
11. Wang, L., Gu, T., Chen, H., Tao, X., Lu, J.: Real-time activity recognition in wireless body sensor networks: From simple gestures to complex activities. In: 2010 IEEE 16th International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA), pp. 43–52 (August 2010)
12. Yan, Z., Subbaraju, V., Chakraborty, D., Misra, A., Aberer, K.: Energy-Efficient Continuous Activity Recognition on Mobile Phones: An Activity-Adaptive Approach. In: 16th International Symposium on Wearable Computers (ISWC) (2012)
13. Yu, X.: Approaches and principles of fall detection for elderly and patient. In: 10th International Conference on e-health Networking, Applications and Services, HealthCom 2008, pp. 42–47 (July 2008)
14. Zappi, P., Lombriser, C., Stiefmeier, T., Farella, E., Roggen, D., Benini, L., Tröster, G.: Activity Recognition from On-Body Sensors: Accuracy-Power Trade-Off by Dynamic Sensor Selection. In: Verdone, R. (ed.) EWSN 2008. LNCS, vol. 4913, pp. 17–33. Springer, Heidelberg (2008)

LumaFluid: A Responsive Environment to Stimulate Social Interaction in Public Spaces

Gianluca Monaci[1], Tommaso Gritti[1], Martine van Beers[1], Ad Vermeulen[1], Bram Nab[2], Inge Thomassen[2], Marigo Heijboer[2], Sandra Suijkerbuijk[2], Wouter Walmink[3], and Maarten Hendriks[4]

[1] Philips Research - Eindhoven, The Netherlands
[2] Department of Industrial Engineering & Innovation Sciences Eindhoven University of Technology, The Netherlands
[3] studio:ludens - Eindhoven, The Netherlands
[4] Little Mountain - Eindhoven, The Netherlands

Abstract. *LumaFluid* is an interactive environment that explores new ways to stimulate emotional and social engagement through light. A vision system localizes people present in the *LumaFluid* square. Colored spotlights highlight each person and connections are drawn between neighboring visitors to underline and stimulate interpersonal communication. Two versions of the concept where deployed during the 2011 STRP Festival. In this paper we describe the conception and realization of the installation, and we discuss the insights collected during the event.

1 Introduction

In large cities, people socialize in squares in open air at evening. *LumaFluid* is an interactive installation that explores ways to stimulate social interactivity between people in outdoor public spaces using light. The principal intents of the installation are twofold: first, we want to attract the attention of the visitors with a colorful and mysterious environment. Then, we want to explore ways to encourage the social interaction of the visitors through light.

Lately artists have proposed a growing number of interactive augmented reality (AR) installations, following the progress and accessibility of technologies [1,2]. Many AR installations involve the combination of light and video projections with the physical space [3,4,5,6]. The design of interaction modalities between passersby or visitors and AR installations is a crucial and relatively novel aspect, in particular in public, urban installations. Through interactivity, the audience becomes an active component of the installation, influencing the course of the events. In most of the above mentioned works [4,5,6], the environment is responsive to visitors' presence, but the response lies in the environment itself. Here, in the path of Snibbe's work [3], we focus on the emotional and social engagement of people, using space and light as means to stimulate interaction.

LumaFluid was installed at the 2011 STRP Festival[1], one of the largest art, music and technology festivals in Europe, with 31.000 visitors in 2011. During

[1] www.strp.nl

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 391–396, 2012.

Fig. 1. *Version 1* (left) and *version 2* (right) of *LumaFluid* at STRP festival

the ten days of festival, two versions of the installation were run. The first version, in Figure 1 (left) focused on having fun together (emotional engagement). Vibrating light particles in shades of green fill the interactive floor. As soon as visitors enter the floor, they attract particles, forming large colored spotlights. Streams of particles appear between people when they move closer together, underlining the possibility of social interaction. When standing very close, the two spotlights morph together and start pulsating in response. The second version, in Figure 1 (right), was intended to have a more explicit connecting element (social engagement). Visitors on the floor are highlighted by a colored spotlight, but in contrast with version 1, the background is completely black. Like the previous version, a visual effect links neighboring visitors: here a line that morphs from one spotlight's color into the other. Through this mechanism, visitors can create colorful patterns which change continuously as people move, join and leave. In both versions, the idea is to use interactive light effects as a stimulus (or a pretext) for people to connect to others in a playful way. We observed and interviewed many of the visitors of the installation during STRP 2011. For a lively impression of *LumaFluid*, please check our video at `www.vimeo.com/34655968`.

2 Installation Setup

The *LumaFluid* installation at STRP is shown in Figure 2. A grayscale camera was installed over an area of 7m×7m at about 4m height, looking downwards (in a red square in the figure). The camera has a wide angle lens to capture the whole installation area and it is mainly sensitive to infrared (IR) radiation. IR illumination was provided by four IR LED light sources mounted at the corners of the area (highlighted by yellow boxes in Figure 2). In this way the camera can capture IR images of the square while filtering out illumination changes produced by the installation, as well as light coming from neighboring luminaires.

The captured images are analyzed by a computer vision software that, using a background subtraction method, localizes the visitors present in the installation area. Their positions are communicated to a rendering software that creates the output visuals. The images from the rendering software are split into nine images that are projected back on the scene by a grid of 3×3 video projectors (highlighted with white circles in Figure 2).

Fig. 2. Setup of *LumaFluid*. A camera (red square) senses the space and provides location information to the rendering software, which projects the graphics back using nine video projectors (white circles). IR LED light sources are highlighted in yellow.

3 Interactions and Impressions

During the first five days of the festival, version 1 of *LumaFluid* was used, while in the last five days version 2 was adopted. We observed the experiences of people who approached and used the two versions of *LumaFluid* and we collected interaction statistics by carrying out three type of observations. Firstly, we observed the installation as a whole 127 times (65 times in version 1 and 62 in version 2), for ten minutes, at regular intervals. Secondly, we analyzed the individual behavior of 199 people from the moment they came close to the installation until the time the exited it: 83 (42 male, 41 female) in version 1 and 116 (69 male, 47 female) in version 2. Finally, we carried out a structured interview with 104 visitors. We interviewed 56 visitors (23 male, 33 female) for version 1 and 48 people (26 male, 22 female) for version 2.

While we managed to keep a certain balance in the sample in terms of number and gender for the two versions, we have to underline that most festival's visitors were Dutch, and many with a background in design, technology or art. Another bias factor lies in the age distribution of the samples. The festival attracts mostly a young audience, but attendance was very heterogeneous across week days. Version 1 was mostly attended by adults in the age range 30-65, while version 2 was mostly visited by young adults between 18 and 30. Notwithstanding these limitations, we believe the collected data was from a sufficiently representative group of attendees, providing insights relevant for the analysis of the installation.

For both versions, we measured that around 80% of the people visiting *LumaFluid* entered the installation area. The concept of having a spotlight

Fig. 3. (Left) Version 1 of *LumaFluid*: an artist using *LumaFluid* for her performance. (Right) Light patterns created with version 2 of *LumaFluid*.

focused on them was generally appreciated by the public and many visitors were surprised by the fact that the spotlight could follow them. One common reaction was to run or zig-zag, and then check if the spotlight was still following. Half of the visitors for version 1 and 40% for version 2 focused on the interactive floor while approaching the installation: in both versions, the floor was the main eye catcher. 36% of the visitors of version 1 and 16% of visitors in version 2 continued to look only at the floor even while inside the installation square. The fact that the participants' focus was on the floor seems to indicate the effectiveness of the visualization, although the main goal of the installation was to stimulate social interaction. However, *LumaFluid* does invoke interaction: 48% of visitors in version 1 and 52% of visitors of version 2 interacted with other persons. From the interviews, it emerged that not all visitors could understand the system's functionalities and its intent. Because of that, in a number of cases people left the interaction square after few seconds. Interestingly, we noticed that people also tried to interpret and attribute a meaning to aspects of the installation that were randomly set, such as the colors of the visualization.

The two versions of the installation had their own character and specificities. In version 1, the first thing we noticed was that people, mostly children, came up with different games to play, using the installation as a tool to create a gameplay. Most children seemed to enjoy themselves when playing their own fictional games using *LumaFluid*. A performing artist used the floor as additional attribute during her performances (Figure 3 (left)). This seemed to have a positive influence on how visitors perceived the performance: many more visitors stayed to watch the performance when the artist used *LumaFluid* than when she did not. Visitors explored what version 1 of the installation could do, for example by walking away or dancing. This however lead people to pay more attention to the installation itself, rather than to other people on the floor. Besides, half of the people indicated they did not understand that interaction with other persons was possible, often because it was not visible to them, or because nobody else was present at the floor.

Concerning version 2, because the visualization was much simpler, people understood the concept of playing together better: visitors created figures and patterns together, as in Figure 3 (right). To create these visuals, people looked at and talked to each other. Where the first version provoked an active, more individual and game-like behavior, the second stimulated a more social and co-creating behavior. People also acted in way that were not anticipated by the authors. Some were seen lying down on the floor, others used objects like a backpack as part of their interaction (Figure 3 (right)). In the interviews, more people reported to find version 2 fun and exciting, although this version did require the presence of other people to remain interesting. Opinions were divided as to whether this type of visualization would lead to actual conversations, or whether it would remain restricted to shallow interaction such as smiling.

When comparing the two versions, we found out that version 1 is more interesting to visit and play with, also when someone is alone, whereas version 2 is only interesting when multiple people are present on the floor. People entering version 1 of *LumaFluid* seemed to understand the interaction possibility intended by the authors only in 14% of the cases. People entering version 2 seemed to understand this in 53% of the cases. However, for both versions about half of the visitors engage in interactions with others. Interestingly, it is not necessary to understand the possibility of interacting to actually start an interaction. In version 1 only 48% of all the interactions involved verbal contact, while in version 2, 83% of interactions involved verbal communication. The difference becomes bigger if we focus on the contact with strangers. In version 1 only a quarter of the people that had contact with strangers had verbal contact. In version 2 this figure raises to over three quarters: people who used version 2 had more direct contact with others, as expressed by verbal communication. This indicates that we are dealing here with two very different types of interaction. Observing the arousal states of the visitors, we noticed that version 1 elicited a more active behavior than version 2. The arousal level of people entering and interacting with version 1 was higher than for version 2. Besides, it was reported that version 1 was perceived to give more creative freedom to visitors. On the other hand, version 2 had a clearer concept, that was easier to grasp for most participants and that stimulated a more direct communication. There is an interesting tension between a fascinating concept, but difficult to understand (version 1) versus a clear concept, but with a limited creative freedom (version 2).

4 Discussion

One of the most important insight about the installation is the importance of both the aesthetic and functional components to keep the experience interesting and meaningful for the visitors. In future concepts we will investigate ways to merge the poetic elements of the visualization in version 1 with the clear links between visitors in version 2 to create meaningful and attractive interactive experiences in the public spaces.

In the interviews, several users indicated that adding an extra dimension would greatly increase the time the system would remain interesting. This extra dimension could be obtained by projecting the visualization also on walls, or adding sound effects that respond to the activities on the installation floor.

Adding another dimension would not only enrich the experience of *LumaFluid*, but it could also solve the current issue of visitors mainly focusing on the floor. While one group of people stated that the present interactive structure would lead to conversations with strangers, others thought it would only lead to superficial interaction (e.g. smiling) and suggested that more triggers besides light would be needed to evoke more interaction. More research is required to investigate how light, graphics and other modalities, such as sound, can enrich and stimulate interactions in different public environments.

Besides using the concept in public places such as squares or large halls, the observations drawn in this paper and the future developments of this work can contribute to applications such as theatrical or dance performances, as was done by one of the artists at STRP. During the interviews, participants repeatedly mentioned two settings where they think a system like *LumaFluid* would be interesting: playgrounds for children and nightclubs for adults. These two ideas were mentioned independently of the version used at STRP festival. These are two interesting directions to be considered in future instances of *LumaFluid*.

5 Conclusions

In this paper we described *LumaFluid*, an interactive installation that investigates new ways to stimulate social interaction in public spaces through light. The installation consisted of a responsive space where people are localized on near-IR images captured by a camera hanging on the ceiling. A rendering software projects light effects over the people present on the floor, highlighting and connecting them. Two versions of *LumaFluid* were deployed during ten days at 2011 STRP festival. By analyzing the behavior of hundreds of participants with the two versions, we concluded that few elements seem to be essential for an installation to stimulate social interaction: firstly, the lighting effects should be capable of grabbing the attention of visitors, to lure them in the installation area. At the same time, the initial effects should be intuitive enough for a person to understand the interactions he or she is having with the environment. To ensure that participation is sustained in time, gradually more sophisticated effects could be introduced. With regards to the social interaction, a similar incremental strategy could be used. This could also stimulate verbal interaction, as the more experienced participants could guide newly arriving people.

References

1. Grau, O.: Virtual Art: From Illusion to Immersion. The MIT Press (2003)
2. Bimber, O., Raskar, R.: Spatial augmented reality: merging real and virtual worlds. A K Peters, Ltd. (2005)
3. Snibbe, Scott: Boundary functions (1998), http://www.snibbe.com/projects/interactive/boundaryfunctions
4. Levin, G.: Collaborators (Motion traces), http://www.flong.com/projects/a1
5. Utterback, C.: (Projects), www.camilleutterback.com/projects
6. Gritti, T., Monaci, G., Vignoli, F., Walmink, W., Hendriks, M.: Flower power. In: Proceedings of ACM Multimedia (2011)

A Cost-Based Model for Service Discovery in Smart Environments

Michele Girolami[1,2], Francesco Furfari[1], and Stefano Chessa[1,2]

[1] Istituto di Scienza e Tecnologie dell'Informazione, via Moruzzi 1, Pisa, Italy
{michele.girolami,francesco.furfari}@isti.cnr.it
[2] Dipartimento di Informatica – Univ. Di Pisa, Largo Pontecorvo 3, Pisa, Italy
{girolami,ste}@di.unipi.it

Abstract. This paper describes the CoDA algorithm for the service discovery in the Smart Environments. CoDA is based on the energy distance metric that allows the clients to select the service providers whose access cost is the lowest in terms of the energy requirements.

Keywords: service discovery, Smart Environments, energy efficiency.

1 Introduction

This work proposes a novel algorithm for the service discovery [1, 2, 3] based on energy costs, in order to minimize the cost of accessing the services. We introduced the *energy distance* as the driving metric for the selection of the most energy efficient service, which is defined as the overall amount of energy consumed by all the devices involved in the service access. To this purpose we first propose a model for the service discovery in the Smart Environments [4](SE), then we propose the Cost-Based Discovery Algorithm (CoDA), which implements a strategy for the service selection minimizing the energy distance. CoDA has been evaluated in an experimental SE and compared with a different algorithm.

2 Cost-Based Model for Service Discovery

A SE can be defined as the tuple $< G, e, \mathcal{S}, \mu >$, where e is the function measuring energy distances among nodes, $\mathcal{S}$ is the set of available services, μ is the service discovery function. Graph $G =< N, A >$ represents the SE as a weighted un-direct graph, where $N = \{n_1, \ldots, n_m\}$ is the set of nodes of the SE and $A = \{a_{n,m}\}$ is the adjacency matrix that collects the edges of G. A node $n_i \in N$ can be configured as service provider sp_i if it provides one or more services, as a service client c_i if it accesses to one or more services, or as directory agent d_i if it stores a subset of the services available in the SE. A node can also be configured to play multiple roles at the same time. Edges in G are weighted by the function $w: A \to \mathbb{R}$ that describes the cost of traversing an edge. The weight of traversing an edge is described by the following

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 397–402, 2012.

relationship: $w(a_{n,m}) = l_{n,m} \cdot k$ where $l_{n,m}$ is the unit-cost of the edge from n to m expressed by using a kind of metric like, for example the energy consumption in $Watt$ of the network interface. k is the traffic sent along the link expressed by using a kind of metric like, for example the time spent in order to send an amount of data along the edge. Our core metric relies on the energy distance, defined as follows:

Definition 1. *Given a client* c_i *and a service provider* sp_j the *energy distance is:*

$$e(c_i, sp_j) = \sum_{a_{n,m} \in p(a_{i,j})} w(a_{n,m}) \tag{1}$$

where $p(a_{i,j})$ *is the path from* c_i *to* sp_j.

The energy distance measures the overall weight of the path from the client to the provider. Such a metric is exploited by the clients in order to select the providers with the most promising path in terms of the energy required for the service access.
Every service provides functionalities with a predictable performance, the function $p: \mathcal{S} \rightarrow \mathbb{R}$ associates to every service an indication of the expected performance during the service access.

Definition 2. *Given a service* $\mathcal{s}_i$ *with* $p(\mathcal{s}_i) = v$, *the function* $r: \mathcal{S} \rightarrow \mathbb{N}$ *provides the number of clients that can be managed with the same performance value. Function r is called service residual capacity.*

We model the service discovery function based on the following operations:

- *Lookup:* it is the function $l: N \rightarrow D$ that, for a given node of the SE, finds a directory agent with whom interact.
- *Registration:* it is the operation denoted by $sp_j \xrightarrow{adv_k} d_i$ and executed by the service provider sp_j to announce the availably of a service to the directory agent d_i. The service $\mathcal{s}_k$ is described by the tuple $adv_k = <DESCR, SP, R>$ (called advertisement) that comprises the description of the service $DESCR$, the service provider SP and the residual capacity R.
- *Query*: is the operation $c_i \xrightarrow{q_k} d_j$ executed by a client c_i to query the directory agent d_j by sending the query $q_k =< DESCR >$.

Definition 3 formally introduces the service discovery in the SE.

Definition 3. *The service discovery function* $\mu: Q \rightarrow ADV$ *associates to a query* $q_k \in Q$ *a set of the service advertisements* $adv_k \in ADV$. *ADV are the candidate services for the query* q_k.

After the execution of the service discovery function μ a client c_i needs to select the most suitable advertisement according to an objective function. To this purpose c_i exploits an objective function that minimizes the energy distance between the client and the service provider and, at the same time, selects the service provider with an admissible residual capacity. Specifically, given the set of candidate advertisements $ADV = \{adv_1, \dots, adv_n\}$ for the query q_k submitted by the client c_i, the problem of selection of the service advertisement can be described by:

$$\begin{cases} \min(e(c_i, adv_j.SP)), \forall j \in ADV\} \\ r(s_j) > 0, \quad \forall j \in ADV \end{cases} \tag{2}$$

that minimizes the energy distance while keeping only the services with an admissible residual capacity.

3 Cost-Based Algorithm for Service Discovery

The subroutines here described implement the model introduced in Section 2. The energy distance between an arbitrary pair of nodes is computed as follows:

1. it finds the path from the client to the service provider providing the service. We use the well-known application-layer tool Traceroute;
2. it gathers the type of the network interfaces used by the intermediated nodes in order to forward the messages to the service provider;
3. it compute the energy distance according to Definition 1 as the sum $e(c_i, sp_j) = \sum_{a_{n,m} \in p(a_{i,j})} w(a_{n,m})$, where $l_{n,m}$is the unit cost of sending *k*=1 unit of data along the edge $a_{n,m}$. The unit cost is obtained by associating to the type of the network interface with the power requirement in Watt for delivering some packets.

