Doris Hausen, Bettina Conradi, Alina Hang,
Fabiant Hennecke, Sven Kratz, Sebastian Löhmann,
Hendrik Richter, Andreas Butz, Heinrich Hussmann
(Editors)

Ubiquitous Computing

Hauptseminar Medieninformatik SS 2011

Technical Report
LMU-MI-2011-1, Aug. 2011
ISSN 1862-5207

University of Munich
Department of Computer Science
Media Informatics Group

Doris Hausen, Bettina Conradi, Alina Hang, Fabian Hennecke,
Sven Kratz, Sebastian Löhmann, Hendrik Richter
Andreas Butz, Heinrich Hussmann (Editors)

Ubiquitous Computing

An overview of current trends, developments, and research in
Ubiquitous Computing

Preface

This report provides an overview of current applications and research trends in the field of ubiquitous computing. There are various applications domains ranging from haptic interfaces, interactive surfaces, mobile devices to ambient information.

During the summer term 2011, students from the Computer Science Department at the Ludwig-Maximilians-University in Munich did research on specific topics related to ubiquitous computing and analyzed various publications. This report comprises a selection of papers that resulted from the seminar.

Each chapter presents a survey of current trends, developments, and research with regard to a specific topic. Although the students' background is computer science, their work includes interdisciplinary viewpoints such as theories, methods, and findings from interaction design, ergonomics, hardware design and many more. Therefore, the report is targeted at anyone who is interested in the various facets of ubiquitous computing.

Munich, August 2011

The Editors

Doris Hausen, Bettina Conradi, Alina Hang, Fabian Hennecke,
Sven Kratz, Sebastian Löhmann, Hendrik Richter
Andreas Butz, Heinrich Hussmann

Contents

Use of Freehand Gestures for Interactions in Automobiles: An Application-Based View

Jennifer Zeiser

Abstract— The use of freehand gestures for interaction in automotive environments is an approach to improve security while driving. Since driving demands a person cognitively and visually, the goal is to reduce driver distraction through gestural input options to perform secondary tasks. This paper provides an introduction to fundamentals of gesture theory, as well as an overview of application areas using gesture-based interaction. In the main part, special conditions of automotive environments are introduced with regard to the use of gesture-based input options. Next it is examined which categories of gestures make sense while driving and in which scenarios the use of gesture-based interaction is especially helpful. Moreover, this work presents existing gesture-based approaches for in-car applications from the world of research. Finally, it is examined whether or not aspects of other application areas can be transferred to automotive environments.

Index Terms—Freehand Gestures, Automotive, Interaction Modality, Ubiquitous Computing

1 Introduction

Driving demands a person cognitively and visually, therefore other activities can distract the driver from controlling the car and concentrating on the traffic. Especially secondary tasks, like controlling an infotainment system, lead to drivers' inattention [2]. Since inattention causes a significant amount of car accidents [8] and interaction with in-car devices becomes more complex due to increased functionality [43] [44] [45], it is reasonable to use input options which cause less driver distraction in both the cognitive and the visual domain. Nowadays, interaction with in-car devices is mostly realized haptically by buttons and rotary knobs [16]. Further input possibilities include speech control [7] and interaction via direct touch interfaces [10]. Except for voice input, all these input options require eye contact with the user interface, again leading to visual distraction of the driver. To overcome this disadvantage, there are many approaches to use gestures as controll option.

Gestures have already been used for interaction in other application areas, e.g. for human-robot interaction (HRI) [25] [41] [19], controlling computer displays [5] [28] [39], sign language interpretation [25] or interaction with medical systems [41].

The aim of this work is to investigate the meaning of gesture-based input options for interaction with in-car devices. Therefore, it will be explored which circumstances limit the use of gestures and which categories of gestures are applicable while driving. Moreover, this work examines if aspects of other application areas using gesture-based interaction can be applied to automotive environments.

This paper is structured as follows: Section 2 gives a general introduction to gestures including different definitions and categorization approaches. Following, different application areas using gesture-based interaction are introduced in section 3. Section 4 covers special conditions in cars, advantages and disadvantages of gesture-based interaction in automotive environments, as well as technologies to realize gesture-based interaction and existing research approaches to use gestures to control in-car devices. In section 5, the tranferability of aspects of other application areas to automotive environments is examined.

2 Gesture Theory

To speak about gestures and their potential use in automotive environments, first some fundamentals of gesture theory including different definitions and categorization schemes, as well as phases of gestures and modeling approaches are introduced.

2.1 Definition of Gestures

Gestures are primarily used in interpersonal communication to support human dialogs [33], but they are increasingly applied to Human-Computer Interaction (HCI) as input options, too [21]. Depending on the context of their usage, gestures can be defined in different ways.

In his book, *"Hand and Mind. What Gestures reveal about Thought"* [26], McNeill speaks of gestures as *"spontaneous movements* [...] *of the hands and arms"*. A similar definition is given by Nehaniv [27], who reveals that gestures are *"phenomena involving human movements, especially of the hands and arms"*. According to Althoff et al. [3], gestures *"correspond to a movement of individual limbs of the body and are used to communicate information"*. However, Kendon [22] defines gestures as *"actions that have the features of manifest deliberate expressiveness"*.

None of these definitions refers either to gestures as a means of communication between humans and machines or includes head gestures. A more suitable definition can be found in Fikkerts work *"Gesture Interaction at a Distance"* [12]. Therein, gestures are defined with relation to HCI. According to Fikkert's classification, gestures are *"motion*[s] *of the hands, facial expressions, gaze tracking, head movements, hand postures, and whole body postures"*.

In the context of this paper, we will refer to gestures as movements of the hand, arm or head which can be used to communicate information, not only in interpersonal relationships, but for the purpose of human-machine interaction.

2.2 Phases of Gestures

Having defined gestures in the context of this work, we will have a closer look at the structure of gestures. Gestures are often perceived as one continuous movement, but they are principally comprised of three phases [26]:

Phase 1 - Preparation (optional): In the first phase of a gesture, the hand is brought to the starting position for the *stroke*.

Phase 2 - Stroke (essential): The *stroke* refers to the main part of a gesture and defines their meaning.

Phase 3 - Retraction (optional): The last phase includes the return of the hand or arm to the resting position.

This distinction primarily refers to hand and arm gestures, but with the limitation that phase 1 and 3 are optional while performing a gesture. It can also be applied to head gestures or facial expressions. Because of the fact that phase 1 and 3 are no essential parts of a gesture, the separation of the particular phases is not always easy. Especially in

• *Jennifer Zeiser is studying Media Informatics at the University of Munich, Germany, E-mail: zeiser@cip.ifi.lmu.de*

• *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

a sequence of hand gestures where two gestures can merge into each other seamlessly.

However, this distinction is important since gesture recognition systems have to detect the preparation and the retraction phase of a gesture or their absence in order to analyze only the meaning of the *stroke*.

2.3 Modeling of Gestures

The modeling of gestures, especially their production and perception, can be described by a simple model [33] *(see figure 1)*.

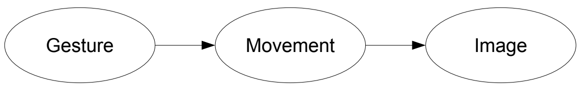


Fig. 1. Model of gesture production and perception [33]

According to this model, gestures arise out of a mental concept which exists in a person's imagination. Based on this mental concept, a person produces a gesture by moving his or her hands or arms. These movements are perceived as visual images by observers and are interpreted subsequently.

With regard to interaction in automotive environments, *"gesture"* describes the intention of the driver or person who interacts with the user interface, *"movement"* constitutes the gestural input and *"image"* is the result of the gesture recognition that causes a reaction of the system.

2.4 Categorization of Gestures

The following section will present four different categorization schemes for gestures. These classifications refer to interpersonal communication, HCI, Human-Robot Interaction (HRI), and automotive environments. Based on these approaches, it is examined which types of gestures can be applied to automotive environments in section 4.

2.4.1 Types of Gestures according to McNeill [26]

McNeill distinguishes between five types of gestures in the context of interpersonal communication.

Iconics: This type of gestures refers to representations of specific objects or events. They typically correspond to the semantic of a person's speech.

Metaphorics: In contrast to *iconics*, *metaphoric* gestures refer to abstract ideas, i.e. they are metaphors which describe a mental concept.

Beats: This category of gestures describes movements which are synchronized along with the rhythm of speech. The special characteristic of *beats* is that they consist of only two phases, e.g. up and down or in and out.

Deictics: These gestures correspond to all types of pointing gestures. They are primarily used to refer to objects and events in the real world.

Cohesives: Gestures of this type are used to relate different parts of a conversation. They can consist of any aforementioned type of gesture, but differ from them by their intention.

2.4.2 Taxonomy of Gestures for HCI by Fikkert [12]

In contrast to McNeill's classification of gestures in human communication, Fikkert describes five types of gestures which highlight the *"process of interacting"* and can be applied to HCI.

Deictics: Fikkert's definition of *deictics* is principally the same as McNeill's. However, he states that *deictics* are *"the basis for communicating with machines as equal partners in communication"*.

Manipulations: *Manipulative* gestures can be used to interact with real objects. Moreover, they can be mapped to objects in a virtual interface.

Semaphores: This type of gestures can be used in a static or dynamic way for signaling purposes. An example for a *semaphore* is a static hand pose.

Gesticulation: This gesture class principally refers to human communication, but can be applied to HCI in combination with speech interfaces. For example, *"spontaneous movements of the hands and arms during speech"* [12] are *gesticulations*.

Language Gestures: Related to HCI, this type of gestures is used in communication applications using not only hand, arm, and head gestures, but also facial expression and body posture.

2.4.3 Categories of Intent by Nehaniv [27]

Nehaniv's [27] classification serves as initial situation to evaluate gestures as a means of human-robot interaction.

Irrelevant/Manipulative gestures are gestures without communicative intentions. In fact, they *"influence on the non-animated environment or human's relationship to it"* [27], e.g. finger-tapping or playing with a pencil.

Side effects of expressive behaviour: This type of gesture has no communicative intentions, too, but they occur in interpersonal communications. For example, gestures which reinforce the variation in the speaker's pitch.

Symbolic gestures serve as signals in interpersonal communication, e.g. nodding for *"yes"* or head-shaking for *"no"*, and correspond to *"discrete actions on an interface, such as clicking a button"* [27].

Interactional gestures: In contrast to *irrelevant* or *manipulative* gestures, these gestures serve to interact with the *"animated environment"* [27]. They can be used to control a conversation and interact with a conversational partner, e.g. nodding can imply that someone is listening.

Referential/Pointing gesture: This type of gesture is equivalent to *deictic* gestures in McNeill's classification.

2.4.4 Categories of Gestures by Geiger [14]

Geiger's [14] classification pays particular attention to the communicative intention of a gesture and refers to gestures in the context of interaction in automotive environments. For this purpose, gestures are divided into primary gestures and secondary gestures. Primary gestures are intended solely for information exchange and the transmission of intentions. On the contrary, secondary gestures do not transfer intentions and have no communicative function. Primary gestures can be divided into the following categories:

Whole-body gestures use not only parts of the body, but the full body to express communicative intentions, e.g. pantomime.

Partial-body gestures describe gestures which are carried out by hand, arm and head movements.

Dynamic gestures are characterized by the transmission of information as part of the motion sequence, e.g nodding. Dynamic gestures can be further subdivided into continuous and discrete gestures. The former refers to direct manipulation whereas the latter stands for indirect manipulation.

Static gestures: In contrast to *dynamic* gestures, the motion sequence is of no relevance in this gesture category. Instead, information is transferred through the form of the gesture or the posture of hand, arm or head.

Based on this classification, Geiger confines himself to dynamic partial-body gestures as a means of human-machine interaction. This gestural type can be further subdivided:

Mimic gestures serve to imitate concrete objects or events and can also be related to their properties. For example, mimic gestures can be used to pick up a phone fictively.

Schematic gestures are a special kind of mimic gestures. They are characterized by the fact that they cannot be understood without knowing their meaning.

Kinemimic gestures refer to gestures in which the direction of motion plays the most important part.

Symbolic gestures refer to abstract features, but they can serve to regulate interpersonal communication, too. Examples for *symbolic* gestures are nodding or head-shaking.

Deictic gestures correspond to *deictics* as described by McNeill and referential/pointing gestures as mentioned by Nehaniv.

Technical gestures are predefined and their meaning differs depending on the application area. They are mainly employed when speech cannot be considered as medium of conversation, e.g. while diving.

Encoded gestures correspond to a language of technical gestures, e.g. sign language.

Since Geiger examines gestures in the context of automotive environments, this work will take his classification as basis for further examinations.

3 Overview of gesture-based applications

Having introduced fundamentals of gesture theory, this section presents an overview of application areas using gesture-based interaction. Based on these fields of application, it is elucidated if aspects of other applications are applicable to gestural interaction in automotive environments in section 5.

3.1 Robotics and Teleoperation

Robotics and teleoperation is one application area using gesture-based interaction. Typical tasks are remote control of robots [25] and robot programming by demonstration (RbD) [34].

Remote control of robots is used to move and control robots through gestures. Possible commands encompass ”go there”, ”grasp” or ”put it down” [41].

Nickel and Stiefelhagen [29] describe an approach to control robots only by deictic gestures. Their HMM[1]-based pointing gesture recognizer detects hands, head, and head orientation by a stereo camera and processes them through a multi-hypothesis tracking framework.

The *GestureDriver* is a remote driving application to control robots introduced by Fong et al. [13]. The recognition of hand movements is realized by a color and stereo vision system. The system provides both direct and indirect control modes which is realized by a virtual joystick. Therefore, the user does not have to wear tracking hardware, such as gloves.

Remote control of robots is also suitable for handicapped persons, e.g. if speech control is no input option [41].

Robot programming by demonstration, also known as *Learning by Imitation* [19], is an approach to control and program humanoid robots [34] [19] through gestural input. Nowadays, robots are mostly controlled by graphical user interfaces, such as teachpanels. However, robot programming by demonstration uses sensor-based or kinesthetic methods *(see figure 2)* to teach new skills to robots [9].

[1]Hidden Markov Model

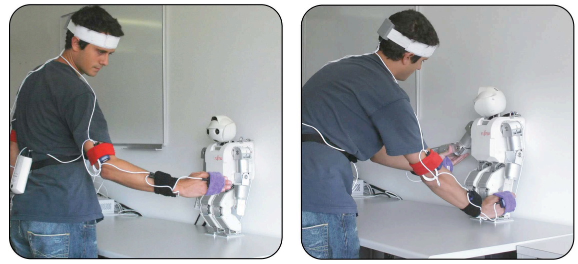


Fig. 2. Left: teaching with motion sensors (sensors record the person's movements during demonstration). Right: kinesthetic teaching (user corrects the movements of the robot) [9].

Calinon and Billard [9] present an approach to incremetally teach a robot human gestures, in particular basketball officials' gestures, using the aforementioned methods.

Iba [19] describes a vacuum-cleaning robot which is handled by an interactive multi-modal robot programming framework. This framework consists of an *Intention Interpretation Module* and a *Prioritized Task Execution Module*. The former is used for task selection, priority attachment, and adaption of the task representation. The latter serves to execute state-based primitives and generates a robot program.

Rogalla et al.[34] developed the humanoid robot *ALBERT*. The mobile platform of the service robot enables navigation via a laserscanner and uses an interactive system, based on *events*, to handle sensor input.

3.2 Medical Systems

Gesture-controlled devices are suitable for the employment in surgery to ensure sterility, as well as for kinesthetic therapies [41]. Moreover, gesture-based interaction can be used to control wheelchairs remotely by head orientation and hand gesture recognition [24].

Remote control in surgery: Graetzel et al. [17] present a system which enables gesture-based interaction between surgeon and operation room equipment during operation. The gestures are recognized by a camera and replace interaction via a computer mouse. Wachs et al. [40] introduce *Gestix* which is a *doctor-computer sterile gesture interface for dynamic environments*. It tracks a surgeon's hand to display magnetic resonance images in surgery. A further approach is the human-machine interface *FAce MOUSe* which is presented by Nishikawa et al. [31] This system tracks a surgeon's face movements to control a laparoscope[2].

Gesture-based therapy: Boian et al. [6] present a virtual reality-based system using gloves to treat post-stroke patients and help them to regain the flexibility of their hands and fingers. Gutierrez et al. [18] introduce a virtual environment-based haptic workstation for kinesthetic therapy. The workstation uses handheld devices for gesture recognition and can be used by patients with disorders of the upper limb.