The energy distance exploits Traceroute to find the path from c_i to each service provider sp_j (step 1) and, for every hop found, it inspects the type of the interface used to forward the message to the next hop (step 2). This information can be obtained by exploiting the SNMP [5] protocol. It defines how to inspect a wide variety of *monitored objects*, while the database of monitored objects, supported by each router (also called MIB), is defined by a number of RFCs. One of the most diffused MIB is the RFC-1213 "Management Information Base for Network Management of TCP/IP-based internets: MIB-II"[1]. It provides the name for a huge number of monitored objects. We use the following SNMP objects for obtaining information about the type of interfaces used among the routers:

– `ipRouteTable`: the routing table will be inspected in order to find the entry associated to the next hop. It contains the index of the network interface used to forward the message to the destination of the entry, named `ipRouteIfIndex`.
– `ifTable:` the network interface table will be inspected in order to find information about the interface used to forward the message along the path. By accessing the `ifTable` at index `ipRouteIfIndex` (see the previous point), it is possible to obtain a number of information about the interface. We are interested in `ifType` that identifies the type of the interface. The list of available types is standardized by the IANA authority list.

Table 3 reports the algorithm. It invokes the Traceroute and for each router *i* in the `path` data structure (denoted `path[i]`), it gathers the following information:

[1] McCloghrie K., Rose M., "Management Information Base for Network Management of TCP/IP-based internets: MIB-II", IETF RFC1213, March 1991.

- from ipRouteTable it extracts the entry matching with the next hop, which is in path[i+1]. To this purpose, the function find(ipRouteTable, path[i+1]) finds the right entry in the routing table matching with the IP address contained in the array path[i+1] (next hop);
- given the ipRouteEntry, it requests to the router *i* in the path the entry of the ifTable that the router uses to forward the message to the next hop;
- given the interface type (ifType), the subroutine evaluates the cost of the next hop by invoking the function compute_cost(c_{link}). Such a function associates the interface type with the unit cost of the interface (i.e. the energy consumption in Watt of the network interface) and sums up the value to the e_distance variable.

Table 1. The Energy Distance pseudo-code

```
function cost e(c_i, sp_j) begin
path = traceroute(i,j);
for i = 0 to path.size-1 begin
  SNMP-MESSAGE.OBEJCT = ipRouteTable;
  Send SNMP-MESSAGE to path[i];
  answer = Receive the response (MaxTime);
  if (answer.PDU == response) then
    ipRouteTable = answer.value;
    ipRouteEntry = find(ipRouteTable, path[i+1]);
    SNMP-MESSAGE.OBJECT = ipRouteTable.ipRouteIfIndex;
    Send SNMP-MESSAGE to path[i];
    answer = Receive the response (MaxTime);
    if (answer.PDU == response)then
      ifEntry = answer.value
      c_link = ifEntry.ifType;
      e_distance += compute_cost(c_link);
  end
end return e_distance; end
```

In order for a client to discover a service, it submits a query to the nearest directory agent (function lookupDirectory() used in the pseudo-code in Table 2). The client gathers all the advertisements matching with the query and it selects the advertisement that minimizes the energy distance with an admissible residual capacity. The Query message is composed by the field DESCRIPTION (used to select the service advertisements matching with the query). The response to Query is an array of advertisements. Each advertisement contains the fields RESIDUAL_CAPACITY, DESCRIPTION, SERVICE_PROVIDER.

Table 2. The service discovery pseudo-code

```
function srv_advertisement service_discovery(q_k) begin
  c_i = myURL;
  directory_agent = lookupDirectory();
  Query.DESCR = q_k.DESCR;
  Send Message to directory_agent;
  ADV[] advertisements = Receive the response (MaxTime);
  for j ← 0 to advertisements.length -1 do
    if(advertisements[j].RESIDUAL_CAPACITY > 0) then
        costs[j] = e(c_i,sp_j); end
  end
  srv_advertisement = min(costs);
  return srv_advertisement; end
```

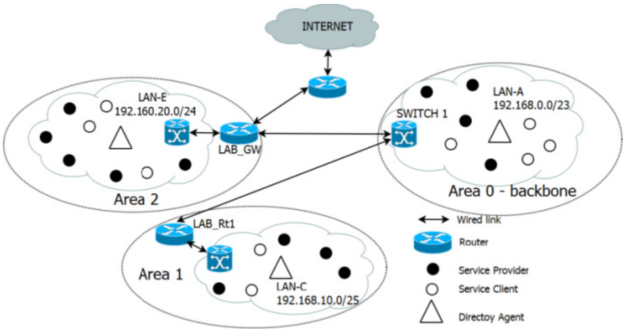


Fig. 1. Network topology of the Smart Environment

4 Experiments and Results

CoDA has been experimented in a real network composed of 3 LANs interconnected by 4 routers as reported in figure 1. The service selection is computed by evaluating the energy distance between a sub-set of clients and the providers placed in the LANs. The energy distance is evaluated according to Definition 1 where: *(i)* the weight function is $w(a_{n,m}) = l_{n,m}$ where $l_{n,m}$ is the power requirements in Watt (W) of the NIC used in order to forward *k*=1 unit of traffic along the edge $a_{n,m}$ (we used Intel® 82579 Gigabit Ethernet and M5 Juniper® Gig.Ethernet PIC) and *(ii)* the path from the client to the provider is computed by running Traceroute. The service selection previously described has been compared with a different heuristic (called SPT, Shortest Processing Time) where the selection of the services is made according to the shortest path from the client to the service provider. Table 3 compares the selection performed by CoDA with respect to SPT. Column *advertisement* reports the service requested by a client, the column *provider* reports the selection performed by CoDA and SPT, the column *data* reports the amount of data sent in order to interact with the service and the column *e(x,y)* reports the energy distance computed for the selection of

CoDA and SPT. The service providers selected by CoDA allow saving a non-negligible amount of energy for the service access. More precisely CoDA saves around 26% and up to 36.47% of energy spent by the routers for one single access performed by one client.

Table 3. Energy distance in Joule (J) for CoDA and SPT

Advertisement	*CoDA*			*SPT*		
	provider	*data*	*e(x,y)* J	*provider*	*Data*	*e(x,y)* J
adv_8	sp_{10}	10MB	3.18 J	sp_{13}	10MB	3.772 J
adv_1	sp_{12}	5MB	0.71 J	sp_1	5MB	1.59 J
adv_3	sp_6	400KB	0.103 J	sp_6	400KB	0.103 J

5 Conclusions

The key contribution of this paper is a service discovery model in the Smart Environments with the objective to optimize the access cost for the services. A first improvement is to elect, in an efficient way, a number of nodes acting as directory agents instead of assuming a pre-existing deployment. This will enable CoDA to resize the dimension of the directory according to the number of clients and providers present in a SE. Another improvement is to adopt fine-grained models for the user mobility in order to reproduce a more accurate evaluation. Future works also include a deeper analysis of the tradeoff between energy consumption and latency and delay and the application of the proposed methodology in gateways for sensor networks [6].

References

1. Ververidis, C.N., Polyzos, G.C.: Service discovery for mobile Ad Hoc networks: a survey of issues and techniques. IEEE Communications Surveys & Tutorials 10(3), 30–45 (2008)
2. Chakraborty, D., et al.: Toward Distributed Service Discovery in Pervasive Computing Environments. IEEE Trans. Mobile Computing 5(2), 97–112 (2006)
3. Rasch, K., et al.: Context-driven personalized service discovery in pervasive environments. World Wide Web 14(4), 295–319 (2012)
4. Weiser, M.: Hot topics-ubiquitous computing. Computer 26(10), 71–72 (1993)
5. Case, J., Fedor, M., Schoffstall, M., Davin, J.: A Simple Network Management Protocol (SNMP). IETF RFC 1157 (May 1990)
6. Chessa, S., Furfari, F., Girolami, M., Lenzi, S.: SAIL: a Sensor Abstraction and Integration Layer for Context Aware Architectures. In: 34th EUROMICRO - Special Session on Software Architecture for Pervasive Systems (SAPS), Parma, Italy, pp. 374–381 (2008)

On the Use of Video Prototyping in Designing Ambient User Experiences

Nikolaos Batalas, Hester Bruikman, Annemiek Van Drunen, He Huang, Dominika Turzynska, Vanessa Vakili, Natalia Voynarovskaya, and Panos Markopoulos

Department of Industrial Design, Eindhoven University of Technology, Den Dolech 2, 5612 AZ, Eindhoven, The Netherlands
{n.batalas,p.markopoulos}@tue.nl

Abstract. We discuss a case study where this technique was used in the design of an ambient intelligence system, highlighting how it impacted the design process both in positive and negative ways. This contextualized account complements related comparative studies that have been conducted outside the context of a design project, and have focused on methodological aspects of video prototyping. We conclude that designers need to be aware of how video as a persuasive medium obfuscates implementation and usability issues, and video prototype production should communicate explicitly the scope of the design issues that it addresses and those it does not.

1 Introduction

Video prototyping is a well established technique in the field of interaction design. It makes use of a range of simple or more complex techniques like stop motion, animations, video editing, narrative voice overs, etc., to bring to life a scenario of use that illustrates key aspects of an interaction design. The method was originally favoured for its ability to provide a dynamic and low cost representation of an interaction. Specifically for the domain of ambient intelligence, video prototypes are particularly appropriate as they make it easy to visualize the particular interaction in different living and working contexts, to show dynamic aspects of interaction, to bridge time spans and explain workings behind the scenes (e.g., adaptation, profiling, communication), with narration and subtitles or even illustrative video effects.

Video prototyping was introduced in the HCI community as a way of obtaining early feedback from users, see for example [8]. However, already from its earliest days, a range of famous vision videos capitalized on the ability of the technique to illustrate futuristic technologies, e.g.,the STARFIRE video by Sun Microsystems[7], or the seminal Knowledge Navigator[5] video by Apple, which have managed to capture features that are now part of the current technological habitat. The critical viewer though, will note how these videos trivialized some technological leaps, which are still not possible with todays technology. It is only too easy to visualize non feasible technologies: in the extreme example of time

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 403–408, 2012.

travel, only a comment by the narrator could be enough to tell us we moved backwards in time. Tognazzini provided a set of guidelines for how these videos should be filmed, assuming the emphasis of the filmaker and the audience is on interaction design. Over time the community has become very much accustomed to these videos as a representation of a system, and to the pitfalls they present[9].

Much of the knowledge on video prototyping can be characterized as anecdotal or craft knowledge. There has been little attempt to provide empirical evidence for the advice given and to consider the applicability, validity and generalizability of such advice. A related line of research has attempted to address this limitation in literature. For example, we have examined the impact of fidelity in representations used in video prototypes. The generally perceived wisdom that low fidelity video prototypes lead test participants to be more critical and to focus on higher level detail, was not confirmed in the case of video prototypes [2]. Different filming techniques were compared, e.g., using actors versus using cut-out animations [6].

These studies, while useful in examining the nature of feedback that users or user representatives may offer as (re)viewers of a video, do not provide insights as to how videos as design representations influence the design process. This paper aims to fill this gap, by considering the impact of video prototyping on the overall design process.

The embedding of a video in the design process is interesting for several reasons. A concise and vivid video representation can have communicative and persuasive uses towards managers, a development team, but can also serve as common ground within a design team. On the other hand, it might draw attention to issues captured well, while ignoring others, and a slick presentation may conceal serious usability end-user experience limitations of the envisioned design.

2 Method

We present a case study of the use of video in the design of an ambient intelligence system, where the focus of the design was on the related user experience. The case at hand was a postgraduate student design project, so many of the contextual and organizational constraints are those common to academic environments, e.g., the bounded duration of the project, the primary motivation of the designers being to learn, and the main motivation for the design being research and innovation rather than profit. The designers were all postgraduate students with different backgrounds: two engineers, one computer scientist and two psychologists, a constitution that can be found in many user experience design teams.

An important disclaimer is added here; both the video prototype and the design process we followed are not put forward as some ideal archetypes that must be followed by all designers. Rather we recognize that every video production and design process has its own strengths and limitations. In presenting these here, we follow a hermeneutic approach for presenting case studies, disclosing and qualifying our own subjective frame of reference in our attempts to reach reliable conclusions.

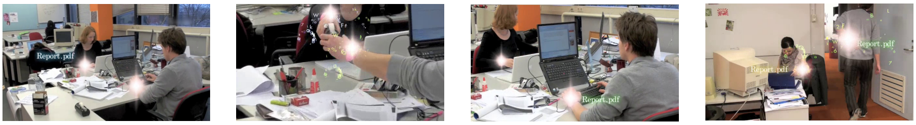
The design was carried out in two phases: the first one (lasting one week) was part of a hands-on course on video prototyping ,emphasizing on developing skills in this technique. The second phase was part of a 12 week team project. Between the two phases, a psychologist and the computer scientist from the original team gave their place to an engineer and a computer scientist. Below we describe the use of video in this design process, and follow the evolution of the design until a working prototype was created and evaluated. We focus specifically on the role of the video prototype, describing how it was used and how it impacted or failed to impact the design process, and reflecting on our successes and failures. The implementation and the user evaluation of the actual application, are reported in [1].

3 Video Prototyping

The design brief was to seek for a technologically plausible system within the scope of ambient intelligence applications, which would facilitate opportunities for informal communication amongst co-located knowledge workers. The following steps we taken:

- Video Contextual Interviews with 4 students (2 PhD and 2 Bachelor's) were conducted, focusing on how informal communication with colleagues takes place.
- A group interpretation session was held to codify and distill the interview material.
- A classic brainstorm session for the purpose of merging ideas was followed by a video-brainstorm. Each team member individually worked out and proposed a concept, showcasing it in a short video or animation.
- After the final concept had been agreed upon, the final video prototype was produced.

The interviews indicated that while people are interested in getting information, they don't explicitly attend to providing information to others, but will respond if specifically asked. It also appeared that people are not keen on making the effort to actively share. They will however do so when it's essential to their work, or when it does not require that much effort. Additionally, face-to-face communication was deemed to be the easiest way to share information, and this usually takes place in rooms designated for common use such as lounges and pantries and around facilities such as coffee machines and water coolers. We note here that the investigation was not very large and we did not seek new insights into informal office communications; as several authors have addressed these before. The educational aim was to familiarize with using video in contextual interviews, while the purpose in the context of the design project was to ground the design and acquire a first-hand understanding of the design challenge, rather than rely on literature only. The final concept is a system that allows the sharing of digital information on the basis of physical proximity of people. Furthermore, it employs a viral manner of transmitting that information, in order to diversify its flow and increase exposure of the shared items to more people.



Worker on the left carries a document on a wearable device, shown by a flare.

Through a typical interaction, devices exchange data, shown by streams of particles.

The document is acquired by the person on the right, without attracting users' attention.

The document is disseminated to another colleague as the protagonist passes her by.

An exchange takes place between the protagonist's device and a 'sink' in the office space (here, a coffee machine).

His device gets more files, left previously by others, and transmits 'Report.pdf' to the machine.

At a later time, an unrelated worker walks up to the machine and an exchange starts again.

She receives 'Report.pdf' because of the viral transmition model, in a 2 step process from the initial carrier

Fig. 1. A walktrhough of the video prototype

The video runs for approximately 5 minutes. It depicts a group of co-workers in a short snapshot of daily routine. They wear wireless communications nodes on their wrists. One individual receives a document from a colleague's node and as he moves on, spreads it to other nodes, either worn by co-workers, or embedded in the surroundings. The video ends with an unrelated member of the staff receiving that same document through the system's mechanism of viral spreading. Figure 1 illustrates the scenario that plays out.

The focus of the video prototyping effort was to convey a high level concept of what the system is about. In order to produce the video, a small storyboard was produced to make sure that the shots and the timeline would sufficiently depict the concept.We avoided complex shots or the use of dialogue. Four "actors" were used from the pool of fellow students, and shooting took place in 2 different rooms. In order to indicate the wireless communications taking place, computer-generated animation was added in post-production, consisting mostly of 2D-particle effects. Recording was completed in less than one hour and post-production took almost a day to complete. Overally, the video was received positively by viewers.

This effort and amount of sophistication is substantially less than the effort put in corporate vision videos like the Knowledge Navigator and the Starfire mentioned earlier. However, it was deliberately representative of how video should be used as a sketching (during the brainstorm) and prototyping technique; see for example the early rationale by [8] or the more recent arguments by Buxton on using video to sketch interaction [3].

4 Realization of the Prototype and Field Test

Based on that video prototype, we moved forward to implement the system and conduct a field trial; these are reported more extensively in [1]. The following design decisions were captured in the video and were implemented in the system:

- The system is operational within an organization, serving workers who use the same physical spaces.
- Communication takes place through radio nodes, which are small, wearable devices.
- The system would also include nodes embedded in places or devices that are frequently visited by workers.
- Communication is wireless and takes place upon what one would consider a casual encounter between two people in the workplace, whether this encounter is acknowledged by the parties involved or not.

5 Reflections on the Use of Video in the Design Process

The video prototype provided an accessible entry point into the next design phase, the design and implementation of a working system. The video is fairly rich in content, implicitly capturing many design decisions and visualizing context and functionality in a concise manner. The design team readily adopted it as a tool for the elicitation of remarks and insights by a focus group of likely users early in this second phase of design. Given that 2 new members of the team had not been involved in the previous phase, its role as common reference point for team discussions should be acknowledged. It helped keep the team's perception of the system aligned and provided a baseline of features upon which to add or take away from. This is in line to the point made about video serving as social glue in [9].

The video prototype assumed a plausible and reliable platform, therefore the particulars of the field test configuration were not in the scope of the consideration of the focus groups that used videos, echoing experiences reported earlier by designers of ubicomp system, e.g., [4]. Essentially, the video prototype provided an idealized view of system use and the emerging user experience. This underlines once more how actual implementation and field testing are irreplaceable in ambient intelligence. It would suggest however that knowledge of such limitations and glitches can inform the scripting of relevant videos. Once designers are aware of technical limitations, e.g., thresholding of sensor networks, initialization issues, etc., these and potential effects on the emerging user experience can be visualized in video when user feedback is solicited.This would allow for a user experience assessment at a very early stage of the design process

Additionally, there was an indirect assumption stemming from the video, that there would be an application to allow users to submit content for sharing, and view content they had picked up from others. However, the video did not depict any actual interaction of the user with such a system. It turned out that this was a critical part of the concept, since it is essentially the single point of explicit interaction for users and application (the remaining of the interaction was implicit).

Clearly the video production focused on the innovative and challenging ambient intelligence aspect of the design, neglecting the complete lifecycle of information and the complete workflow. An opportunity to discuss this aspect with users early in the implementation phase could have been provided by a video prototype that would depict a complete task/workflow with all relevant system components, even if these were assumed to be contrived and standard technologies.