3.3 Sign Language Recognition

Gesture-based applications are suitable for sign language recognition [36]. Early approaches used data gloves to recognize hand gestures. For example, Kramer and Liefer [23] present a system which enables deaf, blind-deaf, and non-vocals to communicate verbally. For this purpose, fingerspelling is recognized using gloves and transformed to speech output and displayed text. A similar approach is introduced by Fels and Hinton [11], therein gestures are transformed to speech using data gloves and a speech synthesizer.

[2]medical instrument used to examine the abdominal cavity

Besides gloves, cameras are used to recognize sign language. Starner [36] presents an attempt to recognize hand gestures of sign language using single view cameras and hidden markov models. Furthermore, Ong and Ranganath [32] introduce a camera-based system which is capable of recognizing hand and head gestures, as well as facial expressions.

3.4 Cockpit applications

There are some approaches to use gesture-based interaction to control in-cockpit devices. Ineson et al. [20] examined camera-based finger recognition to control cockpit applications, e.g. interaction with visual buttons. Nevertheless, they came to the conclusion that speech input is a more suitable input option. Another approach uses the electromagnetic tracker to select targets on a 3D map [25].

Wheeler and Jorgensen [42] present a gesture recognition system using an neuroelectric interface. Therein, the user interacts with virtual joysticks and keyboards wearing a dry-electrode sleeve. In the meantime, electromyogram signals from the muscles are measured. This approach can also be applied to video gaming.

3.5 Entertainment and Multimedia Systems

Gestural input options are particularly attractive for video gaming because they extend the game experience and allow the player to feel as part of the game. Consider the Nintento Wii as an example, it provides a wireless remote control and uses signal sensing via infrared LEDs as well as accelaration sensors for gesture recognition [38] [35]. Another example for this application area is the Microsoft Kinect which uses various cameras and a depth sensor for input recognition. Thereby, whole-body gestures and facial expressions are recognized [38].

Starner, Leibe, and Singletary [37] present a multi-player augmented reality game which uses the Wearable Augmented Reality for Personal, Intelligent, and Networked Gaming (WARPING). This system accepts hand and head gestures, as well as speech as input option. The system incorporates near-infrared light sources and a CCD[3]-camera to recognize hand and head gestures.

3.6 Desktop and Display Interaction

Gestures as input option have the potential to supplement traditional input devices, such as the mouse or the keyboard. Baudel and Beaudouin-Lafon [5] introduced CHARADE, a control interface for computer-aided presentations using hand gestures which allows interaction with the presentation. The system interface defines an active zone and accepts 16 gestural commands for interaction purposes. This is realized by an overhead projector, an LCD-display, data gloves and a tracker.

Ni, McMahan and Bowman [28] presented the *"roll-and-pinch menu"* (rapMenu) for remote control of computer displays. rapMenu accepts freehand gestures as input. The roll movement is used for navigation tasks and the pinch movement performs a selection task. The surface provides an effective zone, similar to the active zone of CHARADE. Gesture recognition is realized using a combination of gloves and a motion tracking system.

Vogel and Balakrishnan [39] examinded freehand gestural input, especially pointing and clicking, for interaction with high resolution displays *(see figure 3)* which can be used for presentation purposes, for example. Their approach uses passive reflective markers and a motion tracking system. Moreover, the system supports different clicking techniques, such as AirTrap and ThumbTrigger, and pointing techniques, like RayCasting, Clutching and Hybrid RayToRelate Pointing (see [39] for further information on these techniques).

4 Gestures in Automobiles

Having introduced application areas using gestural input options, the following section deals with gesture-based interaction in automotive environments. It covers special requirements, as well as advantages and disadvantages of gesture-based control possibilities for in-car devices. Moreover, technologies for gesture recognition and existing gesture-based applications for automotive environments are presented. In addition, this section examines which types of gestures are applicable while driving.

[3]charge-coupled device

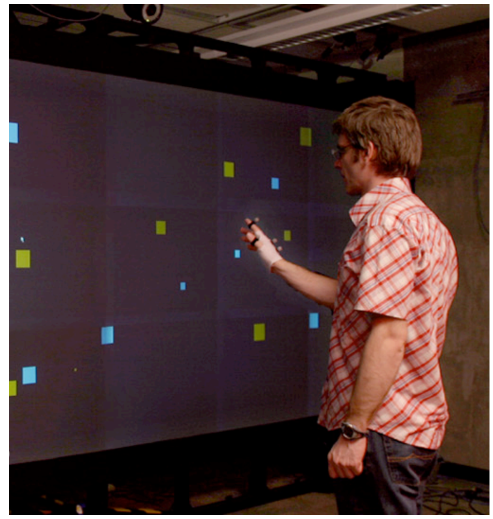


Fig. 3. Freehand gestural interaction with high resolution displays [39]

4.1 Requirements of gesture-based Applications in Automobiles

To use gesture-based applications in automobiles, some special requirements, which result from the automotive environment, need to be taken into account. These requirements and limitations are grouped into user-related and system-related requirements.

4.1.1 User-related Requirements

First of all, the execution space to perform gestures is limited in automobiles. This is a consequence of the sitting position of the driver, as well as the need to keep one hand on the steering wheel. Thus, the freedom of movements is restricted and some types of gestures, such as freehand gestures performed with two hands, cannot be applied [43]. Gestures applicable in automotive environments are discussed in section 4.4.

Due to different drivers, gesture-based applications have to be user-independent. This also results from the possibility that front-seat passengers can control in-car devices, too. For example, the co-driver can control the CD-player while another person is driving. In this case, the system has to recognize both the driver's and the co-driver's gestures. Moreover, a calibration of the system by the user should not be necessary [1].

Furthermore, the number of gesture defined in the interface should be limited and the vocabulary of gestures should be intuitively understandable [1] [2]. This is due to the nature of human-machine communication. In contrast to interpersonal communication, human-machine interaction has to be learned, especially when introducing new interaction possibilities, such as gesture-based input [3]. A small number of gestures can be learned faster than an extensive vocabulary and intuitively understandable gestures can reduce the artificial character of information exchange in human-machine interfaces [1].

4.1.2 System-related Requirements

System-related requirements refer, above all, to varying lighting conditions and unstructured backgrounds [1]. Varying lighting conditions in automotive environments arise as a result of different irradiation angles of the sun due to driving direction and position of the sun. Moreover, environmental lighting can be completely absent, e.g. at night, or at least very limited, for example when driving through a tunnel [14]. Video-based gesture recognition requires a steady background lighting to recognize the position and form of the hand. Furthermore, the contrast between hand and background needs to be sufficient to recognize a gesture at all [1]. Therefore, the system has to be resistant to lighting variations in the environment and lighting determined by the system should not distract the driver.

Another problem of gesture recognition is the detection of the start/end point of a gesture, as well as the distinction between a single gestures and sequences of gestures [21]. Moreover, the optional phases of gestures, preparation and retraction, should not be recognized as gestures. Therefore, it is appropriate to define explicit start and end gestures to frame the interaction with the system's interface.

In addition to varying lighting conditons and general recogniton-related requirements, the system's reaction time plays an important role in gesture-based interaction with in-car devices [1] [14]. As mentioned before, driving demands a person visually, therefore interaction with in-car applications should require only short glances by the driver and control gazes due to missing or delayed system reaction should be prevented [14]. It can be inferred from Geiger et al. [15] that a system reaction time below 100 ms is tolerable.

4.2 Advantages and Disadvantages of gesture-based Interaction in Automobiles

The use of gesture-based interaction with in-car devices has several advantages, as well as disadvantages which are discussed in the following section.

4.2.1 Advantages

In contrast to touch displays or haptical input options, e.g. buttons, gesture-based interaction with in-car devices does not necessarily require eye contact with the system's interface. While touch displays require eye contact to find the requested function on the screen, gestures can be performed in a predefined execution space which does not require eye contact. In combination with acoustic feedback, the user does not have to check for visual system responses. Therefore, gesture-based interaction can reduce the visual and mental distraction of the driver [1]. In addition to that, gesture-based interaction can also be used in noisy environments which is not possible using voice input.

Moreover, gestures are a comfortable and natural form of input known from interpersonal communication [1] [5]. Especially the hand is a natural and flexible input device [21]. Additionally, gestures provide a direct form of interaction without the need to use rotary knobs or buttons [5]. Therefore, mechanical input devices can be conserved [1].

4.2.2 Disadvantages

Disadvantages concerning the use of gestures as input possibility can result from the used gesture vocabulary. Although, gestures are used in interpersonal communication, gesture-based control of human-machine interfaces is not self-explanatory [5] and not intuitively understandable [30]. Therefore, gesture-based interaction has to be learned by the user and the system should provide a help system, e.g. realized by acoustic feedback [15]. Moreover, the user interface has to recognize a large number of gestures to realize a complex system. The problem with an extensive gesture vocabulary is that the user has to remember all gestures and their meaning. This fact has significant influence on the user acceptance of a gesture-based user interface [10].

Another problem with gesture-based interaction in cars is the distinction between user input and gestures occuring during a conversation between vehicle occupants because the recognition system must not interpret the latter as input. Moreover, it is possible that other traffic participant see gestures performed to interact with an in-car infotainment system and misinterpret them as signals or attempt to communicate with them [1].

Furthermore, gesture-based interaction can possibly lead to symptoms of fatigue of hand or arm depending on the number of input actions performed by the user [5].

4.3 Technologies for Gestures Recognition in Automotive Environments

Both video-based and sensor-based systems are suitable for gesture recognition in automotive environments. In this chapter, both methods will be introduced and advantages and disadvantages will be discussed.

4.3.1 Video-based Gesture Recognition

Video-based gesture recognition can be realized by CCD cameras. Since CCD cameras require a constant illumination, a near-infrared lighting source, e.g. a LED-array, and a daylight filter are needed to compensate the varying lighting conditions in automobiles [1] [14]. However, local segmentation meaning the separation of hand and background remains difficult [14]. *Figure 4a* shows the processing sequence of the video-based gesture recognition process.

4.3.2 Sensor-based Gesture Recognition

Sensor-based gesture recognition is better suited for gesture recognition in automotive environments because sensors are independent of varying lighting conditions. Infrared distance measurement sensors are used for sensor-based gesture recognition, thereby different sensors measure the distance to the gesture-performing body part. The local segmentation step works better compared to video-based gesture recognition because the background is farther away from the sensor than the hand and can therefore be blocked out [14]. *Figure 4b* depicts the typical stucture of a sensor-based gesture recognition process.

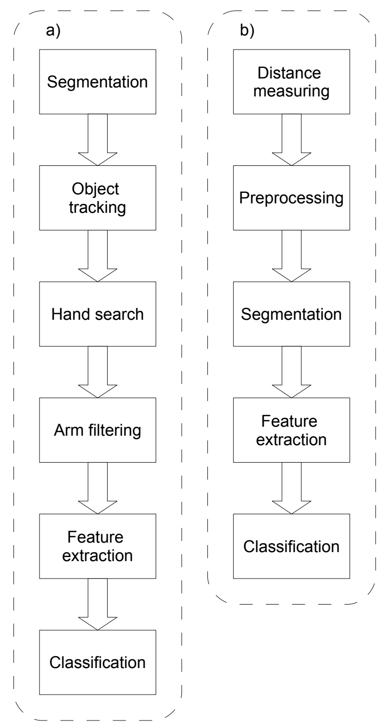


Fig. 4. a) Processing sequence of the video-based gesture recognition process [1]. b) Structure of the sensor-based gesture recognition process [14].

4.4 Use of Gestures in Automobiles

Having introduced special requirements, as well as advantages and disadvantages of gesture-based interaction in automotive environments, this section covers categories of gestures that make sense while driving and scenarios in which gestures are especially helpful.

Gesture-based control of applications can particularly be applied to infotainment systems and telematics[4] [2] [3] [44], as well as general

[4]car information and communication system [2]

vehicle functions and device specific control flows [43]. This includes radio [2], CD-player [45], navigation systems [45], e-mail and telephone [1].

As stated by Geiger [14], dynamic partial-body gestures, especially hand and head gestures, can be adapted to human-machine interaction. However, the use of technical and encoded gestures should be avoided due to their lack of a general applicable gesture vocabulary. Hand gestures on the other hand are limited to one-hand gestures, because one hand should remain on the steering wheel [1] [2]. *Figure 5* summarizes all categories of gestures which are found applicable to automotive environments by Geiger. Following, we will have a closer look at head and hand gestures as input options for in-car devices.

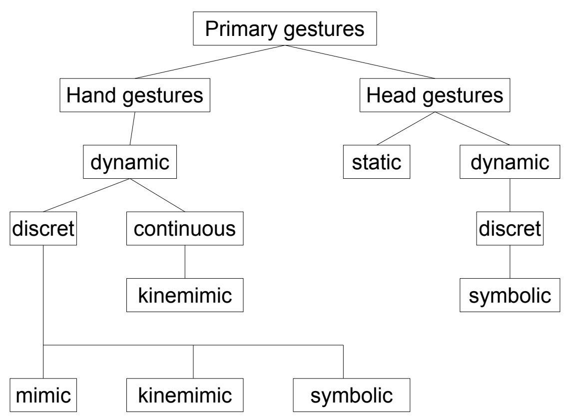


Fig. 5. Applicable gestures for automotive environments [14]

4.4.1 Head gestures

Head gestures are suitable for simple yes-no decisions to communicate agreement or rejection [3] [14]. These can be realized by symbolic gestures, such as nodding and head-shaking, and applied to questions initiated by the system or option dialogs which occur while performing a task. For example, they can be used to accept or reject an incoming telephone call, as well as to agree or disagree with system queries [14]. Besides *"yes"* and *"no"*, a head gesture can stand for *"maybe"* or be used as initial movement to start the system [3].

4.4.2 Hand gestures

Hand gestures are particularly suitable for selection and navigation tasks [3]. The following itemization distinguishes between different categories of gestures and presents possible applications for each category.

Mimic gestures are only partly applicable to automotive environments because they serve to imitate concret objects or events. However, a possible and intuitively understandable application is to accept an incoming call by virtually lifting a telephone receiver [15] [44].

Kinemimic gestures, like waving to the left/right or up and down, can be used to interact with menus and other controls. Kinemimic gestures are further subdivided into continous and discret gestures, continuous ones serve to control steplessly variable controls whereas discret ones can be used to control menus, for example [14].

Symbolic gestures are applied to perform fundamental tasks, such as engaging a device or cancelling [14]. For example, a pointing gestures used in a symbolic way can serve to start an in-car device [44] and wiping can stand for *"cancel"* [15].

Deictic gestures or pointing gestures serve, above all, to select objects or menu items [15]. In contrast to symbolic pointing gestures, which serve to perform a concrete function, deictic gestures are used for selection and navigation. However, Geiger [14] advises against the use of deictic gestures as input option because pointing implies eye contact with the interface and therefore leads to visual distraction.

Hand poses can be used to interact with the system's interface. Hand poses are particularly applicable to be used as start or end gestures. For example, an open palm can serve to activate gestural input. This is necessary because some gestures occur in interpersonal communication, but should not be interpreted by the system [45].

Figure 6 shows an overview of applicable hand gestures for automobiles, as stated by Geiger [14]. The figures a) to f) can be categorized as kinemimic gestures, g) and h) are symbolic gestures and the remaining gestures can be assigned to the category of mimic gestures.

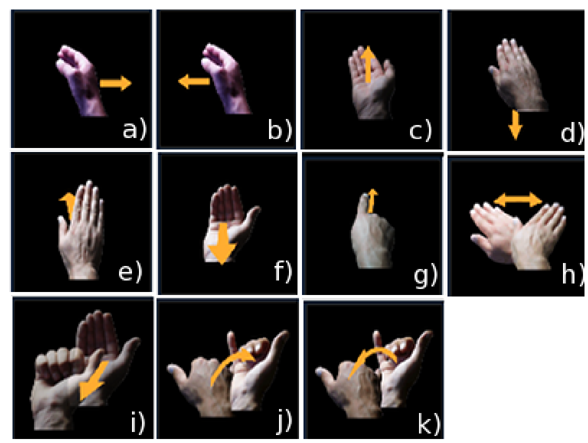


Fig. 6. Overview of applicable hand gestures for automobiles [14]: a+b) Waving to the right/left, c+d) Waving to the top/to the bottom, e+f) Waving forwards/backwards, g) Pointing to the front, h) Horizontal wiping, i) Pulling, j+k) Lifting/Hanging up a receiver

4.5 Existing gesture-based Applications in Automobiles

After presenting categories of gestures which can be applied to automotive environments, this section introduces several applications from the field of research.