6 Conclusion

This paper has discussed the use of video prototyping in the design of ambient intelligence applications, especially focusing on the embedding of this method in the design process. The video production discussed in this paper, was found useful to quickly and emphatically convey a context of use and the core aspects of an immaterial, implicit, and invisible interaction that was the core of the design concept and is quite representative of many ambient intelligence systems. However, it tended to gravitate the attention of the design team to those aspects best visualized on screen and to neglect several aspects that were considered less important during filming. Designers of ambient intelligence can enhance the utility of video as design representations by: showing in the video complete tasks or workflows, even including more mundane aspects of the system, and the effects on interaction of some of the limits of context sensing technologies, that can be known to the design team a priori but are obfuscated often by idealized video illustrations of design concepts.

References

1. Batalas, N., Bruikman, H., Turzynska, D., Vakili, V., Voynarovskaya, N., Markopoulos, P.: SIREN: Mediated Information Communication for Serendipity. In: UBICOMM 2010, pp. 397–405 (2010)
2. Bojic, M., Goulati, A., Szostak, D., Markopoulos, P.: On the effect of visual refinement upon user feedback in the context of video prototyping. In: Proceedings ACM SIGDOC 2011, pp. 115–118. ACM, New York (2011)
3. Buxton, B.: Sketching User Experiences: getting the design right and the right design. Morgan Kaufmann (2007)
4. Dow, S., Saponas, T.S., Li, Y., Landay, J.A.: External representations in ubiquitous computing design and the implications for design tools. In: Proceedings ACM DIS 2006, p. 241 (2006)
5. Dubbcrly, H., Mitch, D.: The Knowledge Navigator. Apple Computer, Inc., video (1987)
6. Dhillon, B., Banach, P., Kocielnik, R., Emparanza, J.P., Politis, I., Pczewska, A., Markopoulos, P.: Visual fidelity of video prototypes and user feedback: a case study. In: Proceedings BCS-HCI 2011, pp. 139–144 (2011)
7. Tognazzini, B.: The Starfire video prototype project: a case history. In: Proceedings CHI 1994, pp. 99–105. ACM, New York (1994)
8. Vertelney, L.: Using video to prototype user interfaces. SIGCHI Bulletin (1989)
9. Ylirisku, S., Buur, J.: Designing with Video: Focusing the user-centred design process, 1st edn. Springer Publishing Company (2007) (incorporated)

Automatic Power-Off for Binaural Hearing Instruments

Bernd Tessendorf[1], Peter Derleth[2], Manuela Feilner[2],
Daniel Roggen[1], Thomas Stiefmeier[1], and Gerhard Tröster[1]

[1] ETH Zurich, Wearable Computing Lab.
Gloriastr. 35, 8092 Zurich, Switzerland
{lastname}@ife.ee.ethz.ch
[2] Phonak AG
Laubisrütistrasse 28, 8712 Stäfa, Switzerland
{firstname.lastname}@phonak.com

Abstract. Users of state-of-the-art hearing instruments (HIs) switch their devices on and off by closing and opening the battery compartment. Switching the HIs off is important for the users to maintain the battery lives of their HIs. However, users currently need to switch off their devices manually, which is easy to forget or which can be difficult, e.g. for elderly with reduced dexterity. In this work, we propose an approach to avoid the need to manually switch off HIs. We assume, that whenever the user's HIs are not moved the same way, they cannot be at the user's ear and are, thus, not in use. We exploit the binaural communication between the user's HIs available in the latest generation of HIs together with the concept of multimodal HIs, which integrates sensors such as accelerometers. On a data set of one hour comprising acceleration data of two HIs worn by three male participants (age 26–31) we achieve a precision of 100% and a recall of 93% in detecting power-off events.

Keywords: hearing instruments, human computer interaction (HCI), multimodal sensing.

1 Introduction

Users of state-of-the-art hearing instruments (HIs) switch their devices off by opening the battery compartment (see Figure 1a–b). Switching the HIs off is important for the users to maintain the battery lives of their HIs. HI users switch off their devices whenever they do not use them. How often a device normally is being switched off is user dependent: Most commonly users switch off their HIs before going to bed and taking a shower, but also in special situations, e.g. for concentrated reading or working. The typical runtime of a HI battery set is about one week. Usually disposable zinc-air batteries are used, mainly because of their high volumetric energy density, flat discharge curve, and low cost.

User's currently need to switch their devices manually, which can be a burden, especially for the elderly due to reduced dexterity. We suggest that the HIs could

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 409–414, 2012.

Fig. 1. A HI with the battery compartment closed (a) and opened (b). The opening and closing mechanism is used to switch the HI off and on, respectively. (c) HI (1) with head movement sensor (2).

be automatically switched off or enter a power saving mode when their movement patterns indicate that they are not in use. We investigate whether recognition of movement patterns can be used to power off the HIs automatically. The movement patterns could be obtained by sharing the use of an existing acceleration sensor, which multimodal HIs use to estimate the user's current hearing wish [9]. Our goal is to provide an automatism, which could be appreciated by HI users similar to automatic hearing program selection, which obviates the need to use HI buttons to change the hearing programs [1,2]. Such a system could ease the handling of HIs, especially for elderly users with reduced dexterity, by reducing the need to operate the battery compartment.

Related Work. Several research groups have worked on correlating movement patterns of at least two moving objects. *Lester et al.* used accelerometers to determine if two devices are carried by the same person [6]. *Mayrhofer* and *Gellersen* suggested to shake devices together to securely pair them for later encrypted data exchange [7]. *Rossi et al.* implemented a pervasive game using smart dice, which integrated acceleration sensors to detect whether they were shaken together [8]. *Wirz et al.* assume a similar walking behavior of pedestrians walking in groups. In their work they propose an approach to identify pedestrians walking together by correlating the acceleration signals from their mobile phones [10]. *Gordon et al.* characterized a ball switch as a wearable vibration sensor for activity recognition, which could be a suitable sensor for integration into HIs for our application. They concluded that this ball switch could be used to effectively improve recognition rates achieved with accelerometers while representing a very low cost sensor in terms of price, device size and power consumption [3].

2 Explorative Data Set

To collect reference data we used a head movement sensor (see Figure 1c), which is part of our multimodal HI prototype that integrates a movement sensor and a state-of-the-art HI. The head movement sensor is based on the BodyANT platform [5] comprising a triaxial accelerometer and data transmission using the wireless ANT+ protocol. The head acceleration data was sampled and transmitted at 32 Hz and could be received by a smartphone (*Sony Ericsson Android Xperia Active*). For our application a reduced sampling rate would be sufficient. However, the main purpose of an HI-integrated accelerometer is also automatic hearing program selection, which requires for the analysis of the user's head movements a higher sampling rate than for detecting power-off events only. The prototype including the head movement sensor and mobile phone served as a research platform to study multimodal HI with the aim of a full integration of the acceleration sensor in the HI housing. In a future generation of HIs, the acceleration sensor and signal processing would be integrated into the HI itself, obviating the need for external devices and radio communication. The primary intended purpose of an HI-integrated acceleration sensor is to support the adaptation of the HI to the user's current hearing wish [9]. One element of the user's hearing context could be "HI is not used", which we could recognize by using the acceleration sensor available in multimodal HIs.

We recorded one hour of acceleration data from three normal hearing male users (age 26–31). Activities included walking around, standing still, office work and walking stairs. We also included activities such as shaking the head and jumping, which could in particular be challenging for our system due to potential

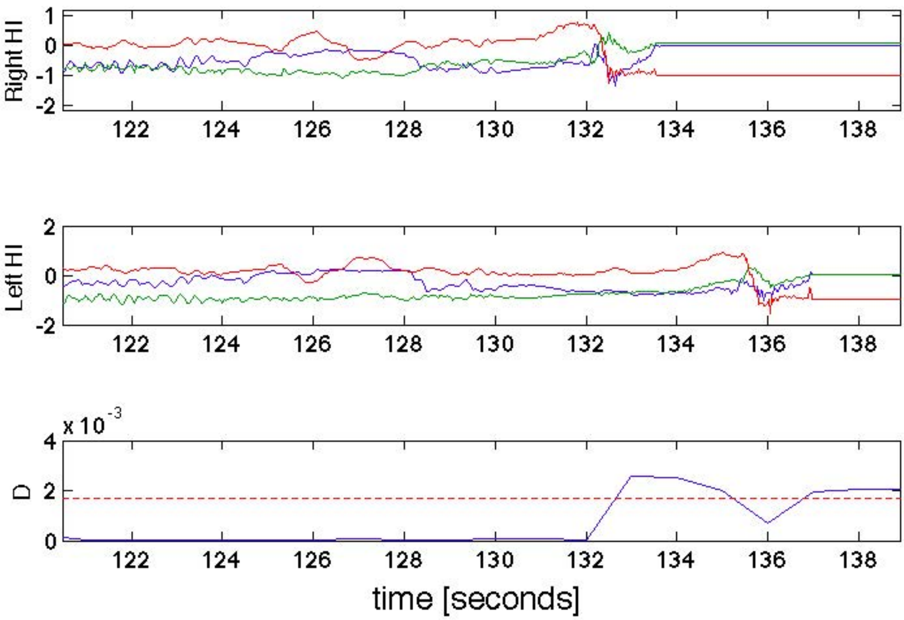


Fig. 2. Recorded acceleration data (in units of g) of an user putting the HIs off his ears, resulting in an increased dissimilarity measure D (in units of g^2/s) as defined in Equation 1. The dashed red line represents the threshold value. The HIs power off as soon as the threshold is exceeded.

differences in the movement patterns of the two HIs. In total, 15 instances power-off events occurred of the users putting their HIs off his ears. Figure 2 depicts recorded acceleration data of a user putting the HIs off his ears. We manually annotated the ground truth describing whether the HI should be switched off.

3 Algorithm for Automatic Power-Off

By comparing the data from the accelerometers at both of the user's HIs, we determined whether the HIs were not in use anymore. Our base assumption is, that whenever the user's HIs were not moved the same way they could not be at the user's ear and were, thus, not in use. E.g., when the user removed his HIs from the ears, the HIs would show a different movement pattern and could be switched off.

We continuously determined whether the HIs were moved or not. We used a sliding window approach with a window size W and a fixed step size of one second. For each window we compared the acceleration data $a = (a_x, a_y, a_z)$ from the two HIs. We calculated for each sample i inside the data window the magnitude of the acceleration $|a_i|$ and calculated the variance of the magnitudes of all data points in the data window, normalized to the window size W:

$$D = \left| \frac{\mathrm{Var}(|a_{left}|) - \mathrm{Var}(|a_{right}|)}{W} \right| \tag{1}$$

We used a threshold T to decide whether the HIs move the same way $(D \leq T)$ or not $(D > T)$.

4 Results and Discussion

With a parameter sweep we found the optimal threshold value of $T = 0.0017$ g^2/s for the window size W of four seconds. These parameters were fixed for all participants. All of the automatically detected power-off events were correctly detected according to the ground truth, corresponding to a precision of 100%. Thus, the HIs would not switch off if the user was still using them. The system was not confused by jumping or arbitrary head shakes. From all 15 correct power-off events we spotted 14, corresponding to a recall of 93%. For the missed power-off event the user put down both HIs at the same time using two hands, in all other events the users put one HI down after the other. In the missed case the HI, however, switched off afterwards, when the user put one HI after the other into the HI carrying box.

The suggested automatic power-off addresses two issues:

- **Saving Energy Due to Automatic Power-Off.** The amount of saved energy depends on the user. If the user would never forget to switch off the HI, then there is no saved energy. If the user would never switch off the HI, e.g. because of a lack of dexterity, the automatic power-off would almost

double the HI runtime, taking into account that the HIs will be switched off over night. We assume, that the additional power consumption due to the accelerometer can be neglected for multimodal HIs, that feature anyway an accelerometer to analyze the user's head movements for automatic hearing program selection.
- **Ease of Use.** Due to automatic power-off the user does not have to open the battery compartment for powering-off. Combined with an automatic power-on the user would not need to perform opening or closing of the battery compartment on a daily basis any more. This can be in particular relevant for elderly people with reduced dexterity. However, for changing batteries, the battery compartment still needs to be operated, which is necessary once a week for common HIs.

The latest acceleration sensors provide an integrated automatic wake-up and power-off functionality[1] However, this functionality is not suited for our application, because it would power-of in the use case, in which a person is not moving the head, but could still be listening to something. It might be possible to overcome this issue with an analysis of the history of the head's movement. Moreover, different features and algorithms could be applied to characterize the differential movement of both of the HIs.

Limitations. Our algorithm required the user to wear binaural HIs (one HI at each ear) and was not suited for users of one HI only. A study by *Kochkin* showed that 74% of the HI users wore two devices [4].

Powering on the HIs still needs to be performed manually by the user. The next step is to enhance the movement analysis to allow for automatic power-on, e.g. by exploiting the wake-up functionality of accelerometers.

To confirm the promising results we need to perform a study with a larger number of participants and include HI users to assess the user acceptance.

5 Conclusion and Outlook

We achieved a precision of 100% and a recall of 93% in automatically detecting power-off events for the HIs by comparing the movement patterns of the devices. The reason for the single missed power-off event was the user removing both HIs at a time, instead of removing one after the other. The recall could be improved by including additional features, which would allow to distinguish the HI movements in a more fine-grained way. Moreover, our proposed approach could be adapted for different sensors, e.g. magnetometers, gyroscopes, or ball switches. We plan to extend our approach with an automatic power-on functionality and conduct large-scale studies with HI users to validate the benefit of the system.

Acknowledgements. This work was part-funded by the Swiss CTI project 10698.1 PFLS-LS. We especially thank all participants in the study, and Hilmar Meier for valuable discussions.

[1] E.g., the Kionix sensors `http://www.kionix.com` offer this kind of functionality.

References

1. Büchler, M.: Nützlichkeit und Akzeptanz einer automatischen Programmwahl in Hörgeräten. In: 27th Annual Convention for Acoustics, DAGA (2001)
2. Büchler, M., Allegro, S., Launer, S., Dillier, N.: Sound classification in hearing aids inspired by auditory scene analysis. EURASIP J. Appl. Signal Process. 2005(1), 2991–3002 (2005)
3. Gordon, D., Von Zengen, G., Schmidtke, H., Beigl, M.: A novel micro-vibration sensor for activity recognition: Potential and limitations. In: Proceedings of the Fourteenth International Symposium on Wearable Computers, ISWC 2010 (2010)
4. Kochkin, S.: MarkeTrak VIII: 25-year trends in the hearing health market. Hear Rev. 16(10), 12–31 (2009)
5. Kusserow, M., Amft, O., Tröster, G.: Bodyant: Miniature wireless sensors for naturalistic monitoring of daily activity. In: BodyNets (2009)
6. Lester, J., Hannaford, B., Borriello, G.: 'Are You with Me?'-Using Accelerometers to Determine If Two Devices Are Carried by the Same Person. Pervasive Computing, 33–50 (2004)
7. Mayrhofer, R., Gellersen, H.: Shake well before use: Intuitive and secure pairing of mobile devices. IEEE Transactions on Mobile Computing 8(6), 792–806 (2009)
8. Rossi, M., Lombriser, C.: Smart dice. In: 3rd European IEEE Conference on Smart Sensing and Context (2008)
9. Tessendorf, B., Bulling, A., Roggen, D., Stiefmeier, T., Feilner, M., Derleth, P., Tröster, G.: Recognition of hearing needs from body and eye movements to improve hearing instruments. Pervasive Computing, 314–331 (2011)
10. Wirz, M., Roggen, D., Troster, G.: Decentralized detection of group formations from wearable acceleration sensors. In: International Conference on Computational Science and Engineering, CSE 2009, vol. 4, pp. 952–959 (August 2009)

Proposal and Demonstration of Equipment Operated by Blinking

Masaki Kato, Tatsuya Kobori, Takayuki Suzuki,
Shigenori Ioroi, and Hiroshi Tanaka

Kanagawa Institute of Technology,
1030 Shimo-ogino, Atsugi-shi, Kanagawa, Japan
{s1185040,s0921049}@cce.kanagawa-it.ac.jp,
{suzuki,ioroi,h_tanaka}@ic.kanagawa-it.ac.jp

Abstract. This paper describes a new input method that makes use of eyelid blinking. We found that the electromyographic (EMG) signal generated by blinking can be detected using a commercially available brain sensor. Since it is impossible to distinguish between voluntary and involuntary blinks, we propose setting a specific time duration between eyelid closing and opening. This duration can be used as a trigger for signal generation and at the same time for selection of a particular operation. The blink pattern is interpreted as a signal pattern for operation and corresponding commands are assigned for the operation selected. We built a demonstration system to evaluate the proposed method. The validity of the method and the effectiveness of the system were confirmed by the experiment using the system.

Keywords: Blink, EMG signal, Brain sensor, Signal generation, Equipment operation.

1 Introduction

Many kinds of schemes such as gesture and voice recognition have been investigated as man machine interface methods with the objective of enhancing convenience or making operation easier [1]-[2]. Some methods that use eye direction or eyelid blinking have been proposed as they are particularly suitable for disabled person and ALS patients. These motions are detected using a camera [3]-[4]. Performance of these systems is affected by the level of brightness in the usage area. The camera's field of view also restricts the usage range of the system. A user must be located in front of the camera within a distance of about one meter. An infra-red system that is attached to glasses [5] disturbs visibility, it is not appropriate for operating television etc. Some schemes have been developed that detects a blink by change in the electromyographic (EMG) signal. However, they are not practical because they require a complex setup for detecting and this causes users to experience a degree of discomfort. We found that a blink can be detected by an inexpensive and commercially available brain wave sensor.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 415–420, 2012.

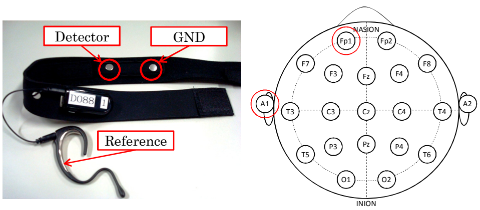
This paper proposes a new signal generation method for an input interface that uses blinks detected by a brain wave sensor. A demonstration system was constructed to evaluate the proposed method. The operation of home electrical appliances such as a TV or LED lights was carried out using the proposed method. Experiments for determining the success ratio and the time required for operation were carried out and the method's performance was clarified. It was verified by experiment that the proposed method shows promise as an input interface.

2 Blink Detection and Its Characteristics

The appearance and detection points of a sensor for blinks are shown in Fig.1. This brain wave sensor is available commercially [6]. The main function of the sensor is to detect the difference in the electric potential between the ear and the forehead (Fp1) as shown in Fig.1 (b) and to analyze the strength of alpha and beta waves.

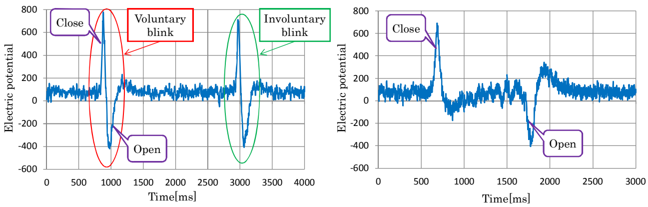
The electric potential in a voluntary blink and an involuntary blink is shown in Fig.2 (a). This means that generating a signal from a blink is a real possibility. However, it is crucial to be able to discriminate between them if blinks are to be used for generating signals because both of them are the same signal pattern. A method by which the start of a signal can be recognized is vital if EMG signals generated by blinks are to be used.