4.5.1 Gesture-based Interaction with a Message Storage

Akyol et al. [1] introduce a gesture-driven message storage for automobiles including control options for traffic news, answering machine and e-mail programme. Acoustical messages can be stored and replayed by the system. An overview of the gesture vocabulary can be seen in *figure 7*. The following functions are assigned to these gestures (from left to right): back, forward, interrupt, cancel, show and idle. The latter does not trigger any system reaction.

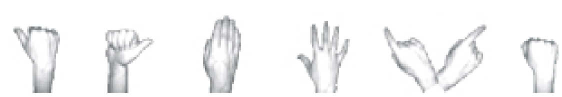


Fig. 7. Gesture vocabulary of the message storage [1]

4.5.2 Gesture-based control of an Infotainment System

Althoff et al. [3] present a gesture-based infotainment system which was implemented in a BMW limousine for demonstration purposes.

Therein, 17 hand and six head gestures are used to control the telephone, the navigation system, the CD-player and the radio. Head gestures (nodding and head-shaking) serve as input options for affirmation and denial. Hand gestures are used for switching, shifting and shortcut functions, for an example see *figure 8*.

Fig. 8. Control of the CD-player: waving to skip between songs [3]

4.5.3 Gesture controlled Man-Machine Interface (GeCoM)

GeCoM is a prototypical gesture-based concept to control devices in automotive environments. The system includes, corresponding to the aforementioned application, radio, CD-player, telephone and navigation system. GeCoM is able to recognize head gestures, continous hand gestures (kinemimics) and discret hand gestures, such as mimics, deictics and symbolics. This is realized by a CCD-camera for image acquisition in combination with a daylight filter and near-infrared LEDs [44] [45]. The interaction with the system is done through horizontal pointing and wiping gestures for back/forward. Vertical movements are used to change control variables, such as volume. *Figure 9* shows the sensor disposal for gesture recognition which is located at the central console of the car.

Additionally, GeCoM supports blindfold-operability by giving acoustical feedback, e.g. in the form of voice output of the selected menu item and signal tones. Moreover, it provides an audiovisual help system, which includes visual presentations and spoken description of implemented gestures [14] [15] [45].

Fig. 9. Sensor disposal of the GeCoM system [14]

5 Discussion

Having introduced gesture-based application areas, as well as special conditions of gesture-based interaction in automobiles, this section examines the transferability of aspects of the aforementioned application areas to automotive environments.

From the field of robotics and teleoperation, no aspects seem to be transferable to cars because there are no robots to control and therefore no robots to train. Moreover, methods for controlling and programming robots are not comparable to interaction possibilities for in-car devices because both systems realize different functionalities.

Gesture-based control of medical instruments in surgery can be compared to gesture-based interaction with in-car devices. In both cases existing devices are controlled by gestural input options. The difference is that medical systems provide support for primary, as well as secondary tasks, whereas automotive systems concentrate on the latter. Concerning primary tasks, gesture-based interaction in cars can be extended to basic vehicle functions. Possible tasks are: controlling windshield wipers by symbolic gestures, indicating direction indicators or opening and closing windows by kinemimic gestures, as well as regulating headlight and interior lighting.

Medical applications concerning therapy purposes are thematically too far away from automotive environments to be applied to them. Furthermore, gesture-based control of transportation devices, like the steering of wheelchairs through hand and head gestures, is not suitable in the context of automotive environments because this technology is not yet mature. As can be seen from [24], the meaning of inaccurate gestures is guessed and the system requires user feedback to correct wrong assumptions. Such a system behaviour can lead to accidents in road traffic due to inaccurate vehicle control. Therefore, gesture-only based control of vehicles is not suitable. On the contrary, such a technique can enable driving for people with impaired walking abilities.

Since joystick-based driving already exists (see [4]), virtual joysticks provide a better approach to gesture-based driving. To which extend this technique are applicable depends on its accuracy. However, gesture-only based driving remains difficult to imagine. In contrast and as mentioned above, gesture-based control of fundamental vehicle functions, such as windshield wipers, direction indicators and headlights, is conceivable.

Although, few gestures are adequate to control infotainment system, it is desirable to extend the gesture vocabulary to operate additional fundamental car functions. Therefore, aspects of sign language recognition systems can be applied to automotive environments since they are able to detect and recognize a large number of gestures.

Due to the sitting position of the driver and the lack of space in cars, the employment of controllers, such as used by Nintendo Wii, is not suitable for automotive environments.

From the area of desktop interaction, interactive menus, like the rapMenu, can be applied to automotive environments. For example, to switch back and forth between different infotainment functions.

To sum up: due to special requirements in automotive environments, the tranferability of aspects of other application areas to them is limited. However, some techniques have potential to be used in cars giving that they meet the security requirements implied by the driving situation.

6 Conclusion

This paper presented an introduction to the use of freehand gestures for interaction in automobiles. Fundamentals of gesture theory were introduced and Geiger's [14] classification of gestures is determined to be most suitable as a fundament for the use in automotive environments.

Following, we presented an overview of application areas using gesture-based interaction, in particular robotics, medical systems, sign language, entertainment, and desktop application, examining whether or not aspects of these application areas can be transferred to automotive environments. It was determined that it is only possible to a limited extent, excluding for virtual joysticks and interactive menus.

In the main part of this work, we presented special conditions in automotive environments, concerning the use of gesture-based interaction, as well as advantages and disadvantages of gesture-based control of in-car devices. Furthermore, two different technologies to realize gesture-based interaction in cars were introduced. Doing so, it was determined that sensor-based gesture recognition is more suitable than video-based gesture recognition due to varying lighting conditions in cars. Moreover, it was examined which categories of gestures make sense while driving. Resulting, head and hand gestures were identified to be applicable. Usable hand gestures can be subdivided into mimic, kinemimic, symbolic and deictic gestures.

In conclusion, gesture-based interaction provides an alternative solution for the control of infotainment systems in automobiles.

REFERENCES

[1] S. Akyol, U. Canzler, K. Bengler, and W. Hahn. Gestensteuerung für fahrzeugbordsysteme. In *Mustererkennung 2000, 22. DAGM-Symposium*, 2000.

[2] M. Alpern and K. Minardo. Developing a car gesture interface for use as a secondary task. In *CHI 2003 extended abstracts on Human factors in computing systems*, 2003.

[3] F. Althoff, R. Lindl, and L. Walchshäusl. Robust multimodal hand- and head gesture recognition for controlling automotive infotainment systems. *VDI Berichte*, 2005.

[4] B. Andonian, W. Rauch, and V. Bhise. Driver steering performance using joystick vs. steering wheel controls. *SAE transactions*, 2003.

[5] T. Baudel and M. Beaudouin-Lafon. Charade: remote control of objects using free-hand gestures. *Communications of the ACM: Special issue on computer augmented environments*, 1993.

[6] R. Boian, A. Sharma, C. Han, A. Merians, G. Burdea, S. Adamovich, M. Recce, M. Tremaine, and H. Poizner. Virtual reality-based post-strokehand rehabilitation virtual reality-based post-strokehand rehabilitation virtual reality-based post-stroke hand rehabilitation virtual reality-bases post-stroke hand rehabilitation. In *Proceedings of Medicine Meets Virtual Reality 2002 Conference*, 2002.

[7] D. Bühler, S. Vignier, P. Heisterkamp, and W. Minker. Safety and operating issues for mobile human-machine interfaces. In *Proceedings of the 8th international conference on Intelligent user interfaces*, 2003.

[8] N. Cairnie, I. W. Ricketts, S. J. McKenna, and G. McAllister. Using finger-pointing to operate secondary controls in automobiles. In *Proceedings of the IEEE Intelligent Vehicles Symposium 2000*, 2000.

[9] S. Calinon and A. Billard. Incremental learning of gestures by imitation in a humanoid robot incremental learning of gestures by imitation in a humanoid robot. In *International Conference on Human-Robot Interaction*, 2007.

[10] R. Ecker, V. Broy, K. Hertzschuch, and A. Butz2. Visual cues supporting direct touch gesture interaction with in-vehicle information systems. In *Proceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2010.

[11] S. S. Fels and G. E. Hinton. Glove-talk: A neural network interface between a data-glove and a speech synthezizer. In *IEEE Transactions on Neural Networks*, 1992.

[12] F. W. Fikkert. *Gesture Interaction at a Distance*. PhD thesis, University of Twente, 2010.

[13] T. Fong, F. Conti, S. Grange, and C. Baur. Novel interfaces for remote driving: gesture, haptic and pda novel interfaces for remote driving: gesture, haptic and pda, 2000.

[14] M. Geiger. *Berührungslose Bedienung von Infotainment-Systemen im Fahrzeug*. PhD thesis, Technische Universität München, 2003.

[15] M. Geiger, R. Nieschulz, M. Zobl, and M. Lang. Gesture-based control concept for in-car devicess. In *VDI-Berichte 1678: USEWARE 2002 Mensch-Maschine-Kommunikation/Design*, 2002.

[16] M. Geiger, M. Zobl, K. Bengler, and M. Lang. Intermodal differences in diistraction effects while controlling automotive user interfaces. In *Proceedings of the International Conference on Human-Computer Interaction*, 2001.

[17] C. Graetzel, T. Fong, S. Grange, and C. Baur. A non-contact mouse for surgeon-computer interaction. *Technology and Health Care, IOS Press*, 2003.

[18] M. Gutierrez, P. Lemoine, D. Thalmann, and F. . ederic Vexo. Telerehabilitation: Controlling haptic virtual environments through handheld interfaces. In *ACM Symposium on Virtual Reality Software and Technology*, 2004.

[19] S. Iba. *Interactive Multi-Modal Robot Programming*. PhD thesis, Garnegie Mellon University, 2004.

[20] J. Ineson, C. C. Parker, and A. Evans. A comparison of head-out and head-in selection mechanisms during simulated flight. Technical report, DERA customer report, 1997.

[21] M. Karam and M. C. Schraefel. A taxonomy of gestures in human computer interaction. In *ACM Transactions on Computer-Human Interactions*, 2005.

[22] A. Kendon. *Gesture: Visible Action as Utterance*. Cambrigde University Press, 2004.

[23] J. Kramer and L. Liefer. The talking glove: An expressive and receptive verbal communication aid for the deaf, deaf-blind, and non-vocal. *ACM SIGCAPH Computers and the Physically Handicapped*, 1988.

[24] Y. Kuno, T. Murashima, N. Shimada, and Y. Shirai. Intelligent wheelchair remotely controlled by interactive gestures. In *5th International Conference on Pattern Recognition (ICPR'00)*, 2000.

[25] G. R. McMillan. The technology and applications of gesture-based control. *RTO educational notes*, 1998.

[26] D. McNeill. *Hand and Mind. What Gestures reveal about Thought.* The University of Chicago Press, 1992.

[27] C. L. Nehaniv. Classifying types of gestures and inferring intent. In *Proceedings of AISB'05 Symposium on Robot Compangions: Hard Problems and Open Challenges in Robot-Human Interaction*, 2005.

[28] T. Ni, R. P. McMahan, and D. A. Bowman. Tech-note: rapmenu: Remote menu selection using freehand gestural input. *3D User Interfaces*, 2008.

[29] K. Nickel and R. Stiefelhagen. Visual recognition of pointing gestures for human–robot interaction. *Image and Vision Computing*, 2006.

[30] R. Nieschulz, M. Geiger, K. Bengler, and M. Lang. An automatic, adaptive help system to support gestural operation of an automotive mmi. In *Proceedings of the International Conference on Human-Computer Interaction*, 2001.

[31] A. Nishikawa, T. Hosoi, K. Koara, D. Negoro, A. Hikita, S. Asano, H. Kakutani, F. Miyazaki, M. .Sekimoto, M. Yasui, Y. Miyake, S. Takiguchi, and M. Monden. Face mouse: A novel human-machine interface for controlling the position of a laparoscope. In *IEEE Transactions on Robotics and Automation*, 2003.

[32] S. C. W. Ong and S. Ranganath. Automatic sign language analysis: A survey and the future beyond lexical meaning. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.

[33] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997.

[34] O. Rogalla, M. Ehrenmann, R. Zollner, R. Becher, and R. Dillmann. Using gesture and speech control for commanding a robot assistant. In *IEEE International Workshop on Robot and Human Interactive Communication*, 2002.

[35] T. Schlömer, B. Poppinga, N. Henze, and S. Boll. Gesture recognition with a wii controller. *Proceedings of the 2nd international conference on Tangible and embedded interaction*, 2008.

[36] T. E. Starner. *Visual Recognition of American Sign Language Using Hidden Markov Models*. PhD thesis, Massachusetts Institute of Technology, 1995.

[37] T. E. Starner, B. Leibe, B. Singletary, and J. Pair. Mind-warping: Towards creating a compelling collaborative augmented reality game. In *Proceedings of the 5th international conference on Intelligent user interfaces*, 2000.

[38] T. Svedström. Gesture interfaces. AALTO University School of Science and Technology, 2010.

[39] D. Vogel and R. Balakrishnan. Distant freehand pointing and clicking on very large, high resolution displays. In *UIST '05 Proceedings of the 18th annual ACM symposium on User interface software and technology*, 2005.

[40] J. Wachs, H. Stern, Y. Edan, M. Gillam, C. Feied, M. Smith, and J. Handler. A hand-gesture sterile tool for browsing mri images in the or. *Journal of the American Medical Informatics Association*, 2008.

[41] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan. Vision-based hand-gesture applications. *Communications of the ACM*, 2011.

[42] K. R. Wheeler and C. C. Jorgensen. Gestures as input: neuroelectric joysticks and keyboards. *Pervasive Computing, IEEE*, 2003.

[43] M. Zobl, M. Geiger, K. Bengler, and M. Lang. A usability study on hand gesture controlled operation of in-car devices. In *Proceedings of HCI International*, 2001.

[44] M. Zobl, M. Geiger, B. Schuller, M. Lang, and G. Rigoll. A real-time system for hand gesture controlled operation of in-car devices. In *ICME '03 Proceedings of the 2003 International Conference on Multimedia and Expo*, 2003.

[45] M. Zobl, R. Nieschulz, M. Geiger, M. Lang, and G. Rigoll. Gesture components for natural interaction with in-car devices. In A. Camurri and G. Volpe, editors, *Gesture-Based Communication in Human-Computer Interaction*. Springer-Verlag Berlin Heidelberg, 2004.

Freehand Gestures for Interactions in Automobiles: a technology-based View

Marion Koelle

Abstract— The work presented covers the use of freehand gestural commands for the interaction with in-car devices. The interaction with traditional in-car devices based on haptic controls often causes distraction effects. Those distraction effects are potentially dangerous for traffic safety, as they anticipate the dictum 'eyes-on-the-road' and 'hands-on-the-wheel'. In contrast gesture-based interfaces draw strength by enabling the driver to interact with the on-board system without eye contact, which increases road safety. Moreover there is strong evidence, that gestural interaction is more intuitive and can easily be adopted. On the other hand gesture-based applications for automobiles have to face specific requirements due to the driving situation as well as to the fabrication and market situation. The main part of this work covers the advantages of gesture-based interaction for the use-case in-car interaction and points out the specific requirements for this application area. Furthermore two existing systems are introduced and analyzed with particularly regard to their compliance of the domain-specific challenges. Additionally this work elaborates on the requirement's technical background and introduces basic principles of gesture recognition.

Index Terms—human-computer-interaction, gesture interfaces, gesture recognition, in-car interaction

1 Introduction

The interaction with traditional in-car devices based on haptic controls often causes distraction effects. Those distraction effects are potentially dangerous for traffic safety, as they anticipate the dictum 'eyes-on-the-road' and 'hands-on-the-wheel'. A total of 5.9% of all accidents can be ascribed to visual distraction by in-car devices[1]. Of these, a percentage of 11.04% is caused by adjusting radio, cassette or CD, 2.89% by adjusting vehicle or climate controls and 1.5% by using/dialing cell phone. Rassl [25] was able to prove, that drivers avert their eyes off the road for up to 16 seconds in order to accomplish tasks demanding visual attention (e.g. the interaction with a navigational system). As the act of driving is a cognitively demanding task, any secondary activity may require only a minimum of cognitive attention. For this reason the aim is to develop interfaces, that do not demand visual attention and that can be controlled in an intuitive an comfortable way.