Only a voluntary action can be used as a trigger for signal generation. Figure 2 (b) shows changes in EMG when the time duration between eyelid-close and eyelid-open is kept by conscious user intention. This signal pattern can be used as a trigger for signal generation. It was confirmed by a pre-examination that the change in the shape of the EMG signal is almost the same. Only the potential amplitude is somewhat different among persons, but this difference is sufficient to discriminate between a blink and the noise floor of the sensor.



(a)Brain wave sensor (b)Sensing position

Fig. 1. Brain wave sensor and sensing position



(a)Voluntary and involuntary blink (b)Eyelid closing / opening

Fig. 2. EMG signal characteristics

3 Triggering and Signal Generation

The signal generated by a blink pattern can be used for the input interface, for example, equipment operation, character input and communication of intention. The following sequence shown in Fig.3 is proposed for signal generation in this study. The sequence is composed of three phases, that is, the trigger, signal generation and decision phases. The blink is identified by the threshold levels (U_{th} and L_{th}). These levels were decided by conducting a pre-examination in which the maximum and minimum potential values for blinks were obtained.

$$U_{th} = \mu_{max} - \sigma_{max} \tag{1}$$

$$L_{th} = \mu_{min} + \sigma_{min} \tag{2}$$

$\mu_{max/min}$: average of max/min values
$\sigma_{max/min}$: standard deviation of max/min values

The trigger phase is created by the duration of eyelid-close/open as shown in Fig.2 (b) since this action can only be carried out consciously. In addition, differences in this duration can be used to supply additional information. Duration time information can be used for equipment selection, for example, 1 second is for a TV, 2 seconds is for a room light.

The second part is the signal generation phase. We propose that the blink pattern is regarded as signal pattern. The signal pattern is created by a combination of blinks as shown in the middle section of Fig.3. Three time slots are introduced in this phase where one time slot a has duration of 1 second. There are 2x2x2 combinations, that is, 8 kinds of signal can be generated in this example. The existence of a single blink is interpreted as 1, no existence is 0. The last part is the decision phase for command transmission. This part is explained in section 4.

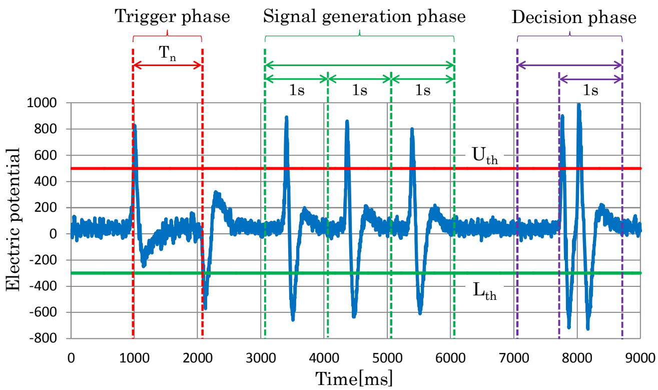


Fig. 3. Sequence for signal generation

4 Application for Equipment Operation

The signal generated by a blink pattern has many areas of application. We demonstrate the effectiveness of the proposed method using equipment operation as an application example. Not only ALS patients but also injured persons, for example, patients immobilized by a cast due to bone fracture etc. cannot operate peripheral electrical appliances such as TV and lights. The system created for the demonstration is shown in Fig.4. The EMG signal is transmitted via Bluetooth from the brain sensor to a PC. The generated signal pattern is interpreted in the PC and the corresponding command is sent to the appliances via an IR transmitter we developed. This unit stores pattern data consisting of infrared signals and these patterns are transmitted to operate a target appliance. This system configuration makes it possible to operate any appliance that can be controlled by an infrared remote controller.

The blink patterns and assigned commands for this demonstration system are shown in Table 1. Three appliances in the laboratory, that is, a TV, an LED light and an electric fan are selected for the demonstration. The IR pattern data are transmitted via the PC based on the blink pattern. The T_n (n = 1, 2, 3) in Table 1 indicates the duration time for appliance selection.

The main feature of this command assignment is that the same signal pattern can be used by using T_n even if the operation target is different. This makes it easy to generate a signal pattern and, at the same time, the limited number of signal patterns can be used effectively. As the number of time slots in the signal generation phase increases, the larger the number of signal patterns that can be generated. But this means that it takes longer to generate a signal and it becomes difficult to memorize the signal pattern sequences and their corresponding commands.

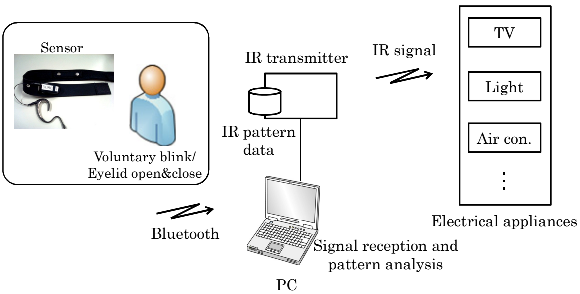


Fig. 4. System configuration

Table 1. Blink patterns and assigned commands

Blink pattern			T_1 = 1000[ms]	T_2 = 2000[ms]	T_3 = 3000[ms]
			LED light	Electric fan	TV
Y	Y	Y	Power On	Power On/Off	Power On/Off
Y	Y	N	Color fluorescent	Rover On/Off	Volume down
Y	N	Y	Color medium	Direction change On/Off	Channel 8
Y	N	N	Max. brightness	Wind up	Channel increment
N	Y	Y	Color bulb	Wind down	Channel decrement
N	Y	N	Medium brightness	Direction change On/Off	Channel 1
N	N	Y	Min. brightness	Rover On/Off	Volume up
N	N	N	Power OFF	Power On/Off	Power On/Off

Y: Blink, N: No Blink

5 System Evaluation

The operation sequence shown in Fig.5 is embedded in the PC. Time duration T_n decides the operation target in the trigger phase. The decision phase has three operations in this implementation. These are signal decision for correct signal generation, return to the signal generation phase for an incorrect signal, and return to the start point that indicates operation cancellation. Two successive blinks, three successive blinks and one or no blink are assigned in the decision phase. A sound to indicate the timing of blinks is produced in order to ensure T_n interval. Time slots of patterns are displayed on the PC. The experimental setup is shown in Fig.6. The sensor is attached on forehead, this is the same way used as brain sensor. The experimenter sits in front of the PC so as to monitor the PC in which the commands generated by blinks are shown.

The results of the experiment are shown in Fig.7. Total 150 times trials by 3 persons were carried out for confirming the validity of the proposed method. The two criteria, that is, success ratio of operation by blinks and required time for operation were evaluated in the experiment. All commands shown in Table 1 were selected for evaluation. Fig.7 (a) shows the success ratio and required time for operation with no repeat of signal generation or cancellation, that is, with no operation error. In the experiment, the success ratio was almost 70% with an operation time of 8 to 10 seconds. The operation error of about 30% is due to the timing error of blinking, not caused by blinking detection error. If we permit operation reentry and operation cancellation due to mis-operation, the success ratios increase to more than 96%. Of course, the operation times are longer where reentry or cancellation occurs when compared to the error-free example. The subjects of this experiment were 3 men accustomed to operating the system. However, it is not difficult to get accustomed to operating this system. This result confirms that the system can be used to operate home electrical appliances because mis-operation does not cause fatal errors.

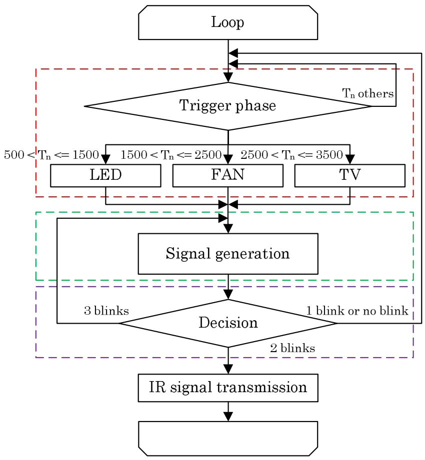


Fig. 5. Operation sequence

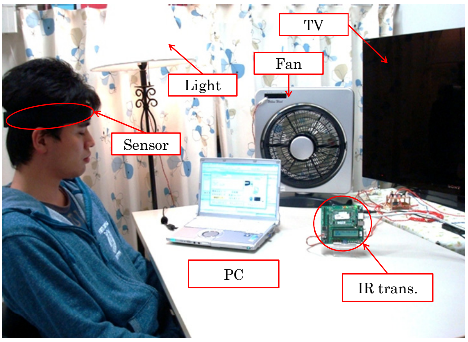
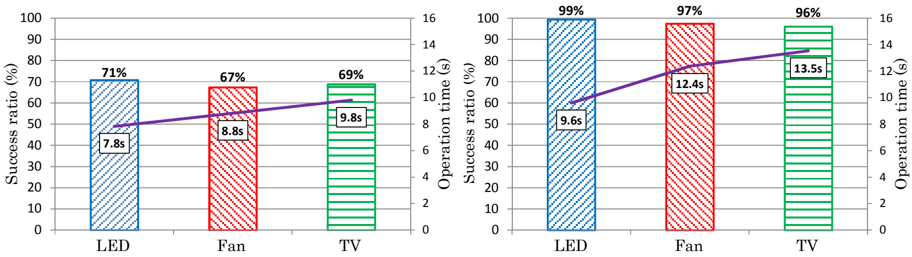


Fig. 6. Experimental setup



(a)Operation with no error (b)Operation with reentry

Fig. 7. Success ratio for pattern generation

6 Conclusion

We have proposed a new scheme for an input method using eyelid blinks. The blink pattern and eyelid close/open duration can be used to generate signals for operating a range of appliances. The EMG signals are detected by a commercially available brain wave sensor. The proposed operation scheme was implemented using a PC. A demonstration system was created for evaluation of the proposed method and its effectiveness was verified. A success ratio exceeding 96% was obtained and it has been confirmed by experiments that the proposed method and system is practicable. More detailed experimental evaluation considering many users operation and various conditions, and investigating the calibration of threshold are remained as further studies.

References

1. Tanaka, H., Kimura, R., Ioroi, S.: Equipment Operation by Motion Recognition with Wearable Wireless Acceleration Sensor. In: Int. Conf. on Next Generation Mobile Applications, Services and Technologies, pp. 111–118 (2009)
2. Siri : `http://www.apple.com/iphone/features/siri.html`
3. Chau, M., Betke, M.: Real Time Eye Tracking and Blink Detection with USB Cameras. Boston University Computer Science Technical Report, No.12, pp. 1–10 (2005)
4. Hori, J., Sakano, K., Saitph, Y.: Development of a Communication Support Device Controlled by Eye Movement and Voluntary Eye Blink. IEICE Trans. Inf. & Syst. E89-D(6), 1790–1797 (2006)
5. Infrared Sensor : `http://www.words-plus.com/website/products/input/istswtch.htm`
6. Brain Sensor : `http://www.neurosky.com/`

CASi – A Generic Context Awareness Simulator for Ambient Systems

Jörg Cassens, Felix Schmitt, Tobias Mende, and Michael Herczeg

Institute for Multimedia and Interactive Systems, University of Lübeck,
Ratzeburger Allee 160, D–23562 Lübeck, Germany
{cassens,schmitt,herczeg}@imis.uni-luebeck.de

Abstract. In this paper, we present CASi (Context Awareness Simulator), a software system for the simulation of context-aware computer systems and environments. CASi provides an abstract framework of components for simulating smart world applications like a smart office or house with ambient sensors and actuators. Agents moving through these application worlds are tracked by sensors and their actions are influenced by actuators, both of which can be programmed to resemble the actual peripherals of the tested system. CASi allows testing ambient, context aware computer systems even in early stages of development without the need for expensive prototyping or real world deployment.

Keywords: Ubiquitous computing, ambient intelligence, context, simulation.

1 Introduction

When designing and building ambient, context aware computer systems, extensive testing will be necessary due to the complexity, the distributed nature and the richness of human interaction of such systems. Deployment of a prototype to its actual target domain is one obvious way to handle this phase of the development process. This approach, however, has severe drawbacks: The deployment of the system consumes time during which development resources are occupied, and the provision of actual sensor and actuator hardware is expensive. Furthermore, human subjects will need some time until they interact with such a system in a natural way. Changes to the system design and even regular development iterations require updates to the system's peripherals and their installation. These problems can be mitigated by implementing virtual sensors, actuators and users in a simulation environment.

MACK [11] is a framework developed by our research group on top of which application systems for different domains can be built. Key requirements for a simulating environment to be used with the development of MACK are:

- applicability across different domains,
- adaptability to new system architectures,
- options to run simulations faster than real time, and
- platform independence.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 421–426, 2012.

In the following sections, we first review and discuss other existing solutions, pointing out their respective advantages and shortcomings. Afterwards, we introduce CASi's design and architecture by describing its key components and the rationale behind it. At the end, first evaluation results are presented before we point out the roadmap for future development of CASi.

2 Related Work

Of particular interest in the field of ubiquitous systems is a special class of deterministic, discrete simulators known as object-oriented simulators [9] and here in particular individual-based or agent-oriented simulators [1]. Basically, an agent-based simulation will be comprised of different simulated actors and their interaction with each other and the world. Instead of describing the probably very complex behavior of the system as a whole, individual agents are described in simpler terms. These agents act simultaneously, and the behavior of the system as a whole is seen as an emergent phenomenon. This approach has been chosen for the CASi simulator. Other approaches and partial solutions for the requirements described can be found:

Siafu [6] is a simulation tool that implements many of the characteristics we demanded in the previous section. The system offers a graphical user interface for visualization purposes which makes it easy for human users to monitor the simulation. Siafu's architecture is rather limited with regard to adaptability and generalizability. Another issue with Siafu is the missing support for plug-in components modeling sensors and actuators. Siafu also requires advanced programming skills not only for implementing new types of sensors, actuators, and protocols, but even for the definition of simulations themselves.

DiaSim [3,4] forms the simulation part of a bigger environment for developing pervasive computing applications. Other components include the DiaSpec language for describing functionality of sensors and actuators in simulated as well as real world systems, and the DiaGen component, which generates an outline of the to-be-written program code for the actual implementation of simulated and real-world components.

In DiaSim, entities like agents, sensors, and actuators are not modeled on an individual level. It uses mathematical density functions to model the probability of persons being present in a particular area of the simulated world and to determine probabilities of relevant actions with regard to the modeled system. The areas from which simulated sensors may record data or across which simulated actuators may convey their output are also described mathematically in an indirect way.

CASS [8] is mainly targeted at detecting rule conflicts in rule-based AI systems for smart-home applications. It focuses on one particular element of ambient intelligent systems, namely a specific kind of reasoning subsystem.

ISS [7], the Interactive Smart Home Simulator, not only models the environment, its inhabitants and various sensors and actuators. Its focus lies on the simulation of home appliances for smart home environments. ISS includes the reasoning components within the simulator itself. Since the reasoners trying to derive context from sensor inputs form a crucial part of any ambient intelligent system, we believe that close ties between simulator and AI lead to reduced flexibility with regard to setting up different testing scenarios.

Similar to ISS, **CAST** [5] also focuses on smart home scenarios. Within its domain CAST aims at simulating data exchange between an ambient system's components not on a functional, but rather on a network and security engineering level.

3 CASi: Concept and Vision

Our development of CASi, the Context Awareness Simulator, is grounded in our work on MACK [11] and MATe [10]. MACK (Modular Awareness Construction Kit) is a framework for constructing ambient intelligent computer systems. MATe (MATe for Awareness in Teams) is an application system built on top of MACK. It aims at supporting team members in knowledge-intensive work environments by fostering awareness of their colleagues' activities and interruptibility status.

One key feature of MACK and MATe is a message passing infrastructure that connects the different actuators and sensors with a hub that does the actual processing, deriving contexts and instigating context-sensitive behavior. Another key feature is the flexible and extensible reasoning subsystem, which can make use of different knowledge representations and reasoning paradigms.

Due to the perceived limitations and different foci of the existing solutions outlined in the previous section, we decided to develop our own simulator. The main focus was set on modularity and encapsulation, since the framework character of MACK demands great flexibility with regard to application domains and, thus, changes in (simulated) peripherals, logics, and communication infrastructure. Generic interfaces between CASi's components also allow modifications of the simulation engine itself.

4 Design and Architecture

In order to develop a highly customizable simulator, we designed an architecture which avoids close bindings between the components. The simulator has been implemented in Java since it is flexible, wide-spread, and platform independent.

The world is a bundle of objects which are needed to describe an environment. The most important element in the world object is a collection of rooms (or, more generally, spaces). Rooms form the map on which agents are able to perform actions. A room is enclosed by multiple walls, which can contain doors. The world also contains a collection of agents and a collection of components like sensors and actuators. It is also possible to embed custom objects. Sensors, actuators, and custom objects can be positioned anywhere on the map, e.g. in any room. Sensors can only monitor a restricted area. Actions and agents outside this area do not influence the component. Actuators in turn cannot influence agents which are outside of their reach.

In CASi, agents are virtual persons who interact with each other and with sensors, actuators and other elements in the simulation independently. They can be used to trigger events in sensors and can be influenced by actuators. Every agent has its own action handler scheduling actions that should be performed next. Actions can be defined with a deadline by which they have to be completed, an earliest start time, duration and a priority. Different action handlers are able to use these parameters to select the most relevant next action for an agent. This results in a dynamic goal stack that guides an agent's activity.

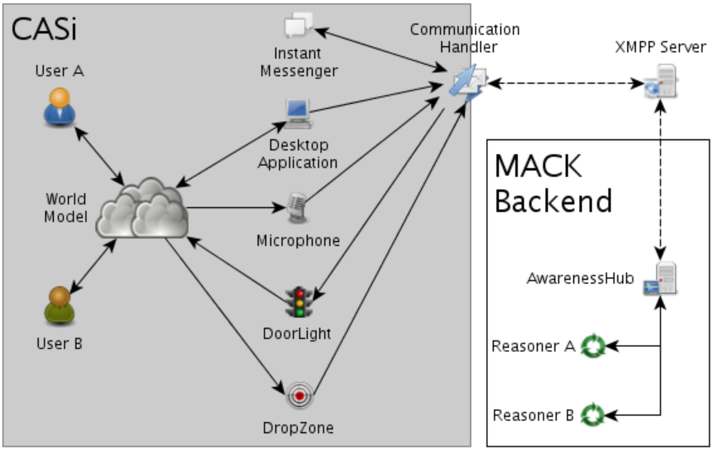


Fig. 1. CASi interfacing with a MACK-based system. CASi simulates the different agents as well as sensors and actuators of the MACK frontend whereas context processing takes place in the actually installed backend. Solid lines denote API calls, dashed lines XMPP connections.

In general, the concept for action handlers defines three different collections. The to-do list contains actions that should be performed once. The real world equivalents are meetings and special tasks. The second collection is a pool containing actions to be performed several times. An example could be short coffee breaks. Furthermore, action handlers may contain an interrupt list containing actions that have to be performed at the next opportunity. This can be used for synchronizing agents, e.g., when they perform an action that involves other agents, like talking. The agent who starts an action schedules another action on the other agent's interrupt list. This prevents the other agent from continuing with other actions before the joint action is completed.

Actions can be postponed for a later execution, e.g. if an agents has to talk to another agent, it may be not allowed to interrupt the agent if it is occupied. The first agent can then schedule the action for a later try.

The communication handler represents connections to external systems, e.g. a network interface which connects to a context middleware or the context aware application itself. It sends sensor values, and receives new values for actuators. Currently, we use the XMPP-based MACK protocol for communication between the simulated sensors and actuators and the MACK backend which contains, amongst other things, reasoners for different contextual aspects (like interruptibility, user activity).

A simple GUI acts as a visual representation of model states. In addition, the GUI allows scaling the time in the simulation and the entire simulation can be paused. Agents, sensors and actuators can be selected to display their state.

An example simulation based on MATe has been developed in order to test modifications and extensions to the architecture on the fly. This simulation models a typical office scenario and consists of several individual office rooms, a meeting room, a coffee kitchen, and restrooms (see Fig. 2). Agents move freely according to their current action list. Virtual sensors and actuators are deployed mimicking a real installation of MATe. Agents have a DropZone on their desk to put a personal token into, a Desktop Activity Analyzer reports on open applications and typing frequency, audio sensors detect the number of persons talking in a room, the state of doors is sensed, and an interactive device called the Cube allows the users to give corrective feedback to the system's reasoners. All these sensory inputs are sent to the MATe system's AwarenessHub and its reasoners (see Fig. 1). The resulting output is forwarded to the simulation again, so the agents are confronted with changing states of actuators like the DoorLight, which signals an office occupant's interruptibility to others, and the Cube, which shows the activity recognized by MATe on its top surface. These actuatory outputs influence the agents' behavior, e.g., if an agent has a talk-to action first on its list, but its counterpart's DoorLight signals non-interruptibility, the agent will postpone the action until later and resume with the next action on its list.

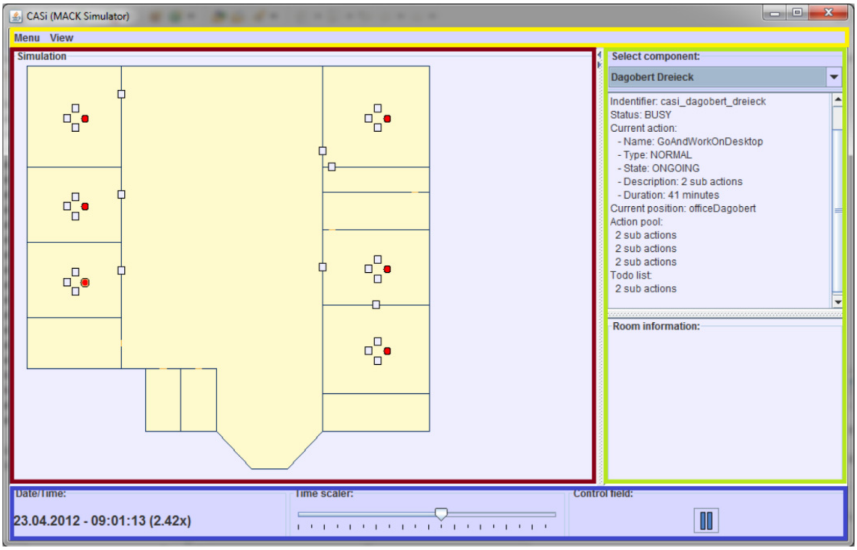


Fig. 2. CASi's SimpleUI visualization interface. Note the detailed information about one agent in the side bar on the right as well as the time scaling slider and the pause button on the bottom.

In addition to the example simulation, which tests the simulator against a known working system, we also ran tests after new components had been added to the system under test. For example, a new case-based reasoner (CBR) for user activities was added to MATe. Simulations were run which included agent activities modeled after results from a field study. Examination of the knowledge base and the system output showed that the reasoner worked as intended. The simulation took only minutes per run, similar data acquisition would have taken half a day each in the real-world.

5 Conclusion

In this paper, we have presented an individual-based simulator for context aware and ambient intelligent systems. While the simulator is customized to interface with the MACK framework, it lends itself to adaptation for other flavors of context middleware. First tests within an ambient application environment have shown the usefulness of CASi for rapid testing of new reasoning components and for testing the effects of adding or removing sensors. In the future, CASi will be tested in further ambient applications and it will be deployed as a modular component to be included in other ambient systems development environments.

References

1. Bonabeau, E.: Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences (2002)
2. Bratman, M.E.: Intention, Plans, and Practical Reason. CSLI Publications, Stanford (1999)
3. Bruneau, J., Jouve, W., Consel, C.: DiaSim: A Parameterized Simulator for Pervasive Computing Applications. In: Proceedings of the 6th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, IEEE (2009)
4. Jouve, W., Bruneau, J., Consel, C.: DiaSim: A parameterized simulator for pervasive computing applications. In: Proceedings of the 2009 IEEE International Conference on Pervasive Computing and Communications, pp. 1–3. IEEE Computer Society (2009)
5. Kim, I., Park, H., Noh, B., Lee, Y., Lee, S., Lee, H.: Design and Implementation of Context-Awareness Simulation Toolkit for Context learning. In: Proceedings of the IEEE In-ternational Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing. Workshops, vol. 2, pp. 96–103. IEEE Computer Society (2006)
6. Martin, M., Nurmi, P.: A Generic Large Scale Simulator for Ubiquitous Computing. In: Annual International Conference on Mobile and Ubiquitous Systems, pp. 1–3 (2006)
7. Nguyen, T.V., Nguyen, H.A., Choi, D.: Development of a Context Aware Virtual Smart Home Simulator. CoRR abs/1007.1274 (2010)
8. Park, J., Moon, M., Hwang, S., Yeom, K.: CASS: A Context-Aware Simulation System for Smart Home. In: Proceedings of the 5th ACIS International Conference on Software Engineering Research, Management & Applications, pp. 461–467. IEEE Computer Society (2007)
9. Roberts, C.A., Dessouky, Y.M.: An Overview of Object-Oriented Simulation. Simulation 70, 359–368 (1998)
10. Ruge, L., Kindsmüller, M.C., Cassens, J., Herczeg, M.: How About a MATe for Awareness in Teams? In: Proceedings of the Seventh International Workshop on Modelling and Reasoning in Context, pp. 58–69 (2011)
11. Schmitt, F., Cassens, J., Kindsmüller, M.C., Herczeg, M.: Mental Models of Ambient Systems: A Modular Research Framework. In: Beigl, M., Christiansen, H., Roth-Berghofer, T.R., Kofod-Petersen, A., Coventry, K.R., Schmidtke, H.R. (eds.) CONTEXT 2011. LNCS, vol. 6967, pp. 278–291. Springer, Heidelberg (2011)

A Conceptual Framework for Supporting Adaptive Personalized Help-on-Demand Services

William Burns[1,*], Liming Chen[1], Chris Nugent[1], Mark Donnelly[1], Kerry-Louise Skillen[1], and Ivar Solheim[2]

[1] Computer Science Research Institute and School of Computing and Mathematics, University of Ulster, United Kingdom
{wp.burns,l.chen,cd.nugent,mp.donnelly}@ulster.ac.uk, skillen-kl@email.ulster.ac.uk
[2] Norwegian Computing Center, Oslo, Norway
Ivar.Solheim@nr.no

Abstract. Mobile applications that encompass personalization and context-aware components are increasingly becoming more prevalent. The ability to offer personalized content and User Interfaces to the users of these applications, however, has still not been fully addressed. In this paper we describe a conceptual framework that establishes a User Profile and aims to monitor the usage patterns of users of a mobile application and, based on these patterns, provide both personalized, context aware content and user interfaces. The framework consists of four components that together contribute towards an overall Help on Demand service that is targeted at older age Smartphone users. A usage scenario is presented to describe the typical usage of the help on demand service.

Keywords: context-aware services, adaptive personalization, smartphone applications.

1 Introduction

With the ever-increasing trend of higher life expectancies coupled with falling birth rates is resulting in a growing ageing society [1]. This ageing society will inevitably have an increasing burden on the social and health services. As a direct result of this, an increasing focus is now being placed on using technology as one possible means to alleviate these burdens. Nevertheless, currently available off-the-shelf technologies do not cater specifically for the needs of the older population, specifically from a personalization perspective.

It is important that the solutions offered by such 'assistive' technologies be personalized to the user's physical, cognitive and age related disabilities. If the end-user cannot use the solution, it is less likely to have a positive effect on their Quality of Life [2].

* Corresponding author.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 427–432, 2012.

In the following Sections we present related work in the area of mobile phone based assistive technologies and the personalization of services. We then present the proposed conceptual adaptation framework that aims to provide help on demand and context-aware support to older, smartphone users [3]. Then a use case scenario is presented detailing an example use of the HoD application and how the proposed framework is used to provide aid this service.

2 Related Work

There is an increasing trend in using mobile technologies in the area of Ambient Assisted Living [4]. In particular, mobile phones, smartphones and tablet computers have witnessed the most uptake. Using these technologies allows the end user to take the assistive technology solution with them wherever they go. The mobility of such devices and the hardware that is built into them such as accelerometers, gyroscopes and GPS allow the opportunity to infer a user's context regarding their current situation.

A number of studies report on the development and evaluation of mobile technology based solutions. The iWander [5] project, for example, aimed to provide a Person with Dementia (PwD) with navigational assistance using a smartphone's in-built GPS hardware. Another example includes the use of a dedicated phone that provides scheduled video reminders to a PwD patient [6]. With this solution family members record video reminders in order to create the notion of a recognizable virtual caregiver.

Peripheral devices can be connected wirelessly to the smartphone in order to provide additional information. One example of a previously developed solution used a smartphone to connect to a number of wireless sensors in order to record the user's heart and respiration rates and activity levels and relay that data to a healthcare professional [7].

To date, few solutions have attempted to offer truly personalized services where the functionality offered attempts to learn from the user's behavior to provide contextual information and services.

3 Conceptual Framework

The MobileSage Project [3] aims to provide older people with context-sensitive, personalised and location-sensitive tools which allow them to carry out and solve everyday tasks and problems, when and were they occur, 'just-in-time'. There are two main components of this project, a mobile application known as Help-on-Demand (HoD) and the Content Management System (CMS). The HoD enables the user to access content on demand that helps them accomplish everyday tasks. The content provided to the user come from the CMS in the form requested based on the User's Profile.

In this paper, we propose a conceptual adaptation framework that aims to provide personalized and context aware [8] content through an adaptable user interface (UI). This application will allow older users to interact with various public devices for example a train station ticket dispenser, and to provide them with personalized information. Other services provided include the ability for navigational help using the device's GPS hardware and personalized content from the CMS. The adaptation framework proposed in this paper will become a core component of the HoD application. The HoD application runs on an NFC (Near Field Communications) enabled device (Samsung Galaxy Nexus), In the following Section, we outline the proposed framework and describe how it will integrate into the HoD application by way of a usage scenario.

Figure 1 presents the proposed framework services and the sub components in the context of the HoD application.

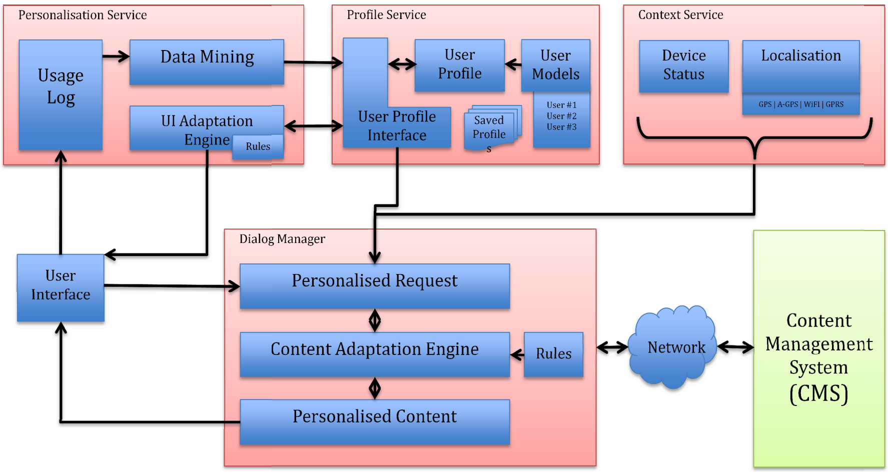


Fig. 1. Flow of information within the Help on Demand application using the proposed adaptation framework. Four core services exist to provide information to the user and to personalise requests to the Content Management System (CMS).

4 Usage Scenario

The following case study will outline a real-world application for the HoD application and the proposed adaptation framework.

Jane is a 65-year-old woman with diminished eyesight who likes to use public transport to travel into town everyday. Jane is not afraid of technology, however, does not want to have to learn a whole new skill set in order to use it. She has a smartphone with the HoD application installed. After reaching the local train station she wishes to buy a ticket. Jane finds it difficult to read the instructions on the ticket kiosk. She holds up her smartphone to an NFC tag on the kiosk, and the HoD application automatically requests personalized content from the CMS. This content consists of personalized instructions on how to use the kiosk, presented in Jane’s preferred feedback mode, as stored in her User Profile. In this instance Jane is ‘Spoken’ through the process of buying a ticket via her mobile phone.

This interaction, location and service usage is recorded in the Usage Log on the smartphone. In subsequent visits to the local train station, the HoD application infers

Jane's context as being in the train station and automatically asks her if she would like to interact with the ticket kiosk, without needing to scan the kiosk's NFC tag. If it transpires that Jane primarily uses the NFC service most, the proposed framework will mine the Usage Log to ascertain this fact and update the UI to make this service button more prominent.

Within the aforementioned scenario, the proposed framework will use its Personalization service to the monitor Jane's usage patterns in order to provide her with personalized content and services. Jane's interactions with the HoD will be stored in the Usage Log component and mined from time to time to ascertain specific patterns, Based on a number of rules, the adaptation engine will update the User Profile in addition to the UI.

5 Framework Components

The User Profile that will be used and updated has been generated during the initialization phase of the HoD, and will be based on one of three User Models, *Technophobic*, *Techno-Static* and *Technophile*. User models will be generated from user personas, generated based on a number of focus groups carried out in three European countries, and will outline the parameters contained within the User Profile based on the type of user. Each of these User Models will generate User Profiles with differing levels of complexity. Technophobic should offer core functionality and require as little cognitive load and user interaction as possible. The Technophile model will offer the greatest complexity of services and options and could be considered an Advanced User Profile. Techno-Static would be an amalgamation of the previous two. The model selected is based on Jane's experience with technology, which will act as a blueprint for the parameters, depicted in figure 2, contained in the User Profile, which is used to personalize the HoD.

```
UserID
FeedbackMode
   - Text
   - Audio
   - Video
   - Video with Closed
   Captions
Locale
   - Language
   - Location
UI
   - Font
   - TextSize
   - Brightness
EmergencyContact
   - Name
   - Number
PositioningModules
   - WiFi
   - GPS
   - A-GPS
```

Fig. 2. An example of the parameters contained within the proposed framework's User Profile component. This will be used to store the user UI and feedback preferences and can be updated based on the user's usage patterns stored and mined from the Usage Log.

The Context Service will provide the core information used to infer the user's and device's context. There are two main components of this service, the first is the Device Status. This provides the system with on-demand information pertaining to the status of the user's device. Information such as the battery level, signal strength and connection speed can be passed to the Dialog Manager to increase the awareness of the user's context.

The second component of the Context Service is Location. Taken from the device's hardware, the location of the user can be ascertained, using GPS, A-GPS or WiFi in addition to language information established in the initialization phase and stored in the User Profile.

The Dialog Manager is the component of the system that handles all requests and responses sent to and from the CMS. Within the Dialog Manager there are three components in which information is passed through in order to personalize the content, these are the Personalized Request, Content Adaptation Engine and Personalized Content.

The Personalized Request component modifies the user request in order to save on network bandwidth in addition to requesting a specific type of content. This should enable the CMS to return the desired content with minimal computational logic required on the server side. This component takes the user's contextual information (Device and Location) along with the preferred feedback modes preferred as defined within the UP.

When the *Dialog Manager* is personalizing the user's request, this component is used to add additional rules to the personalization. These rules take into consideration contexts such as the network speed and battery levels. For example, if the user requests content in the form of a video, however, the device's battery level is below 20%, The request is then modified to request text based responses. In doing so, the battery level of the device is not reduced further by video playback. When the response is received from the CMS into the *Dialog Manager*, its content is adapted to ensure that it matches the user's desired preferences.

This *Personalized Content* component takes the adapted response from the CMS and passes it to the UI of the HoD application.

The roles of each of the adaptation framework's components have been described from the perspective of a typical HoD usage scenario. The framework in its current form, is designed to be scalable to, in the future, to include further components and parameters that will allow the system to provide increased situation and context awareness to the end user.

6 Conclusion

In this paper we have outlined the conceptual adaptation framework that allows an assistive technology to evolve and adapt to the usage patterns of the user. We have identified and described the major components of the framework that allows the creation, adaptation and evolution of a User Profile and UI. Future work will involve the development and evaluation of a prototype application with this presented adaptation framework integrated into the core functionality.

Acknowledgments. This work is supported by the EU AAL MobileSage project. The authors gratefully acknowledge the contributions from all members of the MobileSage consortium.

References

1. Hogan, M.: Physical and Cognitive Activity and Exercise for Older Adults: A Review. J. Ageing and Human Development 60(2), 95–126 (2005)
2. Doukas, C., Metsis, V., Becker, E., Le, Z., Makedon, F., Maglogiannis, I.,: Digital cities of the future: Extending @home assistivetechnologies for the elderly and the disabled. Telematics and Informatics 28(3), 176–190 (2011)
3. MobileSage Group AAL. MobileSage – Situated Adaptive Guidance for the Mobile Elderly, `http://www.mobilesage.eu` (Accessed July 2012)
4. Boulos, M.N., Wheeler, S., Tavares, C., et al.: How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed. Eng. Online 10(24) (2011)
5. Miskelly, F.: Electronic tracking of patients with dementia and wandering using mobile phone technology. Age and Ageing 34, 497–518 (2005)
6. Donnelly, M., Nugent, C., McClean, S., et al.: A Mobile Multimedia Technology to Aid Those with Alzheimer's Disease. IEEE Multimedia 17(2), 42–51 (2010)
7. Boulos, M.N.K., SWheeler, S.,, Tavares, C., Jones, R.: How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. BioMedical Engineering Online 10(24) (2011)
8. Dey, A.K.: Understanding and using context. Personal and Ubiquitous Computing 5(1), 4–7 (2001)

Developing Touchless Interfaces with GestIT

Lucio Davide Spano

ISTI-CNR Via G. Moruzzi 1, 56127, Pisa, Italy
lucio.davide.spano@isti.cnr.it

Abstract. In this paper, we report on the development of touchless interfaces for supporting long lasting tasks, which need an interleaving between the interaction with the system and the focus on other activities. As an example, we considered a dish cooking task, which enables selecting and browsing the information about different recipes while cooking through gestural and vocal interaction. The application demonstrates the advantages offered by the GestIT library, which allows a declarative and compositional definition of reusable gestures.