There have been different approaches in the field of Human Computer Interaction (HCI) to face that challenge by introducing innovative technologies based on the user's natural communication and manipulation skills. Those technologies empower the user to interact with human-computer-interfaces in increasingly natural and intuitive ways. One popular approach is to read the users movements directly and to control the interface by inferring intent from his movements. One of the humans most effective interaction tools is the hand as it can be used for communication with and manipulation of the environment. The human by nature is able to transmit information by performing a wide range of gestures. As conventional input devices, e.g. the mouse, constrain the users ability to use his hand in expressive ways, the use of hand gestures for human computer interfaces has become a very attractive alternative [18].

The work presented is divided into three parts, the first part consists of an academic workup of possible definitions of the term 'gesture'. Furthermore different approaches to classify gestures are covered. This section focusses on taxonomys that have been developed with regard to the exigencies of gesture-based Interfaces in Human-Computer-Interaction (HCI). The second part covers the technical background of gesture recognition. It forms the basis for the subsequent considerations of specific needs for in-car interaction. The third part highlights the advantages of geture-based interaction in automobiles and points out the specific requirements for this application area. Subsequently two existing systems are presented and evaluated concerning *how* the domain-specific requirements are addressed.

2 Conceptual Basics

To introduce the relevant concepts, the term gesture and the classification of gestures are outlined in the following two sections.

2.1 Gestures

The term 'gesture' describes a form of non-verbal communication. In general usage it is defined as "a movement of the hands, arms or head, etc. to express an idea or feeling" (Definition taken from Cambridge Advanced Learners Dictionary [6]). According to Kendon [14] a gesture is defined as a visible action with communicational intent. The presence or absence of a communicative aim is essential for the differentiation between gestures and movements. According to an approach presented in Geiger (2003) [10] gestures can be subdivided into primary and secondary gestures. At this the term 'primary gestures' refers to gestures, that are intentionally used in a communicative way. Whereas 'secondary gestures', though they incidentally do convey information, are not performed consciously and have no communicative intention. Furthermore it can be distinguished between full and partial partial body gestures. In some cases (e.g. pantomime) the whole body serves as communication medium. These are the so-called full body gestures. In contrast movements of a single part of the body are the most common form of gestures used in human communication. These are named partial body gestures [10]. Typical Forms are hand and arm gestures as well as head and face gestures [20]. Regarding the use of gestures as input modality, only primary gestures are considered. Not all types of gestures are applicable for the application scenario 'in-car interaction'. For this reason, the term gesture is limited to one-hand gestures in the following paragraphs. Full body gestures as well as head and face gestures are not considered.

As the humans habitual use of gestures and meaningful interaction is quite complex, there are various scientific approaches that aim to specify this definition more precisely. Whereas psychological and socio-scientific research is concerned with rather semantic topics such as the gesture's role in communication and its co-occurrence with speech in social interaction, in the field of HCI gestures are seen from

- *Marion Koelle is studying Media Informatics at the University of Munich, Germany, E-mail:*
- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

[1] Analysis by Stutts et al (2001, S.11) based on 1995 - 1999 CDS-data obtained from the NHTSA National Center for Statistics and Analysis [27]

a different angle [23, 14, 19]. The general definition of the term 'gesture' is precise by Dorau 2011 [8] with particularly regard to the exigencies of gesture-based HCI-Interfaces. In this scope a gesture is mainly characterized by four features:

1. *spatial and temporal coverage* A gesture is limited to its spatial and temporal coverage. In this context we differ between object-bound and free gestures. The coverage of object-bound gestures is limited to the objects extend, for example a check box on a touch screen. Free gestures do not refer to a specific target but are nevertheless limited due to the interfaces ability to detect gestures and its specifically defined gesture-space.

2. *communicational intent* A movement is, from the performer's point of view, a valid gesture if it is performed with communicational intent. As an illustration: A wiping gesture on a touch screen can be used in order to turn a document's pages as well as to remove cookie crumbs from its surface. Only in the first case we can speak of an communicative intention. Relating to Geiger's [10] definition a valid gesture is consequently a so-called primary gesture.

3. *event profile* A gesture is a characteristic course of motions that is defined through a sequence of events with a specified begin and end. This sequence is called event profile. The event profile is the gestures 'choreography'.

4. *feedback of the system* The systems notifies the user if the gesture had been performed correctly. Every element of the event profile is confirmed by the system through mimic behavior. The feedback signal can be of visual or haptic type. In systems where mainly free gestures are used, the gestures do not refer to a specific object. For this reason, it is challenging to design a user feedback that is clear and coherent [8].

2.2 Gesture Classification

There have been various approaches to classify gesture published by several researchers in computer science. Karam [13] distinguishes between deictic, manipulative and semaphoric Gestures as well as between gesticulation and language gestures. Pavlovic [23] differs between gestures and unintentional movements. A similar taxonomy can be found in Nehaniv (2005) [21]. 'Irrelevant or manipulative gestures' can possibly occur, e.g. when grasping a cup in order to drink its contents. Furthermore 'side effect of expressive behavior' can emerge incidental but without any communicative intention [21].

Geiger's thesis [10] covers gestures of hands and head. He gives a extensive overview over different types of Gestures and subsumes the different types in a schematic diagram (see figure 1). Moreover the different types of gestures are evaluated in regard to their applicability for in-car interaction. Gestures can be classified either dynamic or static. Dynamic gestures are defined by a path of motion which can be either discrete or continuous. Information is mediated by the specific character of the gestures movement. Contrariwise static gestures convey information exclusively through the position of the body part. Zobl et al [33, 34] use a simplified scheme that consists only of hand gesture types that are applicable for the context of gesture-based interaction in automobiles. They propose a set of 22 dynamic gestures and 6 static hand poses as a gesture inventory which is particularly designed for the use case 'in-car interaction'. The gestures contained in this set can be assigned to Geiger's classification groups. In figure 1 those groups of gesture are marked red. As a base for the sequel of this work, these groups are now introduced:

1. *dynamic referencing/deictic gestures* aim to select a certain object in sight. The habitual modus operandi is to point at the intended target [33, 10]. The meaning of the pointing operation depends on the referential value of the selected region or object [19]. It is possible, that deictic gestures are implicit comprised in gestures of other types. Nevertheless many interactions are based on exclusively deictic gestures [13].

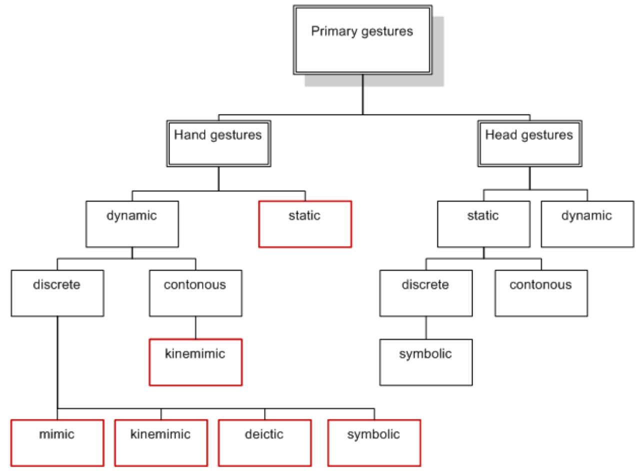


Fig. 1: Schematic gesture classification, adapted from [10] and [34]

2. *kinemimic gestures* are used to illustrate motion by imitating the direction or pattern of a movement. Example: waving to the left/right/up/down [33, 10].

3. *symbolic gestures* aim to visualize abstract ideas or characteristics like emotions or moods. Normally the description of abstract ideas is not linked to a physical object representation. Example: the 'thumbs up' gesture which is used as affirmative answer or to express a positive attitude [33, 10]. As symbolic gestures are conventionalized signals, they are often culture-dependent: The Chinese and the Germans use a different interpretation of a similar set of finger signs for numbers such as 1,2,5 [21].

4. *mimic/pantomimic gestures* imitate typical procedures, certain objects or persons. Example: lifting one hand to illustrate the idea of receiving a phone call [5, 10]. Just as symbolic gestures, mimic gestures are culture-dependent [22].

5. *hand poses* are static gestures. In contrast to the other classes of gestures they are non-dynamic, which means that the hand is not moving and information is conveyed only by its form [10].

3 Technical Background

The following chapter covers the technical background for gesture-based human-computer interaction. The different phases of gesture recognition are addressed and some exemplarily chosen technologies are introduced. This chapter sets out the basis for answering the question which sensoring methods and segmentation methods are suitable for the use-case in-car interaction.

3.1 Gesture Recognition

The term 'gesture recognition' describes the process of detecting intentional movements (gestures) in the systems active zone and identify each as a specific gestural command. The recognition of a gestural command follows basically three steps:

1. *Detection of the intention:* The system is only activated, if the hand is located within the defined interaction space. Gestures (e.g. gesticulation during speech) that are not carried out in the active zone are ignored [4].

2. *Segmentation/feature-extraction:* The process of gesture segmentation aims at obtaining information about the gesture's position, orientation, posture and temporal progression [28]. As a gestural command is described by a specific start- and end position and a sequence of movements in between, the systems task is to recognize the beginning of the gesture as well as the point when the gesture is finished [4, 5].

3. *Classification*[2] Once the start position and the dynamic phase have been recognized by the system, the detected gesture has to be classified [4, 5]. During the process of classification, the features obtained by the proceeding segmentation step are subjected to a number of templates, representing the gestures defined in the system's gesture inventory [12].

If a gesture is sucessfully identified, it is labeled and forwarded to a gesture interpretation module. The recognized gestures are subsequently interpreted and combined with context information [12]. The possibility to integrate context information is in automotive environments especially promising, as a wide range of sensor information is accessible. Useful examples for context knowledge are informations about the current system state (e.g. velocity) and the environment context (e.g. lighting conditions or sound scape). Predefined User preferences could also serve as additional information and can be considered for the systems response, which is initiated afterwards [3].

3.2 Feature extraction

As the captured camera stream is likely to contain not only gestures but also unuseful data such as unintended (hand) movements within the active zone, it is necessary to segment the input frames. Another challenge is to distinguish between different gestures that are typically directly following each other [12]. Kendon [15] describes the ability of humans to segment gestures before recognizing the gesture itself. According to Hassink [12] this is a strong indicator for segmentation being useful. To determine a gesture's position, orientation and posture, precise hand pose tracking is required [5]. Tracking needs to provide real-time access to the six parameters (3 coordinates and 3 angles) that define an objects (here: the hands) position in space. This can be achieved by the use of various sensing devices [18]. The following section introduces different tracking technologies and discusses their principles and background.
According to [26], sensing technologies used for gesture tracking can be classified into

1. *invasive sensing methods* and

2. *non-invasive sensing methods.*

Invasive sensing methods make use of active devices (gloves) or passive markers whereas non-invasive sensing methods do not require modifications of the environment. They are usually based on computer vision [26].

A typical invasive sensing technology are gloves, which measure the flexion of hand and finger joints by dint of goniometers. Such instrumental gloves make use of fiber optics placed on the back of each hand or electrical conductance to retrieve accurate information about the hand pose [26, 5]. The so-called 'Dataglove', the first commercially available tracker, has been described at first in Zimmermann's 'A hand gesture interface device' and had been presented at the CHI-Conference 1987 [31]. Though the dataglove was able to measure finger bends to an accuracy of five to ten degrees, the finger abduction, which is the movement of the fingers to the side, could not be detected [29, 5]. In the following years further glove based input devices that overcome that problem had been developed. Figure 2 shows the Cyberglove which has been developed by Virtual Technologies. Gesture recognition based on wearable gloves provides very accurate results [5] and is applicable for intricate 3D manipulation tasks [16]. A further advantage of datagloves is, that even natural or untrained gestures can be identified. On the other hand they are rather expensive and encumbering as well as they counteract a natural way of interaction [5].

[2]The term (feature) classification in this chapter is used in the technical context and has to be clearly distinguished from the term (gesture) classification (chapter 3.3), which refers to the semantic dimension of gestures.

Fig. 2: Example for an invasive sensing device: the CyberGlove [7]

Contrariwise computer-vision-based interfaces enable the user to interact more naturally, as they do not require any additional instruments attached to the user's hand. Common capturing devices are (stereo) cameras. Promising devices, that are less sensitive to illumination changes are near-IR illuminators, far-IR cameras, capacitive imagers as well as time-of-flight cameras [28]. There are two categories of vision-based tracking methods. The *modelbased* approach creates a 3D-model of the users hand, which is used for tracking and aligned with the hand's pose. The second approach is called *image-based (also: apperancebased)* and calculates the hand's pose based on features that are directly extracted from the source image [5, 23, 17]. Depending on the used approach (model- or imagebased) the parameters applied to describe the hands pose are of different types [23]. The following paragraph describes how these features can be obtained based on the image captured by the camera.

Motion Feature extraction by motion is based on the principle that the movement of the hand is the only motion occurring in the interaction space. As the background in particular is static, dynamic foreground objects can be localized. Thereto the captured camera image is subsequently subtracted frame-per-frame from a learned background image. Systems deploying motion as a discriminatory criterion require certain assumptions. At first, the camera has to be stationary. Secondly, the background has to be static [28].

Depth Approaches relying on depth information to compute the hands pose in the interaction space require specific capturing devices. The required data can be obtained from stereo cameras or direct range (laser) sensors. One major disadvantage of this approach is that it is cost-intensive, as well-calibrated stereo cameras are rather expensive [28].

Color A further method to detect gestures in a captured frame is the distinction by color. The color of human skin is unique and is therefore limited to a certain well-defined area in color spaces [28]. As the most discriminating criteria of skin color is its chromaticity, it is preferable, to operate in color spaces, that separate luminance and chromaticity effectively. A color space model that meets this criterion is the HSV space [30]. Starting from a captured camera frame, it is now searched for areas in the image, that match the chromaticity of skin color. One popular approach is histogram-matching in combination with blob detection [28]. As a consequence of the elimination of the luminance component, a certain degree of robustness to changes in illumination can be achieved. Nevertheless the detection by color is still comparatively sensitive to quickly changing or mixed lighting conditions [30]. One further problem of color-based segmentation methods is, that they are, due to the use of histogram-matching, computationally intensive [23].

Shape The human hand is characterized by its unique shape. Tracking techniques extracting the contours of objects in the images are usually based on edge detecting algorithms which vary in between approaches. Edge detecting processes usually result in a large number of edges containing relevant hand contours as well as irrelevant background contours. To educe the relevant form the irrelevant contours advanced post-processing techniques are essential [30]. Both model- and imagebased approaches can make use of contours as features. An example for a modelbased approach is presented by Gavrila [9]. For this approach multiple edge images of a tree-dimensional hand model, that are a priori synthesized, is used. During the tracking process the edge image of every captured camera frame is compared subsequently to all synthesized edge images of the 3d-model. Chamfer matching is utilized to measure the similarity between synthesized and captured edge images. The process of comparison is a search problem which is solved using the best-fit search algorithm.

As mentioned above, every localization method has certain limitations. To overcome these, in practice often combinations of two or more out of the presented cues are used [23]. The accuracy of e.g. a shape detector, can be improved significantly if it is combined with (skin-) color or motion cues [30].

3.3 Feature classification

The feature data received by segmentation in the previous phase has to be identified as a specific gesture and accordingly labeled by classification. As mentioned above (chapter 3.1) gestures are consisting of a startpositon, a dynamic phase and an endpose. For classification the dynamic phase is interpreted as a sequence of states. Each state corresponds to a static posture. The basic principle of the classification phase is to evaluate the probability that an captured sequence of states matches a certain gesture [12].

A common approach for feature classification are Hidden Markov Models (HMMs), which have been originally developed for automatic speech recognition. The basic principle of HMMs for gesture recognition is, that every gesture is represented by a specific sequence of states. Each state corresponds to a certain hand pose [3]. Each gesture pattern which is comprised by the system's gesture inventory is represented by a differentiated HMM. It is typical for a HMM, that the states are not directly observable, but can be deduced from observable parameters. In a gesture recognition system, those parameters can be obtained by interpreting the features obtained trough segmentation [30].