Keywords: Touchless interfaces, Input and Interaction Technologies, Gestural Interaction, Ambient Intelligence.

1 Introduction

The availability of new interaction devices, which are able to track the position of body joints, opens the possibility to create smart environments that provide the users with enhanced interaction capabilities. Indeed, such hardware is useful when users are performing tasks that do not allow the use of traditional pointing devices or keyboards. For instance, the primary user's task may be the creation of an artefact in the real world, which requires several steps to be completed, like assembling furniture or replacing a part of an appliance. An interactive support that enables the user to browse the information while performing the primary task can be really effective in such situations. For this reason, touchless interaction is gaining the attention of the research community, in particular after the creation of devices that enable it at the industrial level, mainly for the gaming market. Its advantages and its risks have been analyzed in [1], where the authors concluded that a touchless direct manipulation is well accepted by the users, but designers should be careful while choosing the vocabulary, which has to be immediately understandable for them. In literature it is possible to find examples of touchless interfaces for specific appliances in the kitchen environment [2], or for getting a full control of different devices [3], but problems such as gesture reuse or how to distinguish movement aimed to interact with the system from those that are not (the well-known Midas Touch) are still open.

In this paper we consider the kitchen environment as an example for such kind of applications. The case-study scenario envisions the assistance during the dish preparation through a set of information displayed on a screen, which can be browsed by the cooker while touching the food or using kitchen tools. In such situation, the touchless

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 433–438, 2012.

interaction has the advantage of avoiding the contact with the input devices, which can create hygiene problems or the risk of damaging the electronic equipment (e.g. touching it with wet hands).

The application development is based on the GestIT library, which offers a declarative and compositional approach for describing and managing gestural interaction. With this library it is possible to define and reuse high level gestures, composing them starting from basic building blocks through composition operators and assigning handlers at the desired level of granularity.

2 GestIT

The GestIT library has been extensively described in [5], here we summarize the main library features, in order to better describe the implementation of the case-study application. The library supports the definition of gestures according to a specific meta-model, which enables the developers to define high level gestures while maintaining the possibility to decompose them in smaller parts, and to assign handlers to their sub-components. The meta-model is abstract with respect to the recognition platform, thus it can be applied for describing gestures that are recognized by very different devices, such as touch-screens for multi-touch interaction or the Microsoft Kinect for full-body gestures. The library is distributed as an open-source project[1], and it has been implemented in different programming languages (Java, C#, Objective C) and for different OS (Windows, iOS, Android).

The meta-model allows the definition of a complex gesture starting from though *ground terms* and *composition operators.*

The ground terms represent features that can be tracked by developers for recognizing gestures. For instance, in a multi-touch application they are the events that allow tracking the finger positions (usually called touch start, move and end), while for full body gesture they are the events related to joint positions. Ground terms can be optionally associated to a predicate that has to be verified in order to receive the notification of a feature change. For instance, if we consider the movement of a body joint as a feature, it is possible to specify a predicate that computes whether the movement is linear or not. We denote a specific feature change assigning a name to it and optionally specifying the predicate name in square brackets. For instance, $Hl[linear]$ denotes the left hand feature (*Hl*) in which the *linear* predicate checks if the position change is linear or not.

The composition operators allow the connection of both ground terms or composed gestures in order to obtain a complex gesture definition. The operators are a subset of those defined in CTT [4] and their semantics can be summarized as follows:

- *Iterative Operator*, represented by the $*$ symbol, expresses the repetition of a gesture recognition an indefinite number of times.
- *Sequence Operator*, represented by the $\gg$ symbol, expresses that the connected sub-gestures (two or more) have to be performed in sequence, from left to right.

[1] The GestIT library is available at `http://gestit.codeplex.com`

- *Parallel Operator*, represented by the || symbol, expresses that the connected sub-gestures (two or more) can be recognized at the same time.
- *Choice Operator*, represented by the [] symbol, expresses that it is possible to select one among the connected components in order to recognize the whole gesture.
- *Disabling Operator*, represented by the [> symbol, expresses that a gesture stops the recognition of another one, typically used for stopping iteration loops.
- *Order Independence*, represented by the |=| symbol, expresses that the connected sub-gestures can be performed in any order.

The composition result is an expression that defines the temporal relationships among the various low-level device events that are involved in the gesture recognition. Event handlers can be attached to all the expression terms (either ground or complex) and they are separated from the gesture description itself, which is reusable in different graphic controls and for different behaviours. In the following section we provide some examples of such gesture descriptions.

3 Touchless Recipe Browser

For the development of the touchless user interface, we considered a scenario in which the user wants to cook a dish, but s/he does not really master the particular procedure. Therefore, s/he needs a description of the steps to be accomplished in order to complete the preparation, which is usually provided through books or specialised magazines. We try to enhance such experience with an interactive support for delivering the information: the steps are described by the interactive system through a combination of text and video. In order to browse the recipes, the user does not need to touch any particular input device, which has the advantage of supporting the interaction while the cooker is manipulating tools or s/he has dirty hands. Instead s/he controls the application through a multimodal combination of voice and gestures. In order to enable such kind of interaction, we exploited a Microsoft Kinect, together with a computer screen or TV that displays the user interface. The touchless recipe browser supports two tasks: the first one is the recipe selection, while the second one is the presentation of the cooking step. The selection of the recipe consists of two screens: the first one for selecting the recipe category (starter, first course, second course, dessert etc.) and then the selection of the recipe itself.

The presentation of the cooking steps is performed through a combination of text and video. The user can watch the entire video with subtitles that show how to cook the selected dish or s/he can browse back and forth among the different steps with a previous and next function or controlling a timeline.

In order to combine the vocal and the gestural modality, we extended the GestIT library adding the possibility to react to vocal input, representing the different keywords that activate vocal commands as features that can be detected by the Kinect support. Therefore, it is possible to combine in the gesture description expression also vocal inputs.

With respect to the design of the user interface, we decided to assign commands that do not need any argument (e.g. going back to the previous screen) to the vocal modality, while we assigned commands related to object selection and/or manipulation to the gesture modality. The rationale behind this choice is trying to keep the user's focus on his main task (cooking the dish) as much as possible: gestures have an higher cognitive load with respect to speech interaction.

In addition, the design of such kind of user interface must take into account the well-known Midas Touch problem. We exploited the possibility to define the temporal relationships between gestures provided by GestIT in order to mitigate it. Indeed, we chose to enable the interaction with the user interface only if the user stands in front of the screen, while we do not consider any movement or interaction otherwise. The rationale behind this design choice is that, being the dish cooking the main task, we assume that most of the times the user do not want to interact with the application. If the interaction is needed, the user will look at the screen, positioning in front of it. Using the GestIT library, the interaction with the different application presentation follows the schema defined in equation 1. The *Front* gesture enables the *ScreenInteraction*, which represent the allowed gestures or vocal commands for the considered presentation, and it is disabled by the *NotFront* gesture. Such expression term is refined in different ways according to the considered presentation. As it is possible to observe in equation 1, *Front* and *NotFront* are symmetric: they respectively check whether the shoulder position (S_l and S_r) are parallel with respect to the sensor (and screen) plane (the p predicate) or not. This means that as long as the user stays in front of the screen, it is possible to interact with the application. The *Front* and *NotFront* gestures have handlers that provides the user with feedback for signalling whether the application is ready to receive inputs (a green "Tracking" label) or not (a red "Not Tracking" label).

$$\begin{aligned} &Front \gg ScreenInteraction^{*} [> NotFront \\ &Front = \ (S_l[p] \mid\mid S_r[p]) \\ &NotFront = \ (S_l[!\,p] \mid\mid S_r[!\,p]) \end{aligned} \qquad (1)$$

Fig. 1. Recipe category selection

Figure 1 shows the presentation for selecting the recipe category. For the sake of brevity, we do not show here the presentation for selecting the particular recipe, because it is similar to the one in Figure 1 from an interaction point of view. The user can focus on specific category moving his/her hand until the desired icon is magnified with a fisheye effect. After that s/he can select the category closing the hand. As already discussed before, we assigned the commands without arguments to the vocal modality: in this screen it is possible to use the following commands: *back* for going back to the previous screen and *exit* for closing the application. Equation 2 shows the definition of the *ScreenInteraction* definition for the selection presentation (we describe movements only for the right hand for simplicity, but the actual implementation provide a symmetric support also for the left hand). The features marked with *V[word]* are those related to the voice and indicate the pronunciation of the specified word, with the obvious effect on the user interface (respectively going back to the previous screen or closing the application). The *Grab* gesture is used for selecting the recipe category and it is composed by an iterative hand movement (mH_r^*) disabled by a closure of the hand (cH_r). As already explained in section 2, it is possible to attach event handlers not only to the whole gesture completion (which performs the category selection and therefore changes the screen), but also to its sub-parts. In this case the fisheye effect in Figure 1 is driven by an event handler attached to the completion of the hand movement (mH_r^*).

$$\begin{aligned} &ScreenInteraction = V[back][\,]V[exit][\,]Grab \\ &Grab = mH_r^* \;[>\; cH_r \end{aligned} \quad (2)$$

Fig. 2. The dish preparation screen

Figure 2 shows the screen for the preparation of a dish. In the upper part it is possible to read the recipe name, in the centre there is a video tutorial for the preparation[2] together with a text describing the procedure to follow in order to complete the current step. In the lower part a slider represents the video timeline. The interaction for this presentation is defined in Equation 3.The vocal commands *back* and *exit* are still available in this screen. The video playback can be *continuous* or it can stop at each *step*. A vocal command is available for activating both modalities. It is possible to pronounce the words *next* and *previous* respectively to show the previous or the next

[2] The sample recipes included with the application prototype have been created using some videos from the public website of the Italian cooking TV show "I Menu di Benedetta" (`http://www.la7.it/imenudibenedetta/`)

step of the preparation. Such command can be activated also through the *Swipe* gesture, an iterative linear hand movement performed at a certain speed (verified by the properties *linear* and *speed*), disabled by a hand movement that does not have this characteristics (for finishing the iteration loop). If the swipe movement has been performed from left to right, the tutorial proceeds to the next step, if it has been from right to left the tutorial goes back to the previous step. Finally, it is possible to control the video timeline through the *Drag* gesture. The latter is a composition of two sub gestures: the first one is *Grab* (already defined in Equation 1) and the second one is *Release*, which is an iteration of hand movement disabled by its opening (oH_r). Through such gesture description is possible to notice the reuse possibility offered by the library (we defined the *Grab* gesture and reused it for the *Drag* one). Different handlers have been assigned to the different gesture sub-parts, which allow to define easily the user interface reactions while performing the gesture: when the user closes the hand (completion of cH_r in the *Grab* gesture), the user interface changes the colour of the slider knob, after that its position is changed according to the hand movement direction together with the displayed video frame (completion of $.mH_r^*$ in the *Release* gesture) and finally, when the whole gesture is completed, the video playback restarts from the point selected by the user.

$$\begin{aligned} ScreenInteraction &= V[back][\,]V[exit][\,]V[previous][\,] \\ &\quad V[next][\,]V[continuos][\,]V[step] \\ &\quad Swipe[\,]Drag \\ Swipe &= mH_r^*[linear \wedge speed][> mH \\ Drag &= Grab \gg Release \qquad Release = mH_r^*[> oH_r \end{aligned} \tag{3}$$

4 Conclusions and Future Work

In this paper we described the development of a touchless support for providing information to the user while cooking, enhancing the kitchen environment through an interface based on the GestIT library. In the future we will perform an evaluation of the proposed user interface, we will apply the described techniques to other smart environments and to the management of multiple users.

References

1. de la Barré, R., Chojecki, P., Leiner, U., Mühlbach, L., Ruschin, D.: Touchless Interaction-Novel Chances and Challenges. In: Jacko, J.A. (ed.) HCI International 2009, Part II. LNCS, vol. 5611, pp. 161–169. Springer, Heidelberg (2009)
2. Garzotto, F., Valoriani, M.: “Don’t touch the oven”: motion based touchless interaction with household appliances. In: AVI 2012, Capri, Italy, pp. 721–724. ACM (May 2012)
3. Palanger, G.: Kinect in the kitchen: testing depth camera interactions in practical home environments. In: CHI 2012, Austin, Texas, USA, pp. 1985–1990. ACM (May 2012)
4. Paternò, F.: Model-based design and evaluation of interactive applications. Applied Computing (2000)
5. Spano, L.D., Cisternino, A., Paternò, F.: A Compositional Model for Gesture Definition. In: Winckler, M., Forbrig, P., Bernhaupt, R. (eds.) HCSE 2012. LNCS, vol. 7623, pp. 34–52. Springer, Heidelberg (2012)

Tool Support for Probabilistic Intention Recognition Using Plan Synthesis

Frank Krüger, Kristina Yordanova, and Thomas Kirste

University of Rostock
Institute of Computer Science
MMIS Group
Albert-Einstein-Str. 22
18059 Rostock
Germany
{frank.krueger2,kristina.yordanova,thomas.kirste}@uni-rostock.de
http://mmis.informatik.uni-rostock.de

Abstract. To provide assistance in intelligent environments it is necessary to accurately infer the users needs and wishes. In this demonstration we present a probabilistic plan recognition system that is able to track the user and to compare different hypotheses about the users behavior and her goal(s) based on observations of the current activity. Furthermore, the tool provides a probability distribution over the possible goals and selects the most probable hypothesis as the user intention.

1 Introduction

As intelligent environments, such as smart rooms, become more and more complex, their users have to be assisted. To provide such assistance in intelligent environments it is necessary to infer the users needs and wishes. One approach to recognize humans behavior and, based on the intention – that is the underlying goal of her behavior – is plan recognition. The activity of a (group of) user(s) is recognized by assuming a goal "driven" behavior when attempting to reach a given goal. Based on this assumption, plan recognition analyzes whether these activities are prefixes of a plan that is known to lead to a known goal. In the context of intelligent environments, which contain sensors producing noisy and ambiguous sensor observations of users, such plan recognition approach will require the usage of probabilistic plan recognition. Bayesian Filters can be used here to cope with such uncertainties [6].

Plan recognition is based on explicit representations of valid action sequences, or plans, where prior knowledge about interdependencies of actions and the environment is encoded in these plans. However, setting up libraries of all valid plans can be a tedious and time consuming task task [5]. Works such as [3,4] show, that action libraries can be used to automatically set up such plan libraries which are later used for activity recognition.

In this work we go a step further and present how such generated libraries can be used for intention recognition by comparing the different plans based on sensor observations, and choosing the one with the highest likelihood.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 439–444, 2012.

The paper is structured as follows. After giving a short introduction to setting up action libraries for human behavior models, we show how the intention of users can be inferred. Later we present a tool that provides a probability distribution of user goals which can be used for further assistance. Based on this probability distribution we show in the last section how assistance could be provided.

2 Probabilistic Plan Recognition

Plan recognition approaches require all valid plans to be described. This can either be done by explicitly defining them, or by setting up action libraries, where actions are represented in terms of preconditions and effects. In this work the description of the action templates is done with the Computational Causal Behavior Models (CCBM) where preconditions and effects are used to specify the activities using a PDDL[1]-like notation [2]. More precisely an action definition consists of multiple elements, example of which can be seen in Figure 1 where the abstract action *goto* is described. In the example, the following slots can be recognized.

```
(:action goto
 :parameters (?a - user ?x ?y - glocation)
 :agent ?a
 :salience 1.0
 :duration (normal (distance ?x ?y) 2)
 :precondition (and (not (= ?x ?y))
                    (at ?x ?a)
                    (not (moving ?a ?x ?y))
                    (not (busy ?a)))
 :effect (and (moving ?a ?x ?y)
              (not (at ?x ?a)))
 :observation (and (setAction Action_goto)
                   (setStart ?x)
                   (setEnd ?y))
)
```

Fig. 1. CCBM specification of the action template *goto*

:parameters specifies the possible parameters for the operator. The action will later be grounded by applying all possible combinations of parameters to the operator. This element of the action is also typical for the standard PDDL notation.[2]

:agent determines the execution slot for the operator. Typically, the specific agent is given as parameter, which results in one action for each agent, that will be executed in the corresponding execution slot.

[1] Planning Domain Definition Language

[2] The elements *parameters*, *precondition* and *effects* are also present in the standard PDDL notation. The standard PDDL notation also has a *duration* slot, however it is not able to represent probabilistic durations

:salience provides a prior weight for the action template. Each grounded action that was created from this template is assigned the same saliency value.

:duration allows the specification of probabilistic action durations by using a probability density function.

:precondition lists all preconditions to the world state that have to be met in order the action to be applied. (PDDL element)

:effect lists all effects of the action to the world state. Effects will be applied immediately after action selection. (PDDL element)

:observation lists external functions that were implemented in the observation model and will be called if optional conditions to the world state are met.

```
(define (problem office)
 (:domain abc)
 (:objects a b c - job
           alice - user
           p1 - printer)
 (:init
   (forall (?a - user) (and (handsfree ?a) (at outside ?a)))
   (printerjammed ?p1))
 (:goal (and (forall (?j - job) (printed ?j))
             (forall (?a - user) (and (at outside ?a)
               (holds coffee ?a)))))
)
```

Fig. 2. CCBM specification of the initial and goal state

The set of all actions together with the objects type description and the predicates declaration is called domain description. On the other hand, the specification of the initial and the goal state together with the concrete actions durations and other problem-specific information are called problem specification which is shown in Figure 2. Both the domain and problem descriptions are compiled into a Bayesian filter, which is then used to infer the activity of the user. If either the goal, or the initial state are uncertain, multiple instances of the Bayesian filter, each of them compiled from the same domain but with different problem specifications have to be started. For further information about the inference process we refer the interested reader to [2]. A discussion of modeling strategies and its impact on the recognition rate can be found in [7].

3 Tool Support for Probabilistic Intention Recognition

The probabilistic plan recognition tool compiles each domain – problem combination to a single Particle filter (PF) and starts each PF the specified number of times. Whenever new sensor data arrives from the sensors, or in case of the offline mode for each sensor data entry in the file, every running instance of the Particle filter uses this same data. During the correction step the Particle filter uses the observations to weight each single hypothesis respectively. A likelihood is then computed from all hypotheses for each

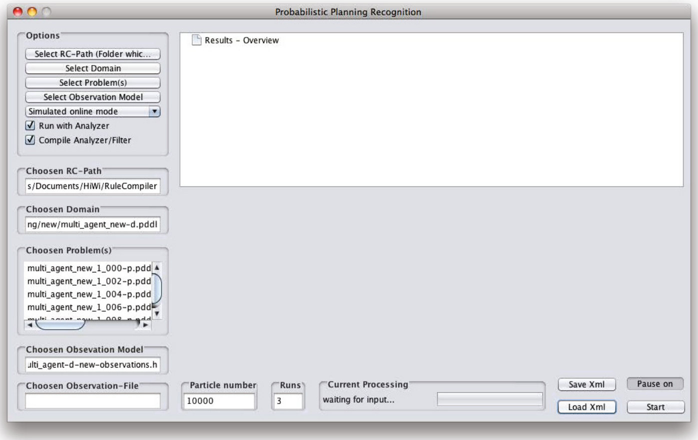


Fig. 3. The demo tool for intention analysis with probabilistic plan recognition

instance and the median from the different instances resulting from the same domain – problem combination is then chosen. These values (one for each problem – domain combination) are used to calculate a categorical distribution over all different problems by using the following strategy. Given an observation y and K mutually exclusive models $M_{1:K}$, the probability that y has been produced by M_k is:

$$p(M_k \mid y) = \frac{p(y \mid M_k)\, p(M_k)}{p(y)} = \frac{p(y \mid M_k)\, p(M_k)}{\sum_{i=1}^{K} p(y \mid M_i)\, p(M_i)} \tag{1}$$

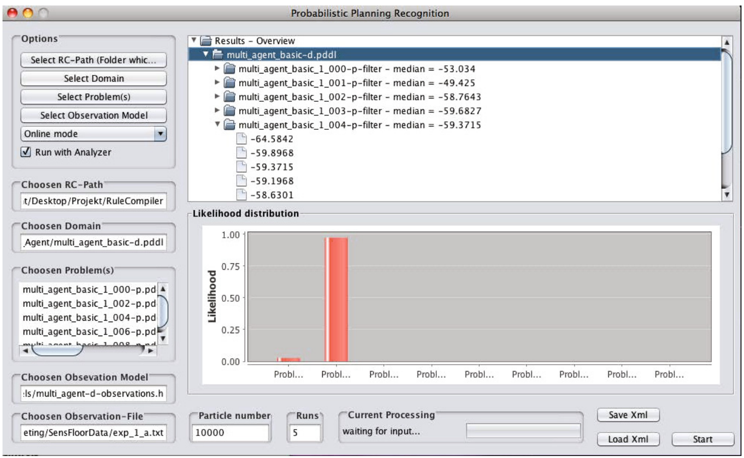


Fig. 4. The graphical illustration of the distribution of the likelihood for the different problems

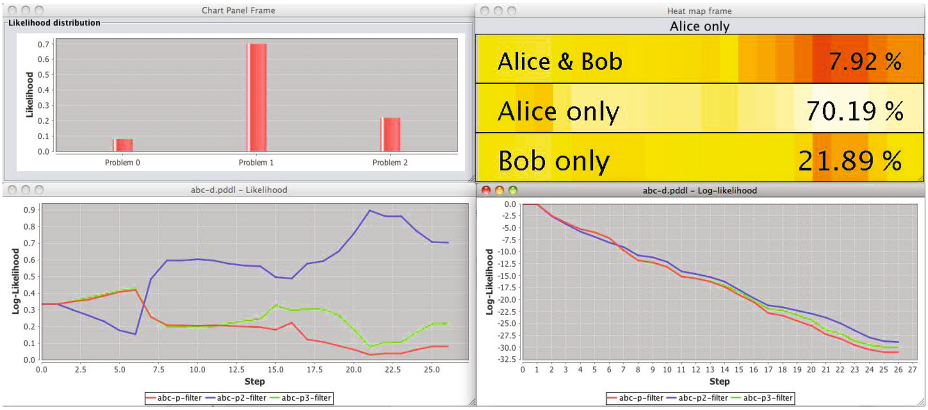


Fig. 5. The temporal course of the probability distribution of different goals of the user

assuming the uniform priors $p(M_k) = K^{-1}$ for all

$$p(M_k \,|\, y) = \frac{p(y \,|\, M_k)}{\sum_{i=1}^{K} p(y \,|\, M_i)} \tag{2}$$

By applying this formula to the likelihoods of all models in each time step this results in a categorical distribution that can be further used.

The graphical user interface for the intention recognition tool is illustrated in Figure 3. As can be seen, the tool requires the user to select one domain specification file, multiple problem files and an observation model, that provides the interface to the sensor data. Two different modes of sensor data usage are supported; an online mode, that reads the data directly from the sensor and infers the intention in realtime and an offline mode, that allows to handle pre-recorded sensor data. In addition the user of the tool can decide whether a goal based heuristic should be used and how many runs of each problem should be executed with the specified number of particles. The specification of all parameters can be saved and later loaded.

A graphical illustration of the log-likelihoods from the different problems as well as the calculated categorical distribution is presented to the user as can be seen in Figure 4. The tool also provides the illustration of the course of distribution of the probabilities of the single problems. Figure 5 shows this course for three different specifications.

4 Application Domains

Intelligent environments consists of a vast variety of application domains and our intention recognition tool could benefit many of them. One such application is user assistance where the system proactively assist her in achieving the user goal. This can be done by using the categorical distribution calculated from the intention recognition. One option would be for the assistive system to take the most likely goal and support the user stepwise until the goal is reached. Another assistance application could be finding an optimal solution where the probability distribution can be used together with cost functions for assistance. A more detailed description of integrating the intention recognition approach into assistance is given in [1].

5 Conclusion

In this work we presented a tool support for intention recognition based on Computational Causal Behavior Models. The models are compiled into a probabilistic inference machine that is not only able to recognize the current user state, but also to follow different supported goals and calculate their likelihood. By comparing the likelihood of different models (combinations of initial and goal state), a probability distribution of possible user intentions is provided. The tool allows probabilistic reasoning about different user goals and execution paths, thus in the future could be used as the link between the activity recognition process described in [3] and proactive assistive systems.

Acknowledgements. We would like to thank André Luthardt, Daniel Koch and Frank Breuel for their implementation support.

References

1. Kirste, T.: Making use of intentions. Technical Report CS-01-11, Institut für Informatik, Universität Rostock, Rostock, Germany (March 2011) ISSN 944-5900
2. Kirste, T., Krüger, F.: Ccbm-a tool for activity recognition using computational causal behavior models. Technical Report CS-01-12, Institut für Informatik, Universität Rostock, Rostock, Germany (May 2012) ISSN 0944-5900
3. Krüger, F., Yordanova, K., Burghardt, C., Kirste, T.: Towards creating assistive software by employing human behavior models. Journal of Ambient Intelligence and Smart Environments 4(3), 209–226 (2012)
4. Ramirez, M., Geffner, H.: Goal Recognition over POMDPs: Inferring the Intention of a POMDP Agent. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence, IJCAI 2011 (2011)
5. Roy, P.C., Giroux, S., Bouchard, B., Bouzouane, A., Phua, C., Tolstikov, A., Biswas, J.: A Possibilistic Approach for Activity Recognition in Smart Homes for Cognitive Assistance to Alzheimer's Patients. In: Chen, L., Nugent, C.D., Biswas, J., Hoey, J., Khalil, I. (eds.) Activity Recognition in Pervasive Intelligent Environments. Atlantis Ambient and Pervasive Intelligence, vol. 4. Atlantis Press (2011)
6. Saria, S., Mahadevan, S.: Probabilistic plan recognition in multiagent systems. In: Proceedings of the International Conference on Automated Planning and Scheduling (2004)
7. Yordanova, K., Krüger, F., Kirste, T.: Context aware approach for activity recognition basedon precondition-effect rules. In: 9th IEEE Workshop on Context Modeling and Reasoning (CoMoRea 2012) at the 10th IEEE International Conference on Pervasive Computing and Communication (PerCom 2012), Lugano, Switzerland (2012)

Aesthetic Intelligence: The Role of Design in Ambient Intelligence

Carsten Röcker[1], Kai Kasugai[1], Daniela Plewe[2], Takashi Kiriyama[3], and Artur Lugmayr[4]

[1] Human-Computer Interaction Center, RWTH Aachen University, Germany
{Kasugai,Roecker}@humtec.rwth-aachen.de
[2] University Scholars Programme, National University of Singapore, Singapore
DanielaPlewe@nus.edu.sg
[3] Graduate School of Film and New Media, Tokyo University of the Arts, Japan
kiriyama@gsfnm.jp
[4] Entertainment and Media Management Lab, Tampere University of Technology, Finland
artur.lugmayr@tut.fi

Abstract. This paper illustrates the rationale behind the second international workshop on *Aesthetic Intelligence*. The workshop addresses the multiple facets of aesthetics in the design process of Ambient Intelligence technologies, especially in the fields of architecture, industrial and interface design as well as human-computer interaction.

Keywords: Ambient Intelligence, Ubiquitous Computing, Smart Spaces, Aesthetics, Design, Architecture, Urban Informatics.

1 Introduction

Ambient Intelligence refers to the integration of information, communication and sensing technologies into architectural spaces and thereby offers the technical basis for providing context-adapted services and assistance in everyday activities [1]. Over the last years, research in the field of Ambient Intelligence came to a point where many technical challenges have been addressed and most fundamental problems have been solved. First commercial Ambient Intelligence applications are already available and more sophisticated systems are likely to follow in the coming years, which will gradually transform our everyday environments into smart and attentive surroundings [4,5]. With the widespread integration of technology into living spaces, aspects of aesthetically pleasing design gain increased importance. Ambient Intelligence systems do not only have to meet technical requirements, but also have to blend into existing environments. Therefore, it is important to bring together research from relevant disciplines and offer them a platform for discussing the relevance of aesthetic values for Ambient Intelligence as well as the role of aesthetically pleasing design for usability, technology acceptance and user well-being.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 445–446, 2012.

2 Research Challenges and Workshop Topics

Previous work in the area of Ambient Intelligence mainly focused on technical aspects and general questions of user interaction in technology-enhanced environments. While those are important aspects, it seems to be time to extend ongoing research activities and also include hedonic and aesthetic dimensions of design and usage. Hence, this workshop explicitly aims at bringing together researchers from adjunct disciplines to discuss the interrelation of functional, architectural, and aesthetic factors and their consequences for the design, use and acceptance of Ambient Intelligence technologies.

This workshop builds on the results and insights gained during the first international workshop on Aesthetic Intelligences [2,3] held in Amsterdam, the Netherlands, on November 16, 2011. The upcoming workshop particularly addresses research challenges originating in the fields of architecture, design and human-computer interaction. In the area of architecture, research questions may include topics like smart buildings and spaces, intelligent and multimedia façades, urban screens and large public displays, innovative building and interaction materials, or novel living concepts. Design-related research challenges may address questions of user experience design, participatory design, emotional and hedonic design, interaction design, multi- and interdisciplinary design, as well as tools and design techniques for Ambient Intelligence environments. Relevant research in the field human-computer interaction includes human aspects of future and emerging technologies, user diversity, human-computer interaction in Ambient Intelligence environments, user- or human-centered design, emotional and affective user interfaces, as well as adaptive and tangible user interfaces for Ambient Intelligence.

References

1. Aarts, E., Marzano, S.: The New Everyday - View of Ambient Intelligence. 010 Publishers (2003)
2. Kasugai, K., Röcker, C., Plewe, D., Kiriyama, T., Oksman, V.: Aesthetic Intelligence – Concepts, Technologies and Applications. In: Wichert, R., Van Laerhoven, K., Gelissen, J. (eds.) AmI 2011. CCIS, vol. 277, pp. 1–4. Springer, Heidelberg (2012)
3. Kasugai, K., Röcker, C., Bongers, B., Plewe, D., Dimmer, C.: Aesthetic Intelligence: Designing Smart and Beautiful Architectural Spaces. In: Keyson, D.V., Maher, M.L., Streitz, N., Cheok, A., Augusto, J.C., Wichert, R., Englebienne, G., Aghajan, H., Kröse, B.J.A. (eds.) AmI 2011. LNCS, vol. 7040, pp. 360–361. Springer, Heidelberg (2011)
4. Röcker, C.: Designing Ambient Assisted Living Applications: An Overview of State-of-the-Art Implementation Concepts. In: Chunxiao, X. (ed.) Modeling, Simulation and Control, Proceedings of the International Conference on Information and Digital Engineering (ICIDE 2011), Singapore, September 16-18, pp. 167–172 (2011)
5. Röcker, C.: Smart Medical Services: A Discussion of State-of-The-Art Approaches. In: Thatcher, S. (ed.) Proceedings of the International IEEE Conference on Machine Learning and Computing (ICMLC 2011), Singapore, February 26 - 28, vol. 1, pp. 334–338 (2011)

Workshop on Ambient Intelligence Infrastructures (WAmIi)

Alina Weffers[1], Johan Lukkien[2], and Tanir Ozcelebi[2]

[1] Philips Research Laboratories, Eindhoven, The Netherlands
alina.albu@philips.com
[2] Technische Universiteit Eindhoven, The Netherlands
{j.j.lukkien,t.ozcelebi}@tue.nl

Abstract. The last two decades have seen a significant amount of results and insights that promote the Ambient Intelligence vision, in particular via the architecture and design of Ambient Intelligence infrastructures supporting interconnected, context aware, personalized devices and services to act as an interactive and intelligent environment. We propose a workshop that would facilitate a systematic overview of the results achieved and work currently done in the context of European projects on the topic. The goal is to identify the white spaces in the domain, and thereby prepare the ground for further work, to be built on thorough understanding of the state-of- the-art.

Keywords: Ambient Intelligence infrastructures, middleware.

1 Introduction

The Ambient Intelligence paradigm embodies a vision of the future that has been studied since the late 1990s with significant contributions in fields such as distributed computing, profiling practices, context awareness, and human-centric computer interaction design. Since then a considerable body of work has been done in creating middleware infrastructures that facilitate connectedness, context awareness, personalization, adaptation and anticipatory execution of the devices and services they support. In the European industrial and academic communities, a large part of the effort in the Ambient Intelligence domain has been made in the context of European projects such as Metaverse1, Ozone, Amigo, WASP and SOFIA. Due to the vastness of the domain and the amount of work done already, current and future endeavors run the risk of not fully benefitting of previous work results and insights, or even partly repeating some of the work. A thorough overview and a good understanding of the concepts, the architectural directions and the existing state of affairs are mandatory. The workshop we propose is intended to address this challenge.

2 Workshop Goal

The goal of the workshop is to bring together a number of representative members of the industrial and scientific communities that have contributed in the past to the

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 447–448, 2012.

development of Ambient Intelligence infrastructures in the context of European and national projects. The input provided by the workshop participants is intended to be used in building a systematic landscape of the work done so far in the domain. This overview will provide further insight into the problems solved at this stage, advantages and disadvantages of the various approaches used, and lessons learned. In the end, the workshop discussions session would allow identifying the “white spaces” that have not yet been investigated at this stage, thereby preparing the ground for future meaningful contributions that draw from a pool of knowledge provided by a whole community rather than a particular team.

3 Topics for Discussion

The workshop discussions are open to topics related to the architecture, design and implementation of Ambient Intelligence software middleware infrastructures as well as the devices and services they support. Examples include but are not limited to:

- Architecture and design criteria leading to decisions regarding differentiating between application specific and infrastructure functionality,
- Ambient Intelligence services,
- Multi-device communication and interaction,
- Ethical aspects, privacy, security and trust,
- User interaction, embedded intelligence and learning behavior,
- Resource management and Quality of Service (QoS) management

4 Workshop Organization and Outcome

We propose a full day workshop consisting of two parts:

- The first half of the day (9.00-12.00) will be dedicated to presentations of the work described in the workshop (invited) position papers that are based on concrete results of projects.
- The second half of the day (13.00-16.00) will be dedicated to (solicited) presentations that propose new directions.

The overall result of the workshop is an overview of the landscape and a clear direction for further work to realize the AmI vision. Provided high quality of the outcome a journal paper summarizing the landscape, white spaces and opportunities for future work will be authored by the workshop participants. The article could subsequently serve as a basis for a new (FP7/Artemis) European project proposal intended to continue the work in the Ambient Intelligence domain.

Sixth International Workshop on Human Aspects in Ambient Intelligence (HAI 2012)

Juan Carlos Augusto, Tibor Bosse[1], Cristiano Castelfranchi, Diane Cook, Mark Neerincx, and Fariba Sadri

[1] Vrije Universiteit Amsterdam, Agent Systems Research Group
de Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands (Contact Person)
tbosse@few.vu.nl
http://www.cs.vu.nl/~tbosse

1 Background

Recent developments within Ambient Intelligence (AmI) provide new possibilities to contribute to personal care. For example, our car may monitor us and warn us when we are falling asleep while driving or take measures when we are too drunk to drive. As another example, an elderly person may wear a device that monitors his or her wellbeing and offers support when a dangerous situation is noticed. Such applications can be realised partly because of advances in acquiring sensor information about humans and their functioning. However, their full realisation depends crucially on the availability of adequate knowledge for analysis of such information about human functioning. If such knowledge about human functioning is computationally available within devices in the environment, these systems can show more human-like understanding and contribute to personal care based on this understanding [1].

In recent years, scientific areas focusing on human functioning such as cognitive science, psychology, social sciences, neuroscience and biomedical sciences have made substantial progress in providing increasing insight in the various physical and mental aspects of human functioning. Although much work still remains to be done, models have been developed for a variety of such aspects and the way in which humans (try to) manage or regulate them. Examples of biomedical aspects are (management of) heart functioning, diabetes, eating regulation disorders, and HIV-infection. Examples of psychological and social aspects are emotion regulation, emotion contagion, attention regulation, addiction management, trust management, and stress management.

If models of human processes and their management are represented in a formal and computational format, and incorporated in the human environment in systems that monitor the physical and mental state of the human, then such ambient systems are able to perform a more in-depth analysis of the human's functioning. An ambience is created that has a human-like understanding of humans, based on computationally formalised knowledge from the human-oriented disciplines, and that may be more effective in assisting humans by offering support in a knowledgeable manner that may improve their wellbeing and/or performance, without reducing them in their freedom. This may concern elderly people, medical patients, but also humans in highly demanding circumstances or tasks. For example, the workspaces of naval officers

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 449–450, 2012.

may include systems that track their eye movements and characteristics of incoming stimuli (e.g., airplanes on a radar screen), and use this information in a computational model that is able to estimate where their attention is focussed at. When it turns out that an officer neglects parts of a radar screen, such a system can either indicate this to the person, or arrange that another person or computer system takes care of this neglected part. Similarly, such intelligent assistants may play a role in providing support to groups of people, e.g., to help coordinate the evacuation of large crowds in case of an incident, or to optimise the performance of teams in sports or organisations.

2 Workshop Goals

This workshop addresses multidisciplinary aspects of AmI with human-directed disciplines such as psychology, social science, neuroscience and biomedical sciences. The aim is to bring people together from these disciplines, as well as researchers working on cross connections of AmI with these disciplines. The focus is on the use of knowledge from these disciplines in AmI applications, in order to take care of and support in a knowledgeable manner humans in their daily living in medical, psychological and social respects. The workshop can play an important role, for example, to get modellers in the psychological, neurological, social or biomedical disciplines interested in AmI as a high-potential application area for their models, and, for example, get inspiration for problem areas to be addressed for further developments in their disciplines. From the other side, the workshop may make researchers in Computer Science and Artificial and Ambient Intelligence more aware of the possibilities to incorporate more substantial knowledge from the psychological, neurological, social and biomedical disciplines in AmI architectures and applications. As part of the interaction, specifications may be generated for experiments to be addressed by the human-directed sciences.