For the recognition of an temporal process (like e.g. a gesture) so-called semicontinuous HMMs are used. Semicontinuous means in this context, that from a current state s_i only some states $s_{i+j}|0 < j < k$ are reachable. The amount of reachable states is referred as k. The transitions a_{ij} equate the probability to migrate from state s_i at time t to state s_j. An example of a semicontinuous HMM is shown in figure 3. Every gesture is represented by a stochastic model (HMM), deduced from its trajectory, velocity and hand form.

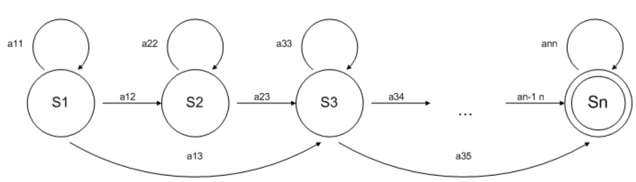


Fig. 3: Left-right topology of a semicontinuous Hidden Markov Model (HMM), adapted from [20, 3]

Every incoming gesture is converted into a sequence of feature vectors $v_1 ... v_s$, each describing the change in position and orientation of the hand. To classify an observed sequence O from an input sequence the probability $P(O,\lambda)$ has to be determined for every defined HMM. $P(O,\lambda)$ is the probability that the corresponding HMM constitutes the observed gesture. The input sequence is identified and labeled as the HMM with the highest probability [3, 20]. It is possible to vary the predefined HMMs depending on the system context, e.g. previous gestures or cross-modal information [30]. The use of HMMs for feature classification furthermore allows to retain "garbage-classes". This method enables the rejection of non-gestural or unintended movements [26].

Another promising approach to classify an input sequence is the Dynamic Time Warping algorithm (DTW). The DTW is a distance classifier. Alike the HMM approach, the DTW algorithm compares the captured input sequence to a range of stored reference patterns and computes a confidence measure (score) that defines the probability that the observed pattern represents a certain gesture class. The special feature of the DTW approach is, that it is able to cope with non linear temporal distortion. This is important as the duration and execution dynamics of the same gesture pattern can vary, even if it is carried out by the same user. This factor is considered during the classification process. As a result the major advantage of the DTW algorithm is, that it requires a far smaller number of reference patterns per class than the HMMs. Therefore the training expenses can be minimized [10].

4 Gestures for in-car Interaction

The utilization of hand postures in automotive environments has several benefits as well as it requires specific prerequisites. The following chapter goes into details about these aspects and presents two gesture-based systems. Subsequently both systems are evaluated concerning how the domain-specific requirements are addressed.

4.1 Advantages and Benefits

The use of gesture based interaction in automotive environments has considerable advantages over traditional input methods. Compared to interfaces based on speech interaction, gesture based input can even be used in noisy environments. Especially in the context of driving, where the load by ambient noise is experiential high, this is a convincing benefit [3].

Unlike the use of conventional controls, gesture based interaction requires less visual and mental attention [1] and can therefore be used as a secondary task. The concentration on primary tasks like steering, accelerating and braking is not adversely affected [3]. A Study conducted by M. Geiger in 2003 could proof, that gesture based input causes significantly less distraction effects than haptic input. Geiger's analysis measured the modality-specific control error during the accomplishment of different secondary tasks. The control error of the haptic control exceeds the gestural input modality by 36%. In addition, participants were asked to indicate their subjective estimation of the distraction effect: 94% of the participants perceived the gestural input as less distracting [10]. This result could be confirmed in a further study by M. Alpern [2]. The measurement showed, that drivers using gesture based input make less total errors than a control group using a physical interface. Additional qualitative interviews also support Geiger's results, as can be discerned from this statement of a participant:

> "[The gesture interface] helped me keep my attention on the driving more because I didnt have to take my eyes off the road."

Gesture based interaction is more natural and comfortable than haptic controls [1]. Moreover a better user acceptance is supposable. Geiger's [10] interviews allow to draw conclusions on the user acceptance as 76% of the participants rated the gestural input as the more convenient input method. Gestural commands require less time than haptic commands. Geiger [10] could proof, that physical input is significantly slower - by a factor of 1,4 - than gesture input.

4.2 Specific Requirements

The practical use of gesture recognition for interaction with in-car devices demands domain-specific conditions.

Robustness   As a result of the driving situation, the periphery of those interfaces is characterized by intensive changes in illumination: the range includes direct sunlight as well as sparse lightning during night rides; the direction and color of the incident light is variable and abrupt changes of illumination can possibly occur (e.g. when entering or leaving a tunnel). For this specific use-case it is of particularly interest that the detection is robust against any interference in the environment [1, 10].

Non perceptible lighting   Furthermore the distraction of the driver or other traffic participants has to be avoided. As additional lighting could possibly deflect the driver, any extra illumination used for detection has to be not perceptible to the user [1].

Limited interaction space   In order not to impair road safety the interaction space has to be limited. Gestures performed in sight of other traffic participants could possibly lead to distraction effects. Consequently the execution area has to be not perceptible to other traffic participants on the one hand and easily accessible for the driver on the other hand. Figure 4 shows an appropriate interaction space which is situated on the center console [1].

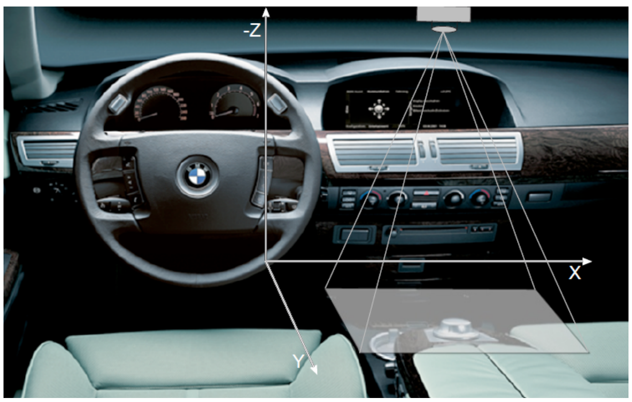


Fig. 4: Example of an adequate interaction space with the sensing device positioned at the roof lining [3]

Latency   The systems response time (latency) has to be minimized due to two factors. First a high latency makes the system unacceptable for practical purpose and affects the users acceptance negatively [1, 28]. Second a minimal latency is an important prerequisite for the traffic safety. While in most domains, the response time is only effective on the acceptance, in automobiles the absence of a system reaction is a potential source of danger. Studies [11] could proof, that a few 100 ms without a system response seduce the driver to control views and distract him from traffic. As a consequence the challenge to face with systems for in-car gesture interaction is to process the detected gestures in real-time.

Intuitivity   The gesture inventory has to be intuitive and natural [1]. As most drivers are used to the traditional input devices, like adjusting knobs, buttons and other haptic controls, they are often unwilling to adopt alternative interfaces that require a period of learning. To minimize the required adoption-time, the gesture inventory must be designed in a way that makes the gestural commands easy to perform and remember. Complex hand shapes or unnatural finger configurations are difficult to learn and unlikely to be remembered. Therefore it is advisable, to build on the users preexisting knowledge and to make use of gesture patterns that are familiar and intuitive to the user. In addition an ideal gesture pattern has to be simple and temporally short to minimize the user's mental load [28]. The factors 'intuitiveness'and 'ease of learning' are in the use-case interaction in automobiles of particular importance. Unlike e.g. the player of a gesture-based computer game, the driver uses the interface not as an end in itself and does not want to spent time to assimilate.

Profitability and market acceptance   Beside aspects of safety and usability, economic aspects have to be considered. From the vehicle manufacturers as well as from the customers perspective it is important, that the costs for additional components for gesture detection do not exceed a certain range [10]. The market acceptance of the product depends on the price the customer would pay for additional features. "The cost of more advanced sensors and sensor setups must be weighed against any potential performance benefit." [28] Furthermore the limited space on offer in the vehicle has to be considered. As the existing electronic components already occupy space, the remaining zone is constricted. Therefore the hardware components of a gesture detection system may only require a minimum of space [10].

'Come as you are'   A further prerequisite for the applicability of a gesture-based system for the standard integration in automobiles is the 'Come as you are' aspect. The driver should be able to control the interface without any additional instruments. He does not have to wear markers, gloves, or long sleeves. Consequently, a system used for gesture recognition in an vehicle has to be based on a tracking solution that is non-invasive [28].

4.3 Systems and Application Scenarios

iGest   The in-car application *iGest* developed by Akyol et al [1] is a gesture-controlled message store, which provides access to audio messages from different sources. The system stores incoming messages such as traffic news, e-mail and messages recorded by a phone answering machine. The driver can interact with the storage by gestures and play the archived messages via a speech synthesizer. It is possible to control the application exclusively via gestures, no eye contact is required. Nevertheless, the system contains an additional screen, that displays the identifier of the current message. The visual notification furthermore aims to facilitate the familiarization of first-time users. The user can play the stored messages sequentially or navigate within the archive using the 'next'- or 'back'- gestures or alternatively point 'to-the-left' or 'to-the-right'. The 'stop' gesture stops the playback. The interaction mode can be ended by signalizing 'chancel'. The 'idle' gesture pattern serves as garbage model. If it is detected no system response occurs. Figure 5 provides an overview of the uses hand poses and the assigned functions.

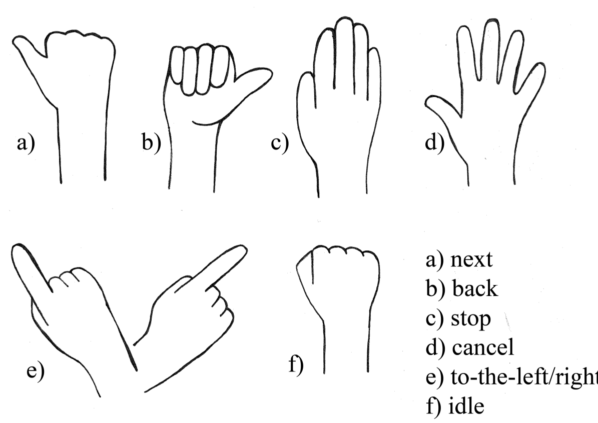


Fig. 5: iGest: hand poses and assigned functions, adapted from [1]

GeCom   The *GeCom* (Gesture Controlled MMI) system, presented by Geiger [10], is a MMI that covers domain-typical infotainment components such as audio, navigation and phone. The interface is controlled via discrete and continuous hand gestures in combination with head gestures [3]. Figure 6 shows the system integrated into the center console. The system utilizes acoustic feedback (speech output as well as signal tones) to inform the user about recognized gestures and to provide operability without eye contact. Selected menu-items are read out aloud. Additionally tones are used to signalize the begin and end of a gestural input. This method aims to anticipate control views by the driver, which are - as already mentioned in chapter 4.2 - a potential danger of distraction. All control options are accessible via gesture input as well as conventional haptic controls. The user can chose his preferred input modality at will or depending on the situation. Haptical input using a console terminal is also possible. GeCom can be classified as a *multimodal* operational concept, as it uses multiple communication channels. The application is based on an inventory of discrete and continuous hand gestures, which had been developed by M. Zobl (see [32] and [33]). The MMI instruments can be accessed by *indirect manipulation* modes and *direct manipulation* modes. We speak of *indirect manipulation* if the execution of a certain discrete gesture results in exactly one specific system response. On the other hand *direct manipulation* describes the use of continuous hand gestures to manipulate one specific parameter (e.g. the musics volume). Moreover the systems ability to recognize head gestures enables the driver to answer questions (e.g. "Start navigational unit?") in an intuitive way by nodding or shaking the head. The system integrates twelve hand gesture pattern (eleven gestural commands and one action) as well as eight head gesture classes (2 gestural commands and six actions) [33, 34, 10]. According to Geiger [10] an *action* is an garbage class that does not trigger a system response. The occurrence of those patterns is very frequent during driving. From a technical view, *actions* are treated like gestural commands, they only differ in their impact on the system's response. The GeCom system defines e.g. the actions "tripping the gear shift" as a hand gesture or "shoulder check" as a head gesture. Table 1 contains an exemplary excerpt of the gesture inventory for hand gestures. The full list can be taken from [10], page 143.

Table 1: Gesture inventory of GeCom (excerpt) [10]

Gesture (classification)	GeCom response
"Wave to the right" (kinemimic)	Navigate to the next menu item
"Wave to the left" (kinemimic)	Navigate to the previous menu item
"Point" (deictic)	Select current menu item
"pick up virtual phone" (mimic)	Accept incoming call
"tripping the gear shift" (action)	none

Furthermore, an adaptive help system is provided. The primary aim is to assist drivers which are not yet used to gesture control by visualizing the gesture patterns on the display as well as explaining them via speech output.

4.4 Comparison and Evaluation

In the following the systems presented in chapter 4.3 are compared with regards to the specific requirements of in-car interaction. The assessment criteria is based on the exigencies described in chapter 4.2. In detail, those are:

1. Robust tracking under driving conditions
2. Real-time capability
3. Intuitiveness and ease-of-learning

[3] Head gestures are addressed in the system's description for reasons of completeness. The evaluation and comparison to other systems (see chapter 4.4) will focus on hand gestures.

Fig. 6: Integrated GeCom system with IR sensor array (yellow) [10]

Both systems have been evaluated in depth regarding applicability for in-car interaction and feasibility during development. The aim of this analysis is not to verify *if* those requirements are satisfied but rather to point out *how* the systems face those challenges.

Robust tracking To consider the aspect robust tracking under driving conditions, we have a look in detail at the image acquisition method as well as at the system-specific illumination. The capturing devices embedded into the *iGest* system is a modified customary CCD-camera. *iGest* additionally makes use of a LED-array in the near-IR spectrum (950nm). This illumination in not perceptible to humans, therefore the distraction of the driver is avoided. The additional illumination is necessary as CCDs (charge coupled devices) are not sufficiently capable of processing varying lighting conditions. Additionally a daylight filter is applied and intensity fluctuations are compensated by the use of an automatic range control. A major disadvantage of this image acquisition method is the lack of color information due to the near-IR illumination. As a result color-based segmentation can not be utilized. Tests showed that incorrect segmentation caused by direct sunlight or overlapping objects can occasionally emerge [1].
Contrariwise the *GeCom* system avoids the use of a camera. Instead an array of distance measurement sensors operating in the infrared spectrum is utilized. The sensor array gathers 3D-information about objects in the the interactions space. The distance values are measured by triangulation. The result forwarded to the gesture recognition system consists of quasi-continuous range-data for every pixel. The sensor array is integrated into the gearshift leaver and covers the interaction space in front of the center console (see figure 6). The image acquisition method used by GeCom is robust against sudden changes in illumination, the objects color as well as the appearance of shadows. A further benefit is, that range-data can be processed with comparatively low computational effort, which enables real-time capability [10].

Real-time capability As stated in chapter 4.2, the real-time capability of an in-car gesture interface is essential for traffic safety. Geiger [10] showed that a response time > 100ms is potentially dangerous. The *GeCom* system had been tested with regard to its response time. Thereby a typical driving situation was assumed. As most drivers are listening to music while driving, a multimedia application was running in the background. Due to this, the computers processor load was relatively high. Nevertheless it could be proved, that even under those circumstances response times in real-time (< 100ms) are achieved. The *GeCom* system achieves its real-time operability by computing the segmentation directly based on range-data and by utilizing feature classification based on the DTW-approach, which had been introduced in chapter 3.3. The DTW (Dynamic-Time-Warping) approach measures the similarity between sequences that can vary in speed or time. It is less computationally intensive than approaches based on HMMs and requires less training [10].
The tracking method of the *iGest* system reduces computation efforts

Table 2: Comparison of the systems iGest and GeCom (summary)

System	gesture types	robust image acquisition	real-time capability	intuitiveness and ease-of-learning
iGest [1]	static hand poses	• modified CCD-camera • additional near-IR lighting	• preselection of objects	• familiar interface design • visual help system
GeCom [10]	dynamic hand gestures	• Infrared distance measurement sensors	• range-data acquisition • DWT-classification	• adaptive help system • audiovisual support

by making a preselection of objects in the active zone. Objects, that are unlikely to be a human hand are excluded in advance [1]. For the *iGest* system no empiric data is available. For this reason it is not possible to draw valid conclusions concerning its real-time capability.