Since 2007, the workshop has been held annually, and has always attracted a substantial amount of high quality submissions and interested visitors. The first version of the workshop was organised in 2007 within the European Conference on Ambient Intelligence (AmI) in Darmstadt (Germany). From 2008 until 2011, the workshop has been organised within the International Conference on Intelligent Agent Technology (IAT), in Sydney (Australia), Milan (Italy), Toronto (Canada), and Lyon (France), respectively. The sixth edition of the workshop [2] takes place on November 13th, 2012, in Pisa, Italy, and is hosted by AmI 2012. Seven high quality papers on various topics related to human aspects in Ambient Intelligence have been accepted for presentation at the workshop.

References

1. Treur, J.: On Human Aspects in Ambient Intelligence. In: Mühlhäuser, M., Ferscha, A., Aitenbichler, E. (eds.) AMI 2007 Workshops. CCIS, vol. 11, pp. 262–267. Springer (2008)
2. http://www.few.vu.nl/~tbosse/HAI12/

Context-Aware Adaptation of Service Front-Ends

Francisco Javier Caminero Gil[1], Fabio Paternò[2], and Jean Vanderdonckt[3]

[1] TID, Telefónica I+D
Distrito Telefónica, Edificio Oeste 1, Planta 5, Ronda de la Comunicación, s/n
E-28050 Madrid, Spain
fjcg@tid.es
[2] CNR-ISTI, HIIS Laboratory
Via Moruzzi 1 56124 Pisa, Italy
fabio.paterno@isti.cnr.it
[3] Université catholique de Louvain
Place des Doyens, 1, B-1348 Louvain-la-Neuve (Belgium)
jean.vanderdonckt@uclouvain.be

Abstract. Ambient Intelligence implies the need for context-aware adaptation of user interfaces. This adaptation with respect to the context of use is applicable to a wide spectrum of interactive applications ranging from front ends of web services, information systems to multimedia and multimodal applications. Although the ultimate goal of this adaptation is always for the ultimate benefit of the end user, many approaches and techniques have been used to various degrees of experience and maturity that effectively and efficiently support context-aware adaptation. This workshop is intended to review the state of the art in this domain, while looking at a broad range of applications, to discuss positive and negative experiences of context-aware adaptation, and to come up with criteria and requirements for driving such adaptation.

1 Theme, Goals, Relevance and Format

Ambient Intelligence needs the support of multi-device user interfaces, which are sensitive and responsive to people and their behaviors, and able to deliver advanced functions, services and experiences. The context of use can vary in terms of many aspects: users, technology, environment, social relations. Many different pieces of work have been conducted in order to address the challenge of context-aware adaptation with valuable inputs from various disciplines such as, but not limited to ubiquitous computing, pervasive applications, ambient intelligence, artificial intelligence, engineering interactive computing systems, mobile human-computer interaction. These contributions are often expressed in a format that prevents them to compare them to each other, mainly because they are heterogeneous and their format is inconsistent. This workshop is aimed at addressing context-aware adaptation of user interfaces via an interaction model according to three dimensions (adapted from [1]): a descriptive power that consists in characterizing a sufficiently large spectrum of adaptation techniques according to a unified format, an evaluative power that consists in comparing different adaptation techniques based on the same format, and a generative virtue that is intended to identify holes in the resulting design space in order to foster further

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 451–452, 2012.

research and development of new interaction techniques. "A good interaction model must strike a balance between generality (for descriptive power), concreteness (for evaluative power) and openness (for generative power)" [1]. This means that there is a need for creating a design space for context-aware adaptation of any kind of user interface that should exhibit enough expressiveness (for guarantying enough descriptive power), decidability (for ensuring the ability to assess any adaptation technique), and flexibility (for accommodating new adaption techniques that were previously unforeseen). There is also a need for discussing and reviewing the state of the art in the domain of UI adaptation under the view-point of context-awareness since most surveys are either obsolete [2] or do not address context-awareness [3]. This state of the art will be namely focusing on models, methods, and tools that support context-aware adaptation of UIs, in particular in order to address existing challenges.

The workshop will be a half day workshop that will result into discussing and reviewing the state of the art in the area of context-aware adaptation: a design space for context-aware adaptation of user interfaces, criteria and requirement for an Adaptation Specification Language (ASL), and a common format for expressing such adaptation rules.

We intend to discuss at the workshop the following topics:

- How can we obtain an ambient intelligence where the user interfaces of multiple devices are sensitive and responsive to people and their behaviors, and able to deliver advanced functions, services and experiences?
- What are the major challenges (e.g., conceptual, methodological, technical, organizational) for developing context-aware adaptation of user interfaces?
- For which kinds of systems or applications are context-aware user interfaces particularly useful?
- When and how could we measure the effectiveness, the efficiency of context-aware adaptation?
- How could we measure the quality of the user interface resulting from a context-aware adaptation process?
- In which ways will context-aware adaptation affect user interfaces in the future and how will they evolve?
- What kinds of context-aware adaptation do you see as particularly promising?

References

1. Beaudouin-Lafon, M.: Designing interaction, not interfaces. In: Proc. of ACM Working Conf. on Advanced Visual Interfaces AVI 2004, Gallipoli, May 25-28, pp. 15–22. ACM Press, New York (2004)
2. Dieterich, H., Malinowski, U., Kuhme, T., Schneider-Hufschmidt, M.: State of the art in adaptive user interfaces. In: Schneider-Hufschmidt, M., Kuhme, T., Malinowski, U. (eds.) Adaptive User Interfaces Principles and Practice, pp. 13–48. Elsevier Science Publishers B.V., Amsterdam (1993)
3. López-Jaquero, V., Vanderdonckt, J., Montero, F., González, P.: Towards an Extended Model of User Interface Adaptation: The ISATINE Framework. In: Gulliksen, J., Harning, M.B., van der Veer, G.C., Wesson, J. (eds.) EIS 2007. LNCS, vol. 4940, pp. 374–392. Springer, Heidelberg (2008)

2nd International Workshop on Ambient Gaming

Janienke Sturm[1,2], Pepijn Rijnbout[1], and Ben Schouten[1,2]

[1] Dept. Industrial Design, Eindhoven University of Technology, P.O. Box 513, 5600MB Eindhoven, The Netherlands
[2] Serious Game Design lectorate, Fontys University of Applied Sciences, P.O. Box 347, 5600AH Eindhoven, The Netherlands
{j.sturm,bschouten,p.rijnbout}@tue.nl

Abstract. Ambient games are games and playful activities that offer context-aware and personalized features. Because ambient play and games can be incorporated in everyday objects and routines, they allow players to play throughout the day. Ambient gaming offers promising opportunities for creating novel and unique player experiences. However, there are still many unanswered questions related to this new field of research, for instance related to gamification, personalisation and adaptation, aspects and issues of control and privacy. In this 2nd workshop on Ambient Gaming we intend to further discuss the opportunities and challenges in the field of ambient gaming and play with people from different disciplines (designers, researchers, and developers) who are active in this field.

Keywords: Keywords: Ambient Gaming, Playful Interactions, Design, Ambient Technology.

1 Workshop Description

In many historical works about play, the definition of play is restricted to a specific 'time and place', separated from ordinary life (i.e. play takes place in a 'magic circle') [1]. Digital play, however, can be more integrated in a spatial, temporal and social sense [2] owing to new media, social networks, modern technology and (social) interaction. This enables us to design for playful activities that are seamlessly integrated within our daily lives in such a way that the boundaries between other activities and play disappear or blur; we call this ambient games and play. Ambient games blend the virtual and real world and are interacted with through multiple ubiquitous devices. They incorporate ambient intelligence characteristics, such as personalization, adaptation and anticipation [3]. Ambient intelligence environments may sense who is present, where they are, what they are doing, and when and why they are doing it. In line with this, ambient games offer context-aware and personalized features. They also allow players to move around freely, without being bound by a computer screen or another device, by using information coming from sensors embedded in the environment. By their nature, they allow players to play throughout the day, as play and games may be incorporated in everyday objects and routines [4].

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 453–454, 2012.

Ambient gaming offers promising opportunities for creating novel and unique player experiences. However, there are still many unanswered questions related to this new field of research, for instance related to gamification, personalisation and adaptation, aspects and issues of control and privacy, as we concluded from the first workshop on Ambient Gaming, which was held at AmI-11 in Amsterdam. In this 2nd workshop on Ambient Gaming we intend to further discuss the opportunities and challenges in the field of ambient gaming and play with people, from different disciplines, active in this field (designers, researchers, and developers). The workshop objectives are:

- To share experiences, research, insights and best practices regarding issues related to design, theory, technology, and methodology;
- To unfold the challenges and opportunities of this interesting emerging area and develop a common research agenda for future studies;
- To come to a strategy and research agenda in the field of Ambient Gaming for the forthcoming years
- To stimulate participants to form collaborations to advance the field further in a multidisciplinary manner.

2 Workshop Organisation

The workshop is a half-day workshop, with a maximum of 20 participants (designers, researchers and technologists in the field of games and play).

Participants will be selected on the basis of a position paper describing their area of research and their specific interest in the topic of ambient gaming. All position papers will be peer reviewed by an international program committee, consisting of experts in the field of technology and design for games and playful interactions. Each paper will be assigned to two reviewers.

To minimize the amount of overhead during the workshop, we will make sure that abstracts of the position papers are available to attendants prior to the workshop. In addition, all workshop attendants will be asked to post one or more questions or discussion statements on the workshop website before the workshop starts. These will be used to kick-start our discussions during the workshop. The workshop has its own website: http://www.playfitproject.nl/ambientgamingworkshop.

References

1. Huizinga, L.: Homo Ludens. A Study of the Play Element in Culture. Beacon Press, Boston (1955)
2. Montola, M., Stenros, J., Waern, A.: Pervasive games: theory and design. Morgan Kaufmann (2005)
3. Aarts, E., Marzano, S. (eds.): The New Everyday: Visions of Ambient Intelligence. 010 Publishers, Rotterdam (2003)
4. Sturm, J., Tieben, R., Deen, M., Bekker, T., Schouten, B.: PlayFit: Designing playful activity interventions for teenagers. In: Proceedings of Digra 2011, Hilversum, The Netherlands,, September 14-17 (2011)

Designing Persuasive Interactive Environments

Marco Rozendaal[1], Aadjan van der Helm[1], Walter Aprile[1], Arnold Vermeeren[1], Tilde Bekker[2], Marije Kanis[3], and Wouter Middendorf[4]

[1] Delft University of Technology, Faculty of Industrial Design Engineering, The Netherlands
{m.c.rozendaal,a.j.c.vanderhelm,w.a.aprile, a.p.o.s.vermeeren}@tudelft.nl
[2] Eindhoven University of Technology, Department of Industrial Design, The Netherlands
m.m.bekker@tue.nl
[3] Amsterdam University of Applied Sciences, CREATE-IT Applied Research, The Netherlands
m.kanis@hva.nl
[4] Fabrique - Brands, Design and Interaction, The Netherlands
wouterm@fabrique.nl

Abstract. Ambient Intelligent environments are interactive environments that can sense human behaviour and can respond intelligently. This workshop explores how interactive environments can be designed with persuasive quality, hereby influencing human experience and behaviour. The workshop follows a research-through-design approach in which practise-relevant insights are gained while designing. The focus will be on intuitive and rational decision-making, the role of aesthetics in persuasion, social and spatial influences as well as technological influences on persuasion, evaluation methods for persuasion, and the ethics of designing for persuasion.

Keywords: Ambient intelligence, persuasion, aesthetics, decision-making research-through-design, interactive prototyping.

Introduction and Workshop Goals

Ambient Intelligence allows for adaptive environments that sense human behaviour and that react intelligently through lighting, audio-visual media and physical adaptations [1]. Much research focuses on the technical feasibility of adaptive environments, such as sensor accuracy, image recognition, and so on. The aim of this workshop is to investigate the persuasive quality of adaptive environments by investigating its transforming effect on human experience and behaviour.

This workshop intends to bring together designers, researchers and technologists in the field of ambient intelligence who want to explore the complexities and subtleties involved when designing adaptive environments that influence behaviour [2, 3]. The workshop will provide a unique opportunity for practical design activities. This means that during the workshop, interactive prototypes will be created with the intent to influence behaviour on conference participants. During the conference, visitors can experience the prototypes in the conference commons in order to assess whether the anticipated behavioural effects are actually realized.

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 455–456, 2012.

The workshop follows a *research-through-design* approach [4] consisting of two rapid design cycles where participants both design and reflect. The merit of such an approach is that it leads to practise-relevant insights in a short period of time. An interactive sketching tool will be developed for the workshop, so to enable participants to build interactive prototypes quickly and intuitively without requiring much technical expertise [5].

When designing for persuasion, the capabilities of an interactive sketching tool, the characteristics of the physical environment and peoples' inherent social behaviours need to be matched and considered. Reflecting on design involves discussing how aspects of persuasion are embodied in a prototype's form and interactivity. The aim is to publish the results of the workshop in a shared publication, targeting the ambient intelligence community.

Topics to Be Reflected on

The aim is to explore the persuasive quality of interactive environments on human experience and behaviour. We connect to the work of Fogg [6] with a specific emphasis on design related issues as discussed by Lockton et al [7]. We invite participants to reflect on the following aspects of persuasion in relation to form and interactivity:
Intuitive and rational decision making, the role of aesthetics in persuasion, social and spatial influences on persuasion, technological influences on persuasion strategies, evaluation methods for persuasion and the ethics of designing for persuasion.

References

1. Aarts, E., Marzano, S.: The New Everyday: Views on Ambient Intelligence, 010 Publishers. Rotterdam, The Netherlands (2003)
2. Bekker, M.M., Sturm, J., Eggen, B.: Designing playful interactions for social interaction and physical play. Personal and Ubiquitous Computing 14(5), 385–396 (2010)
3. Rozendaal, M., Vermeeren, A., Bekker, T., de Ridder, H.: A Research Framework for Playful Persuasion Based on Psychological Needs and Bodily Interaction. In: Salah, A.A., Lepri, B. (eds.) HBU 2011. LNCS, vol. 7065, pp. 116–123. Springer, Heidelberg (2011)
4. Zimmerman, J., Forlizzi, J., Evenson, S.: Research through design as a method for interaction design research in HCI. In: Proceedings of the SIGCHI Con-ference on Human Factors in Computing Systems, pp. 493–502. ACM, New York (2007)
5. Aprile, W.A., van der Helm, A.J.: Interactive technology design at the Delft University of Technology - a course about how to design interactive products. In: International Conference on Engineering and Product Design Education, City University, London, UK, September 8 & 9 (2011)
6. Fogg, B.J.: Persuasive technology: Using computers to change what we think and do. Morgan Kaufmann, New York (2010)
7. Lockton, D., Harrison, D., Stanton, N.A.: The Design with Intent Method: A design tool for influencing user behaviour. Applied Ergonomics 41(3), 382–392 (2010)

Applying AmI Technologies to Crisis Management

Workshop at AmI2012

Monica Divitini[1], Babak Farshchian[2], Jacqueline Floch[2], Ragnhild Halvorsrud[2], Simone Mora[1], and Michael Stiso[2]

[1] Dept. of Information and Computer Science, NTNU, Trondheim, Norway
{divitini,simonem}@idi.ntnu.no
[2] SINTEF ICT, Oslo, Trondheim, Norway
{Babak.Farshchian,Jacqueline.Floch,Ragnhild.Halvorsrud,
Michael.Stiso}@sintef.no

Abstract. The workshop aims to bring together researchers and practitioners working on the application of AmI to crisis and disaster management. Because of their pervasiveness and ease of use, AmI technologies hold a great potential to support crisis management in an efficient and effective way. Focus will be on better understanding (1) the strengths of the AmI paradigm, (2) challenges to its application, and (3) its potential in the development of innovative solutions. The workshop is open to papers from different standpoints, including platform and user interaction issues, methodological approaches, and specific applications.

Keywords: Ambient Intelligence, Crisis Management, innovative AmI solutions.

1 Introduction

Natural and man-made disasters are on the rise, with sources reporting on a five-fold increase of natural disasters in the last 35 years[1]. In 2010, DG ECHO (the EU Directorate for Humanitarian Aid and Civil Protection) reported an EU expenditure of €1115 million to respond to new or protracted crises, and 373 natural disasters killing around 300000 people[2].

Recent years have seen an increasingly rapid introduction and advancement of technologies that allow unprecedented levels of effectiveness in and control over a crisis response. ICT solutions proposed for supporting crisis management vary considerably in scope and complexity, ranging from organizational workflow systems to coordinate different forces that are employed in specific events up to platforms like Ushahidi (http://ushahidi.com/) for crowd sourcing and the usage of Twitter (twitter.com) to share information among the population.

---

[1] http://www.euractiv.com/foreign-affairs/europe-beef-response-natural-disasters-news-499193

[2] EU DG ECHO, Annual Report 2010, available at http://ec.europa.eu/echo/files/media/publications/annual_report/annual_report_2010.pdf

F. Paternò et al. (Eds.): AmI 2012, LNCS 7683, pp. 457–458, 2012.

The workshop will facilitate knowledge exchange among the community of researchers interested in AmI for crisis management, creating opportunities for synergies and for transferring knowledge across institutional and geographical boundaries.

2 Relevant Topics

Crisis management is a particularly relevant application domain for AmI. AmI has the potential to increase the efficiency and effectiveness of crisis management, and thereby to contribute to saving lives, reducing risks for rescue teams and lowering costs. Several example solutions are described in the research literature, such as monitoring of environmental data under hard conditions, impact of information presentation on decision-making, rescuer teams management supported with physiological data monitoring, situational awareness support for rapid crowd evacuation.

The workshop aims at offering to researchers and practitioners a space to reflect on where these increasingly pervasive and ambient technologies are going, what they will make possible, and how they will be used. Focus is on challenges connected to the use of AmI in crisis management as well as the opportunities to use AmI to conceive innovative solutions, e.g. empowering not only traditional actors, but also the population at large; supporting not only management, but also promoting continuous learning and training. Relevant topics include platforms issues, user interaction in challenging environments, methodologies and applications.

3 Organization

The workshop is jointly organized by three EU IST research projects that investigate from different perspectives ICT support for crisis management:

- *BRIDGE* (http://www.bridgeproject.eu) aims at building a system to support technical and social interoperability in large-scale emergency management.
- *MIRROR* (http://www.mirror-project.eu/) aims at developing ICT tools for supporting workplace reflection and learning. Training of crisis workers is a core application domain of the project.
- *SOCIETIES* (http://www.ict-societies.eu/) aims at extending the application of pervasive computing beyond the individual to communities of users. Disaster management is chosen as one area for the evaluation of the proposed solutions.

More information about the workshop is available at the workshop website: http://ami4cm.wordpress.com/

Author Index