Intuitiveness and ease-of-learning Especially for the use-case in-car interaction the intuitiveness and ease-of-learning of the application is essential as the users primary interest is driving. (see chapter 4.2) The use concept of the *iGest* system is derived from a radio's type of use. The application builds on the user's preexisting knowledge and supplements it with gestural control. Additionally the system is designed to assist the unexperienced user. Graphical information containing the current position in the archive as well as the gesture inventory is presented on a display to help on familiarization. If a certain gesture is recognized the corresponding graphical representation is highlighted. A redraw of that system is that acoustical feedback is not provided. Furthermore the system's intuitiveness could be improved if instead or in addition to static hand postures also dynamic gestures - which are more natural - can be recognized [1].
The *GeCom* system includes an adaptive help system, which provides contextual information. The system is adaptive as it takes account of the user's current state of knowledge and his training level. The aids are presented in graphical form and as audio information. If a gestural command is recognized it is confirmed by an signal tone. The gestural commands themselves are not visualized. Nevertheless, extensive studies have been conducted, to design the UI in an intuitive way, that illustrates the gestural commands at the best. At this gestural commands are visualized by graphical tools, that represent the gestures nature [10].

All in all can be noted, that the two presented systems use different approaches to face the domain-specific requirements. Both systems provide a substantiated base for future scientific research as well as for possible commercial development. An overview of the two systems and their specific methods of resolution is provided in Table 2.

5 Summary and Future Prospects

Recapitulatory there is to say that gesture-based in-car devices involve significant benefits, including an intuitive and natural handling, good learnability as well as the possibility to increase traffic safety by avoiding visual distraction. Despite these advantages and though there has been fruitful research in this field , only few prototypes of in-vehicle systems based on gestural input have been developed yet, and no system made it to serial production and the commercial market. One can only speculate about possible explanations. Some of the aspects presented in chapter 4.2 could possibly shed light on this question. One point that has to be investigated is economic efficiency. The prerequisite for successful market entry are profitability and user acceptance. Regarding that issue Pickering stated:

> "The rate of introduction of any automotive gesture recognition system is more likely to be dictated by the rate of user acceptability and not the timing of technical issue resolution." [24]

For this reason it is likely that future research concerning non-contact gesture-based vehicle interaction will focus on increasing the usability and user acceptance of those systems. The research focus will be moved from the field of computer vision, where the technical foundations were developed, into the area of usability research. The aim has to be to develop a marketable, user-optimized system, that convinces customers and manufacturers in equal measure.

References

[1] S. Akyol, U. Canzler, K. Bengler, and W. Hahn. Gestensteuerung für Fahrzeugbordsysteme. In *Mustererkennung 2000, 22. DAGM-Symposium*, pages 139–146, London, UK, 2000. Springer-Verlag.

[2] M. Alpern and K. Minardo. Developing a Car Gesture Interface for Use as a Secondary Task. In *CHI EA '03 CHI '03 extended abstracts on Human factors in computing systems*, pages 932–933, New York, US, 2003.

[3] F. Althoff, R. Lindl, and L. Walchshusl. Robust Multimodal Hand- and Head Gesture Recognition for Controlling Automotive Infotainment Systems. *VDI-Tagung: Der Fahrer im 21. Jahrhundert*, 2005.

[4] T. Baudel and M. Beaudouin-Lafon. CHARADE: Remote Control of Objects using FreeHand Gestures. *Communications of the ACM*, 36:28–35, 1993.

[5] W. Buxton and M. Billighurst. *Gesture Based Interaction*. 2009. (to appear).

[6] Cambridge University Press. Cambridge advanced learner's dictionary, 2011. accessed on 2011/07/13.

[7] Cyber Glove Systems. CyberGlove II, 2011. accessed on 2011/07/13.

[8] R. Dorau. *Emotionales Interaktionsdesign*. SpringerVerlag, Heidelberg, Germany, 2011.

[9] D. M. Gavrila and L. S. Davis. Towards 3-D Model-Based Tracking and Recognition of Human Movement: a Multi-View Approach. In *In International Workshop on Automatic Face- and Gesture-Recognition. IEEE Computer Society*, pages 272–277, 1995.

[10] M. Geiger. *Berührungslose Bedienung von Infotainment-Systemen im Fahrzeug*. PhD thesis, Technische Universität Mnchen, München, Germany, October 2003.

[11] M. Geiger, R. Nieschulz, M. Zobl, and M. Langl. Bedienkonzept zur Gestenbasierten Interaktion mit Geräten im Automobil - Gesture-Based Control Concept for In-Car Devices. *USEWARE 2002 Mensch-Maschine-Kommunikation/Design*, 1678:299–303, 2002.

[12] N. Hassink and M. Schopman. Gesture recognition in a meeting environment. Master's thesis, University of Twente, Eschede, Netherlands, February 2006.

[13] M. Karam and M. C. Schraefel. A Taxonomy of Gestures in Human Computer Interactions. Technical Report TR ECSTR-IAM05-009, University of Southampton, 2005.

[14] A. Kendon. *Current Issues in the Study of Gesture*, pages 32–46. Lawrence Erlbaum Associates, Hillensdale, Newjesey, US, 1986.

[15] A. Kendon. *How gestures can become like words*, pages 131–141. C.J. Hogrefe, Toronto, US, 1988.

[16] S. Lenman, L. Bretzner, and B. Thuresson. Computer Vision Based Hand Gesture Interfaces for Human-Computer Interaction. Technical Report CID-172, Center for User Oriented IT Design, 2005.

[17] J. Lin. *Visual Hand Tracking and Gesture Analysis*. PhD thesis, University of Illinois, Urbana, Illinois, US, 2004.

[18] G. R. McMillan. The technology and applications of gesture-based control. *RTO educational notes*, 3, 1998.

[19] D. Mcneill. *Hand and Mind: what gestures reveal about thought*. University of Chicago Press, Chicago, US, 1992.

[20] S. Mitra and T. Acharya. Gesture Recognition: A Survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37:311–324, 2007.

[21] C. L. Nehaniv. Classifying types of gestures and inferring intent. In *Proceedings of AISB'05 Symposium on Robot Compangions: Hard Problems and Open Challenges in Robot-Human Interaction*, pages 74–81, Hatfield, England, 2005.

[22] J.-L. Nespoulous and A. R. Lecours. *Gestures: Nature and Function*, pages 49–62. Lawrence Erlbaum Associates, Hillensdale, New Jersey, US, 1986.

[23] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:677–695, 1997.

[24] C. A. Pickering, K. J. Burnham, and M. J. Richardson. A research study of hand gesture recognition technologies and applications for human vehicle interaction. pages 15–30, 2007.

[25] R. Rassl. *Ablenkungswirkung tertiärer Aufgaben im Pkw - Systemergonomische Analyse und Prognose*. PhD thesis, Technische Universität München, München, Germany, July 2004.

[26] T. Sowa. The recognition and comprehension of hand gestures: a review and research agenda. In *Proceedings of the Embodied communication in humans and machines, 2nd ZiF research group international conference on Modeling communication with robots and virtual humans*, ZiF'06, pages 38–56, Berlin, Heidelberg, Germay, 2008. Springer-Verlag.

[27] J. Stutts and D. Reinfurt. The role of driver distraction in traffic crashes. Technical Report 202/638-5944, AAA Foundation for Traffic Safety, Washington, DC, US, 2001.

[28] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan. Vision-based hand-gesture applications. *Commun. ACM*, 54:60–71, February 2011.

[29] S. Wise, W. Gardner, E. Sabelman, E. Valainis, Y. Wong, K. Glass, and J. Rosen. Evaluation of a fiber optic glove for semi-automated goniometric measurements. *Journal on Rehabilitation Research and Development*, 27:411–424, 1990.

[30] X. Zabulis, H. Baltzakis, and A. Argyros. *Vision-based hand gesture recognition for human-computer interaction*. Lawrence Erlbaum Associates, Inc (LEA), 2009.

[31] T. G. Zimmerman, J. Lanier, C. Blanchard, S. Bryson, and Y. Harvill. A hand gesture interface device. In *Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface*, CHI '87, pages 189–192, New York, NY, USA, 1987. ACM.

[32] M. Zobl, M. Geiger, K. Bengler, and M. Lang. A usability study on hand gesture controlled operation of in-car devices. In *Human-Computer Interaction*, New Orleans, Louisiana, USA, 2001.

[33] M. Zobl, M. Geiger, B. Schuller, M. Lang, and G. Rigoll. A real-time system for hand gesture controlled operation of in-car devices. *Multimedia and Expo, IEEE International Conference on*, 3:541–544, 2003.

[34] M. Zobl, R. Nieschulz, M. Geiger, M. Lang, and G. Rigoll. Gesture components for natural interaction with in-car devices. In *International Gesture Workshop in Gesture-Based Communication in Human-Computer Interaction*, pages 448–459, Gif-sur-Yvette, France, 2004.

Energy awareness through ubiquitous computing in modern households

Felix Reitberger

Abstract— While the limited availability of non-renewable resources like oil, gas and coal is known at least since the 1970ies and the risk of nuclear power is present in the media after the Fukushima reactor disaster, the issue of energy saving is a good example of the "not in my backyard" mindset. In theory, we like the idea of saving energy, but we are not aware which concrete steps we could take or, on the other hand, not willing to take them as this would require some effort and rethinking. This problem of indecision and lack of awareness might be solved by ubiquitous computing as we can provide several variants of feedback for the user and thus teach him how to save energy. In this paper, I will evaluate several existing and potential devices using criteria defined by Anton Gustafsson. These include subjective aspects like aesthetics and social interaction but also measurable and therefore objective aspects of time effort, cost and the amount of energy saved. Some of the devices are well known because they are already available on the market. Others are prototypes only evaluated in single studies. After drawing conclusions evaluating these two groups, I will propose designs for fictional devices which excel existing devices in single aspects. The coincounter for example removes the abstract unit "Watt" from the information display and shows energy consumption as "money". They remain without evaluation for now as they might be built in the future.

Index Terms—Energy Awareness, Devices for Energy measurement, User Feedback

1 Introduction

Do you know how much electricity are you using right now? Lets say in your room? There are some easy calculations. One compact fluorescent lamp uses 8-12 Watt and there are only two states "on" and "off". But most appliances' consumption is not that easy to calculate. The consumption of a fridge for example varies strongly. Newer fridges tend to consume less electricity then older ones, but the consumption also depends on the user interaction. A fridge with open doors and maximal cooling enabled consumes significantly more electricity than a fridge with low cooling and closed doors. So the central question this paper deals with is: How can ubiquitous computing raise energy awareness and make people change their habits in consuming energy with the main intention and goal being to save precious energy on a long term basis?

In this paper, I mainly refer to and discuss devices and ideas outlined by Anton Gustafsson as research done regarding sources dealing with this topic led to the conclusion that Gustafssons work currently is the most comprehensive one for designing new devices because it defines criteria that can be used to evaluate new devices.

We can find out from different publications the carbon footprint of our several activities (dependent on provider, for example in the "how dirty is your data?" report [6]). Valid information on plain energy electricity consumption on the other hand is harder to get. We need to know how much energy is being consumed right now in order to be able to tell whether consumption should be reduced and also because "awareness" starts with being interested in, feeling concerned and knowing the consequences, in this case of one's own behavior. This is why we, assuming that change is always triggered by an individual's awareness and willingness, focus on the home and not on the commercial sector. In the commercial sector thIn a second step, we will try and show ways of easing energy saving and thus creating a more sustainable way of living. Let us stick to the fridge example. Building a fridge with smaller doors will most likely result in a reduced consumption [9]. It will not result in raised energy awareness, though, except if the door talks to us after being open for to long. Here we can see why this example contrasts with the positive feedback design criteria. A fridge door that yells at us is mostly annoying and will lead to a variety of behaviors answering that. So we stick to the approach of Gustafsson providing "positive feedback" or if that is not enough create a strong emotional attachment to the device [12]. With this example we illustrated the difficulty of designing household devices that raise energy awareness, but let us return to the beginning of the thought process. Be it that we want to know how much energy the devices used in our household consume. Nowadays, we have two commonly known ways of finding this out. Either we use the product information itself or we measure the energy those devices consume. Product information may be available on several levels of detail. The data sheet or the handbook may contain accurate information about the minimal / maximal consumption, the EU Label [16] provides easy to use information. These labels suggest that "green" is good and "red" is bad which is a well known color scheme from traffic lights. Measuring, on the other hand, is not as simple in terms of time efficiency or cost efficiency. There are devices which can be plugged in in order to then measure the current drain, but data filing is time consuming. Therefore, it is hard to maintain a regular schedule while measuring energy consumption as the latter requires 24 hours-monitoring. The awareness resulting from this kind of monitoring is very abstract, we get a number of Watt. So we can learn that our fridge consumes more after it lost cold because we opened the door or turned up the oven next to it. We already receive delayed feedback regarding our energy consumption with our utility bill, but it is to "inconspicuous" to have any persuasive effect towards changing our minds so that we revise our ways of consuming energy. This shows the necessity of interfaces providing a better (meaning more persuasive) feedback. Regarding energy awareness in general, Gustafsson states that in domestic environments, we seldom recognize that appliances are consuming electrical energy. Since the home appliances do not communicate their consumption behavior, it becomes difficult to understand and make rational decisions about energy-related issues [7] p. 156.

Although this might hold true for some devices in our households, people do pay attention to their electrical bills especially when they are unusually high or consciously by, for example, a refrigerator promoted as consuming relatively little energy. But there definitely is room for improvement for both people who already have a certain level of energy awareness and those who become aware of consuming energy in their homes only when being deprived of it, for instance in the rare event of a blackout. In the following, I would like to define some of the most important terms used here and thus also the scope of this paper's discussion.

- *Felix Reitberger is studying Media Informatics at the LMU Munich, Germany, E-mail: felix.reitberger@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

1.1 Definitions

In this section the design goals are defined. We want to accomplish energy saving as a behavioral pattern. And we are only allowed to design sustainable devices and use positive feedback.

1.1.1 energy consumption

The term "energy consumption" requires a definition because "energy" in the original meaning cannot be consumed only transitioned from one form into another [11]. For this paper we assume energy equals electricity and therefore deal with the consumption of electricity and ways of reducing it.

1.1.2 energy saving

"Energy saving" is the reduction of energy consumption. It should be achieved without loss of functionality of the devices concerned, start "at the bottom", i.e. in the single household and be based on awareness regarding individual energy needs and consumption which very often are not in line. This implies that energy saving not necessarily means deprivation or loss of comfort, but might, on the contrary, even be a pleasure or offer satisfaction once a certain degree of awareness and an understanding of what "sustainability" means, is achieved. Applications like the Shower that visualizes the amount of water used [10] do not fall in the category of this paper because they focus on water consumption.

1.1.3 sustainability

Another important concept regarding reasonable energy consumption is "sustainability". The term is derived from the Latin verb "sustinere", meaning to maintain" or "endure". Since the 1980s, the nowadays commonly known concept, also used as "sustainable development" refers to a way of living and thinking which attempts to combine the fulfillment of needs of the present without compromising the possibility of future generations to meet their own needs. Energy saving is an integral part of this concept[14].

1.1.4 feedback

One of the big gains from additional devices in the household that should help to save energy is the possibility of real time feedback. More frequent feedback is more convincing both on a longterm and on a shortterm basis [17]. So today feedback can be seen as a learning tool for energy users allowing them to teach themselves [2]. In order to achieve that energy consumption is based on awareness and energy is not just carelessly wasted, Gustafsson holds that the user should receive a "reward" for saving energy and being aware of his energy consumption. This can be entertainment, decorative value or self-reflection [7]. The purpose of this so called "positive feedback" is the acceptance of an additional device in the household and furthermore, persuading the user to save energy. In another context, negative feedback might provide a powerful leverage to force energy saving [8], but we would rather encourage the user to do so (passive feedback) than forcing him into it as negative feedback. Fining or punishing people in different ways, definitely is efficient. But the more important aspect of it is that we are discussing design criteria. Who would like a fridge that yells at you every time you leave the door open too long [9]? In exceptional cases we accept negative feedback from a machine telling us what to do. In a car or an airplane for example where the beeping noise tells us to fasten the belt this feedback from the machine is supposed to make our travel saver, so we accept it. If combined with positive feedback we can still stand some negative feedback (compare The Erratic Appliances [1] or the EnergyLife App [8]). But these remain the exception. As most devices provide only positive feedback we will discuss if the persuasion is strong or passive if this aspect is evaluated [4].

2 Criteria for Feedback Devices

In order to find or develop devices which provide information regarding an individual's or a household's energy consumption, we need to define criteria/ requirements those devices should meet.

2.1 general Criteria

If a device is to be used in everyday life, it has to fulfill several criteria. Above all, the cost of the device and the cost of implementation have to be in a reasonable relation regarding the possible amount of saved energy. So reasonable cost The information displayed can have several levels of complexity. It may be an ambient information display providing general information or require the center of the attention because it displays details on a mobile phone. Therefor providing information about the consumption of energy is another requirement and also a criteria that can be evaluated [20].

2.2 criteria defined by Gustafsson

Gustafsson defined a list of criteria that describe in which case a device gets accepted as an addition to the household. A device in the household should

- fulfill aesthetic standards, meaning that it needs to make its use a pleasure starting with what it looks like. Why this aspect is so important is to be outlined later on.
- fit into the personal taste, so that people are likely to easily get accustomed to it and may even be able to customize it so that it becomes individual and "unique", this being quite an important factor in our society
- catalyze social interaction, implying that this leads to higher motivation based on a certain sense of competition and communication which also means creating a longer lasting learning effect and possibly questioning and, as a consequence, changing behaviorial patterns

In order to convince the user through positive feedback, the device should also

- provide entertainment which means that the user readily deals with it or even perceives energy consumption as some kind of game and therefore pleasure, not duty or constraint. At the same time, it should not be pure entertainment and leisure, though as this would mean that no real learning effect is achieved.
- promote self-reflection, so that the user questions his own habits and behavior instead of feeling forced into rules and renunciation (see also above)

3 Device Features

3.1 positive persuasion

For Anton Gustafsson, the central aspect in encouraging people to save energy is "positive persuasion" through so called persuasive technology, the definition of which being

"any interactive computing system designed to change people's attitudes or behaviors." [5] This means that technology is used in order to (this term here being without any negative connotation) manipulate its users, i.e. being a certain intent behind it, an intent or purpose by whom conceives, realizes and implements it. There are three categories to be distinguished: technology can be used to persuade people by acting 1) as a tool, 2) as a medium or 3) as a social actor:

1. Acting as a tool means that technology or more precisely: computers are are used for increasing people's capabilities and help them to perform certain tasks or actions, so that this ease or enabling convinces people of doing something they usually either would not be capable of or not willing to do because they consider it too difficult.
2. When acting as a medium, the user is persuaded, for example, by making him experience cause-and-effect relations in simulations. In this case, he can "try out" what would happen if...." virtually before opting to actually perform the action.

3 In this case, computers simulate human behavior and let the user explore social interaction virtually [5]. In all those functions, persuasion is, ideally, being achieved by positive feedback, i.e. a positive experience the user has and therefore decides to follow what he probably does not know being the original intent of who is behind the persuasive technology he uses for his own purposes. I do not aim at evaluating or judging this concept regarding ethical aspects (which it has been strongly criticized for), I merely use the categories presented above for giving solid ground to the concepts/devices presented in this paper. All of them belong to one or more than one of the categories described above as people first have to be made aware of the level and the consequences of their energy consumption and secondly need to be persuaded to modify it, maybe substantially. This understandably is much easily achieved through persuasion (and reward) then through coercion (and punishment). We argue that the first is to be preferred, i.e. positive behavior is to be rewarded, so that "negative" behavior (in our case high energy consumption) ideally is reduced or even avoided.

Part of positive persuasion (basically falling into the categories 2) and 3), that is, medium and social interaction) is playing games which have an underlying pedagogical purpose. There are four tenets which need to be taken into consideration when thinking up such a game:

1 Community: The concept of learning based on playing games in this context implies that the learning effect is, primarily, achieved due to interaction and learning by doing and getting feedback from or competing with other people, i.e. learning is mediated by peers in the game.

2 Autonomy: Those playing are in control of what they learn by playing and are in control of the learning process.

3 Locationality: The learning process happens at times and in places which allow the gamer to relate the learning experience to his environment and thus might trigger relevant action in "real" life.

4 Relationality: The gamers themselves make their learning happen in contexts which allow them to transfer knowledge and experiences gained in a virtual world into their own and thus profit from it, understanding also the implications of certain actions and thus facilitating the construction of meaning ([7] p. 164 also in the following). This concept of learning by using persuasive games is based on the assumption that the possibility of knowledge transfer cannot be taken for granted but that, in this kind of games or games in general, learning actually does occur, just that it is a consequence of social interaction, be it with a community, a real or simulated peer or a machine. Nevertheless, it is essential to create a virtual environment that allows cognitive relating to a player's individual situation and surroundings and obviously also to what he is supposed to learn about. Gustafsson refers to the Social Learning Theory and in particular situated action, but in this short paper, we cannot go into detail regarding this concept, so the ideas and notions outlined above should suffice ([7] p. 161 and in the following).

3.2 aesthetics and design

One other essential aspect Gustafsson deals with is design or what he calls "aesthetics":

"Design shapes our everyday material world, thereby inhibiting some actions and affording others. During the 20the century, design has been used successfully to increase our energy consumption through the creation of electrical appliances." ([7] p. 150 also in the following) He argues that design nowadays has not only to do with outward appearance and the purely aesthetic, i.e. "nice-to have" aspect of "our material world", but that its importance has (been) increased so considerably that designers have the power to change people's mindset regarding energy use and efficiency. Even though I do not fully agree with this point of view as I consider design one of many aspects but not the most important and at least in this area not the decisive one, I would like to present the devices thought up and realized by Gustafsson. They are not only well-conceived and based on an idea I support, namely the crucial importance of changing people's awareness and not just trying to "scratch at the surface" by, for example, propose them a new energy-saving device to use without making them understand the reasons behind.

4 PROTOTYPES

In this section we will describe the different areas of prototypes. Well known and well evaluated ones are described first, some from current literature and two I thought of.

4.1 known today available devices

4.1.1 electricity meter interfaces

Having defined and outlined the criteria I used to find suitable devices, I have a look at what is commonly available so far:

Simple electricity meters can monitor energy consumption of one specific task [20] or one household. There are devices to be plugged into an electric socket for monitoring a device. They are relatively cheap, but on the other hand raise low attention. Webservices provided by power suppliers for monitoring household consumption are a progress which could easily open up competition [3]. The best imaginable solution so far would be a household in which every single power socket is controlled through a homeserver application, monitoring energy consumption through a network. If it is possible to turn off devices in the household without moving around, it is comfortable enough, so that we are more likely to do it as we are provided with easily usable information and interaction.

4.1.2 energy calculators

Energy calculators are provided by a variety of hosts and usually suggest consuming new devices because they require less energy. Even though they obviously offer accurate information on energy consumed in a household, they nevertheless do not fulfill the requirements defined here. Most importantly, they might raise awareness only with those who are already interested in saving energy and probably also know how to interpret those figures. One more aspect why energy calculators are not sufficient in my view is that energy cost for creating the new device and energy cost of disposal are rarely taken into account, and therefore, these tools provide no reliable information.

4.2 Gustafsson prototypes

In general, Gustafsson proposes two kinds of prototypes meeting his requirements: design objects and games, both supposed to be pervasive, even though in different ways. They are to be presented and criticized in the following:

4.2.1 Power-Aware-Cord

The "Power-Aware-Cord" ([7] p. 82 and following) is a power cord glowing depending on the amount of electricity transported *(see figure 1)*, more precisely electroluminescent wire containing a semiconductor layer glowing when alternating current runs through it. This provides a real-time visualization of the relative energy consumption of one device or a group of devices connected by the power cord. This device, in Gustafssons words, lets people's actions

"immediately result in a response from the cord, giving the user direct feedback and the feeling of both seeing and interacting with electricity. This approach might inspire users of the Power-Aware-Cord to explore and reflect upon the energy consumption of other electrical devices in their home. Since this design has no added action or functionality, it does not become 'yet another gadget for their home', but a product people already buy." ([7] p. 150).

The idea as such and also its realization make sense and received overall positive feedback in testing as outlined by Gustafsson ([7] p. 151 f). However, the primarily design-oriented shape and function can be questioned as well as the fact that no accurate information

Fig. 1. Gustafsson's Power-Aware-Cord

is provided, the cord is just glowing, giving no scale or orientation. Moreover, ”inspiring” users actually is not sufficient, they need to be convinced of the necessity to do so and not merely temporarily play around with a device perceived as design element. The latter might also be questionable as one might argue that, if this cord has no other functionality, people living in the 21st century used to be surrounded by design and very sophisticated ”gadgets” like smartphones, they might see no need to buy this Power Aware Cord which, after all, might be perceived as adding no substantial value to what they already possess or use in their homes. In this context, it would rather be some kind of App on, for example, people's iPhones to be appealing to them and willingly be used as it fits into their habits and lifestyle.

4.2.2 The Element

”The Element” (Fig. 2) is an electric heater built with light bulbs. It consumes bright visible amounts of energy and reminds the user not to leave it switched on. It visualizes the room climate and the consequences of actions like opening the window in which case the temperature in the room would drop and the ”Element” heat up and at the same time emit more light to balance and make up for this by increasing temperature to the level set by adjusting the external sensor. The

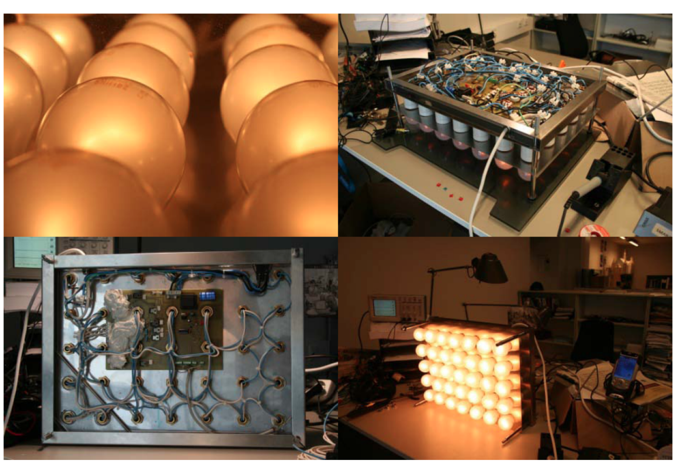


Fig. 2. Gustafsson's Element

argumentation behind this concept and prototype is that heating is a considerable expenditure in almost every household, but that people are not aware of this. If we categorize this device using foggs scheme, the ”Element” can be seen as a tool balancing room temperature, a medium allowing the user to play around with room climate and visualizing the effect this has and if not actually as a social actor, at least as a catalyst for social interaction, providing some of the charm of a campfire, although this might only be regarded as a pleasant side-effect. In Gustafsson's argumentation, though, the latter is crucial as he attributes high importance to object's design and the influence this might have on their users. Gustafsson rates this device as very valuable:

”The aim was to blur the product semantics of lamps and radiators to create an engaging object that discloses hidden properties of heat and energy. users found the ambient display to be an intriguing and interesting way to present energy consumption. To conclude, engaging objects such as our Element can influence users to reflect on energy and render such intangible phenomena more understandable.” ([7] p. 159 f).

Again, the idea behind this prototype fits the original intent to a high extent. Nevertheless, it can be stated that the concept is not so original any longer with wood oven imitations being rather common nowadays (even though they obviously lack the main functionality of the ”Element”, the visualization of room temperature, but they are, as design objects, far more appealing as they seem much more authentic, displaying a ”real” fire). Also, as with the ”Power-Aware-Cord”, the fact that luminosity increases along with the amount of energy being consumed, can be questioned as this might have an effect contrary to what actually should be reached: psychologically, it might make more sense, decreasing light when more energy is consumed and vice-versa as the ”reward”-effect would, in most people's perception, be more light, whereas deprivation of light and sitting in a darker room would rather have a negative connotation. Additionally, it might again be criticized that for visualizing a high consumption of energy, a lot of energy is employed which is a contradiction in adjecto apart from being counter-efficient.

4.2.3 Power Agent

Gustafsson, in his introduction to this persuasive learning game, writes that

”he hypothesis underlying our work is that persuasive games have the potential to strengthen situated learning and promote behavioral changes by reframing familiar activity and social systems.” ([7] p. 162 f).

As mentioned above when introducing Thomas' four essential aspects of such games, the main intent behind them is to provide the means for playful learning, that is, in the best case, learning without making a considerable effort or without even realizing it, the results of which can then be used in everyday life. In our case, the purpose clearly is to increase energy awareness and possibly also to show people how to modify energy consumption patterns in their homes.

The first persuasive learning game introduced is ”Power Agent”, a mobile game in which two households compete as agents. Each household has to fulfill energy saving quests. Real-world missions like cooking food or washing clothes need to be performed by the gamer, the game primarily being aimed at teenagers, and the focus is on saving electricity in the home, mostly in conjunction with everyday activities and tasks. Each day, the supervisor character in the game provides feedback about who is leading in the competition.

One positive aspect of ”Power Agent” is the possibly high learning effect achieved by letting gamers perform real action and giving them feedback on the consequences of those actions, so that they are able to generalize the knowledge gained. Plus, this game is supposed to quite accurately simulate the playing individual's real world, so that the criteria of locationality and relationality are cleraly fulfilled. The competition factor refers to the community aspect, and autonomy is granted by the possibility to start end stop or perform actions or not when the player wants to do so.

It is to be criticized, though, that this game lasts 24-48 hours and is, thus quite engaging. Apart from this, the game being aimed at teenagers, not all household members might be pleased by the impact the gaming activity of one or more underage member has on their everyday life and surroundings, not to mention possible risks switching on and off electrical devices or actively influence on household activities might entail. When the game is involves the whole household and is a ”one-off”, providing an entertaining learning experience, though, it is a valuable concept which can be realized for certain target groups.

4.2.4 Power Explorer

The second game, ”Power Explorer”, again belongs to the category of so called ”pervasive games”, i.e. games, which imply some kind of reality expansion, usually either Spatial Expansion (the game is not limited to a cellphone or computer screen but takes to the streets of the ”real world” and makes players act their), Temporal Expansion (it might mean, among others, that the game cannot be switched on or off, that it ”runs” in parallel to ”real life”) or Social Expansion (meaning that the game includes people who do not play the game themselves and may not even be aware of being part of a game) ([7] p. 203 f). ”Power Explorer” is a

”pervasive action-oriented multiplayer game where the overall goal is to explore the household, learn about its electricity consuming devices and develop a positive attitude towards conserving electricity.” ([7] p. 206 f).

The player has to feed a blob by reducing energy consumption at home. There is a duel mode in which the players have to consume energy in a specific way in order to control their avatar. The user group for tests were teenagers who were supposed to develop awareness regarding their use of energy which resulted in relatively high energy savings in the participating households. This shows that the game achieved its original goal. In intent, idea, target group and realization it is quite similar to ”Power Agent”, with one difference being that it tries to transform ordinary environments into ”playgrounds”, thus altering perception and ideally raising awareness of devices, facts and behavioral patterns which usually are either not considered or not even consciously perceived. Moreover, it renders tasks and responsibilities often seen as tedious and tiring more interesting, giving them, even if only temporarily, a playful character which is enhanced by the social factor of interaction with other gamers and the aspect of competition and somehow combining the necessary and useful with pleasure and entertainment. What can be improved or added, though, is background information on what is done why and how in order to foster a deeper understanding and learning that goes beyond the boundaries of the game. The experience made might be discussed, reflection and joint understanding of events strengthening the positive learning effect and making experience and knowledge achieved less purely intuitive and more based on facts, reason and understanding. This would lead to a long term change in behavior and ideally also allow this knowledge to be passed on to others.

4.3 additional prototypes

This section briefly presents two more devices supposed to raise energy awareness and question consumption behavior. As they can be compared to devices discussed above, we just mention them, adding some minor aspects and underlining the statement that so far, no really convincing device has been realized and successfully promoted.

4.3.1 The Flower Lamp

[1] p 17 The ”Flower Lamp” is a lamp in the shape of a flower which opens when there is little energy consumption and closes at high energy consumption. The feedback provided draws attention through its passive visualization, and the lamp has to be used before it operates. I mention this device as one more example of a design product which has the additional functionality of visualizing energy consumption, but I will not discuss it as it is very similar in intent, advantages and disadvantages and impact as the ”Power-Aware-Cord” presented above, with the main difference being its more aesthetic aspect which means that it might primarily attract attention and interest because it ”looks good”, ”reacts” and might seem an original idea or even fit some flat’s interior design. However, the ”Flower Lamp” is very likely to not considerably raise energy awareness and even less to make people change their habits or think about it for more than a short period of time when the device is still seen as new and enjoyable before it falls into oblivion.

4.3.2 The Energy Curtain

[1] p 7 The ”Energy Curtain” collects energy when light falls on it. It reverses the learned usage of curtains. We are used to closing a curtain in order to keep light out and open it in order to let light in. In this case, though, the curtain needs to be closed in order to collect the light. This, again, makes us think of one device already discussed: the ”Element”. Quite similarly, light plays the predominant role in visualizing electricity consumption, and there is a reversed notion of the common function the ”original” household device has. The main aspect of criticism here is that in my view, it is not likely to be very appealing to people unless they want some more gadget in their households, but even then it might become uninteresting and boring after a short while. It does not really provide valuable information and has no considerable learning effect.

4.4 future devices

In addition to the devices which either already exist or have been conceived, I would like to to propose two new ones which have the advantages and features needed according to Gustafsson in order to enhance energy saving awareness and encourage people to adapt a more sustainable lifestyle, but at the same time make up for the disadvantages I will outline in detail in the ”critique”-section.

4.4.1 coincounter

The so called ”Coincounter” visualizes energy consumption through falling coins on a conveyer belt in a defined fixed time interval (once an hour). Depending on the amount of energy consumed the coin used to display the amount changes so we can use a well learned technique (counting pocket money) to estimate the amount of consumed energy. Everyone understands money [19], for example while going to work you check the coincounter and there are only cents on the belt except for the time between 2 and 3 am. This is when you realize that your midnight snack might have consumed more then the usual cycle of the fridge, that is cooling repeatedly during the whole day [20].

Loosing money is a symbol understandable for everyone and has a strong pedagogical effect. This means, that the first impact when consuming a lot of energy very probably is negative feedback as the loss of money visualized by the ”Coincounter”, the coins symbolizing energy consumption passing by and vanishing ”hurts”, whereas, over time, positive feedback is provided when, by learning to use less energy, the amount and value of the coins ”spent” or ”wasted” is reduced. Thus, saving energy becomes equal to saving money which makes the idea of energy saving not only more interesting but might even give it a positive image. In addition, similarly to Gustafsson’s games, this device might lead to some sort of competition within a household with those being ahead who manage to save money by considerably reducing their use of energy. Moreover, in this case, related social games can be added, and self-reflection and social interaction are encouraged which makes the learning effect even higher. This idea can be further developed and improved, for example by providing statistics, overviews and graphs regarding energy consumption in a household over time, broken down on single devices, household members/users, potential for improvement and so forth. What makes this device more suitable and meets better the previously defined requirements are ease-of use, visualizing energy consumption in a commonly understandable way, positive feedback for using less energy, its playful and yet serious character and its potential to raise awareness plus possibly pointing at potential for improvement.

4.4.2 endless bar - an infinite recycle cup

The second device offering the features we are trying to combine would be the so called ”endless bar”. In order to demonstrate the overall energy consumption of a product like a plastic cup, use a 3d-printer to print the cup and a ’cut and melt device’ to recycle it. The infinite recycle cup selects one throw-away product to monitor the energy consumption of the whole product cycle it runs through and visualizes with high accuracy the required energy especially of its disposal. The basic idea behind this concept is that, by visualizing the energy consumed by/for a simple and everyday product (production and disposal), people are induced to think about a usually unconscious action, i.e. the waste of energy in everyday life. One could argue that most throw away products today are not produced by a 3d printer, fabrication in large numbers is much more efficient and therefor this is not a valid comparison. But in the production of termoplast products the most significant amount of energy used goes to heating up the plastic [15]. So we can argue that if the volume of the 3d printed cup equals the volume of the fabricated cup the amount of energy used for heating the plastic is the same. So we can measure the overall energy consumption of the product cycle if we accept the approximation that "performant manufacturing + long distance transportation" equals "3d printing + short distance transportation". The overall energy consumption can, in a second step, be put in relation to a commonly known and easily imaginable value like for example by stating that the energy used for 10 "coffe2go"-cups equals the energy required for making a pizza, heating the water in a bath tub or watching a movie on TV. Those examples are easily understandable and not only do raise or enhance awareness, but will very probably also encourage energy saving as one might, for example, think of using one's own cup instead of buying a throw-away cup 10 times, and then "rewarding" oneself by watching a movie. Additionally to the learning effect caused, this concept, if realized in the way mentioned, might also be somewhat entertaining and raise curiosity - this being a positive side effect.

Referring to the categories and criteria mentioned above, it can be stated that the two newly presented devices can neither be defined as "games" nor as mere "design products". Their main focus lies on learning by raising awareness in a somewhat pedagogical way. There is no interaction required nor is it possible, but they strongly suggest to proactively change energy consumption behavior after realizing its impact. They offer information, are "fun" in the way they are presented, are designed for dealing with them not only once or growing easily tired of their use, but for being an "energy saving companion" in a household, a school or educational institution, a public space or even an office. This also means that the criticized limitation to only some target groups is overcome here as those two devices are more generally and globally applicable. They also fulfill the criteria mentioned for persuasive learning games, i.e. locationality (being directly related to the environment they are used in and its electrical appliances), autonomy (each user being able to learn from it as he pleases and make use of it when he desires to do so), community (as referring to a whole household, comprising all members and devices and triggering social interaction in saving energy) and last but not least, relationally (the user being immediately able to understand, relate and possible take action regarding energy consumption in his household. I therefore state that even though the two devices proposed might not be perfect either, they offer some considerable advantages the already existing devices do not and at the same time lack their most important disadvantages. But in the following section, evaluation is to be made and visualized so that the overall picture becomes clearer.

5 Evaluation

Evaluation of learning effects that appear rather longterm is costly and difficult.

5.1 Methods of Evaluation

Only few devices have been evaluated. The criteria used for evaluation are enumerated and briefly explained in the following:

5.1.1 Measurable amount of energy saved

Some devices like the mobile games or meter interfaces [8] can be evaluated by the amount of energy saved.

5.1.2 Promoted self reflection

Some devices like the Power aware cord have shown a promoted self reflection. They have shown in user studies that users are intrigued by it and understand it intuitively.

5.1.3 Promoted social interaction

Devices like The Element can only act as social catalyst, which can be a strong leverage to promote a behavior.

5.2 Difficulties of Evaluation

For several of the criteria there is no standard scale. Which makes attempts of evaluation either expensive because they would require long term studies with a significant amount of participants or just impossible because they would simply take to long.

5.2.1 longterm learning

How can we evaluate the impact of glowing flowers on the wall on learning about energy? Their design will only work on people who are already aware of energy usage in their households. Therefore it is questionable whether the devices presented (like the Energy Curtain for example) do not serve just as pleasurable distractions which might, at the most, point at the issue of energy saving in the household, but have no long term effect or maybe not even any learning effect at all (giving no additional information, there main feature being aestheticism).

5.2.2 aesthetics

Aesthetic decisions vary not only through gender and age groups they may also differ strongly within one household. In general, it may also be stated that "design" in a family with younger children might play a far less important role than in a single household with high income. In the latter, decision for buying a design product very probably will be taken in favor of a mere design-oriented item, without any additional functionality or effect and possibly of high quality, i.e. the "market" for devices like the "Power-Aware-Cord" or the "Element" might be narrow, although this obviously also depends on other factors like marketing which are not to be discussed in this paper.

5.3 cultural and personal differentiation

The aspect of "personal involvement" is important for any kind of action to be taken or learning effect as it considerably enhances or even only just triggers changes in behavior: if you feel personally involved, the issue becomes also "yours" and you feel the need to act much stronger than if it where "only" about others. A good example regarding energy is the fact that we do not experience power cuts anymore in Europe. So we do not really feel the necessity that each one of us has to participate in the effort of saving energy which is called rebound Effect [13]. Even if, occasionally, a blackout occurs in a single household or even more rarely, in a whole city, this usually has no long term effect on our energy consumption habits. Moreover, our idea of "energy" in most cases is very abstract and vague, so that we do not really feel the need to deal with it as awareness only comes with and is directly related to concrete ideas, i.e. we need to perceive with one or more senses in order to become aware. One more aspect to be taken into consideration is the high living standards we are used to in Western societies. Warm showers, heated car seats and 24/7 television broadcasting form an integral part of our lives which we could not do without, actually, we cannot even imagine life without all the commodities we are accustomed to. This makes raising energy awareness even more difficult as in order to save energy, we might have to refrain from some of the aspects of our everyday life luxuries we take for granted. In countries still experiencing power cuts each member of the family knows where in the house the candles are to be found. In developing countries where electricity is not a available in abundance or maybe for some people not at all, everyone is aware that energy has a price and therefore can be worth saving which makes the aspect of "everyone understands money" even more important [19]. In general, it can be argued that a prepaid electricity contract (that lasts one month for the average consumer) will have a bigger impact on energy awareness [18] (regarding overall society) then a game that only interested people will play.

5.4 abstract feedback

Devices like the power aware cord are likely to have a counterproductive effect on some people. The more energy it consumes the brighter it glows which might encourage to enhance energy consumption in order to make it glow. So it will require long periods of evaluation to show which devices are the best let users learn about their consumption [2].

6 CRITIQUE

Although the devices proposed by Gustafsson and others definitely do have features which serve the main purpose, i.e. encouraging energy saving, they also have some disadvantages which, in my view, either outweigh the positive aspects or at least make them much less efficient than they are supposed to be. The "Power Aware Cord", for instance, might have the counter-effect of making people want to use more energy then less as it seems playful and fun making it glow ever more brightly by enhancing energy consumption. As Gustafsson says "At this stage, the Power-Aware Cord is meant to be a conceptual design statement, mostly used to test peoples reactions and provoke thoughts around the area of energy consumption". Moreover, the concept of "The Element" as such is rather questionable as an electrical heating cycles twice the conversion efficiency of heat to electricity and again form electricity to heat which means that it is rather inefficient as such and thus contradicts the original intention of designing sustainably.

Games like "Power agent" at first sight seem to be a good and original idea and might work as they distract and make aware at the same time, but nevertheless, it has to be criticized that games are not everybody's cup of tea. For less game-prone people and in general the elder generation or people with less time at their disposal, "Power agent" might just seem a waste of time and the playful character outweighing the serious purpose and intention. Even a comparably simple application like the EnergyLife proved to very hard to handle in some use cases [8].

One more aspect regarding energy measuring devices or flowmeters in the household is the social interaction and dynamics this might entail: a flowmeter, set up in a network in the household which visualizes energy consumption in detail, meaning that consumption can be tracked down to the device used and to the person using it, can lead to enhancing awareness. But this might also imply that, instead of having the effect of opening people's eyes and making them want to save energy, blame is attributed or accusations are being made against those who consume more or too much energy. This might seem a negative impact, but on the other hand can also lead to positive feedback for those who consume less energy or, after a learning effect caused by enhanced awareness, make an effort to do so.

Moreover, in my opinion, one aspect which also needs to be thought of is whom we design/ conceive a device for. "Household" in central Europe very often means single households, but can also imply a family, i.e. more than one generation, gender and age group. This means that we have to consider different needs, perceptions shares of and participation in energy consumption etc. To put it differently: the target group has a substantial impact on what a device looks like, on its use and purpose. The use and perception of Gustafsson's "Energy Aware Cord", for example, might differ strongly between a parent and a child - the latter might want to play with it and, as mentioned above, regard high luminosity as attractive and joyful whereas the parent understands the idea behind and might therefore try to keep luminosity low. This goes for all the devices presented here, though and is one crucial aspect not to be neglected, even though it is quite obviously hard if not impossible to design a device meeting, on the one hand, all our requirements and, on the other, appealing to and having the desired effect on all possible target groups (and the gamut is very vast as energy saving should, ideally, concern each of us).

Last but not least, a general aspect of criticism which Gustafsson, in my view, neglects, is the fact that the energy consumed by producing and, in the end, disposing of a device, is not sufficiently taken into consideration or not considered at all.

This is why I outlined two ideas which embrace all the essential aspects of enhancing energy saving awareness identified in this paper, the most important one being providing positive feedback and thus encouraging changing energy consumption habits on a long term basis.

7 CONCLUSION

While providing information about energy consumption for people who are already interested in the topic is a challenging but solvable task for interface designers, those who are not interested in saving energy will be hard to convince of the necessity and ease of doing so. Providing appeal to save energy will require more effort than providing just a variety of well designed devices. The raising cost of energy is a strong argument when the cost of recharging a laptop equals the cost of being online. Even if this might seem pure common sense, it still needs to be mentioned as it holds true especially in this case: As soon as people become aware how much money they are wasting by wasting energy, they might want to deal with energy saving if this helps them to save money, possibly without considerable changing their habits, lifestyle or depriving them or comfort they are used to. This can be rated as both "negative" and "positive feedback", the negative part being the deterrent effect of being confronted with loosing money and the positive side being the "reward" dangled by making people aware of money to be saved as well as precious energy. This combined might lead to a more sustainable lifestyle.

REFERENCES

[1] S. Backlund, M. Gyllenswrd, A. Gustafsson, S. Ilstedt Hjelm, R. Maz, and J. Redstrm. Static! the aesthetics of energy in everyday things. Technical report, SICS publications database [http://eprints.sics.se/perl/oai2] (Sweden), 2007.

[2] S. Darby. The effectiveness of feedback on energy consumption. *A Review for DEFRA of the Literature on Metering, Billing and direct Displays, April*, 2006.

[3] S. Darby. Smart metering: what potential for householder engagement? *Building Research Information*, 38(5):442–457, 2010.

[4] C. DiSalvo, P. Sengers, and H. Brynjarsdóttir. Mapping the landscape of sustainable hci. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 1975–1984, New York, NY, USA, 2010. ACM.

[5] B. Fogg. *Persuasive technology: using computers to change what we think and do*. Morgan Kaufmann series in interactive technologies. Morgan Kaufmann Publishers, 2003.

[6] J. V. H. Gary Cook. How dirty is your data?, 2011.

[7] A. Gustafsson. Designing enjoyable energy feedback experiences in the home. In *Positive Persuasion Designing enjoyable energy feedback experiences in the home.*, pages 3–245, Göteborg, SWEDEN, 2010. ACM Press.

[8] G. Jacucci, A. Spagnolli, L. Gamberini, A. Chalambalakis, C. Björkskog, M. Bertoncini, C. Torstensson, and P. Monti. Designing effective feedback of electricity consumption for mobile user interfaces. *PsychNology Journal*, 7(3):265–289, 2009.

[9] E. P. James Pierce. Redesigning everyday interactions, relationships and practices with energy. BECC, 2010.

[10] K. Kappel and T. Grechenig. "show-me": water consumption at a glance to promote water conservation in the shower. In *Proceedings of the 4th International Conference on Persuasive Technology*, Persuasive '09, pages 26:1–26:6, New York, NY, USA, 2009. ACM.

[11] J. Moan and Z. Smith. *Energy use worldwide: a reference handbook*. Contemporary world issues. ABC-CLIO, 2007.

[12] J. Pierce and E. Paulos. Designing for emotional attachment to energy, 2010.

[13] J. Roy. The rebound effect: some empirical evidence from india. *Energy Policy*, 28(6-7):433–438, June 2000.

[14] A. Stevenson and L. Brown. *Shorter Oxford English dictionary on historical principles*. Shorter Oxford English Dictionary. Oxford University Press, 2007.

[15] K. Stoeckhert, W. Woebcken, and H. Gupta. *Kunststoff-Lexikon*. Hanser, 1998.

[16] E. Union. Directive 2010/30/eu of the european parliament and of the council of 19 may 2010 on the indication by labelling and standard product information of the consumption of energy and other resources by energy-related products (text with eea relevance). 2010.

[17] J. H. van Houwelingen and W. F. van Raaij. The effect of goal-setting and daily electronic feedback on in-home energy use. *Journal of Consumer Research: An Interdisciplinary Quarterly*, 16(1):98–105, June 1989.

[18] H. Wilhite and L. Lutzenhiser. Social loading and sustainable consumption. *Advances in Consumer Research*, 26:281–287, 1999.

[19] V. W. Will Anderson. Exploring consumer preferences for home energy display functionality. 2009.

[20] G. Wood and M. Newborough. Dynamic energy-consumption indicators for domestic appliances: environment, behaviour and design. *Energy and Buildings*, 35(8):821–841, 2003.

Chances for computer-aided energy-awareness in private homes

Benedikt Zierer

Abstract— Conserving natural resources and using electric power in an energy-aware way are challenging tasks of this century; recent events like nuclear catastrophes in Japan and global warming, resulting in melting of the poles and rise of water levels, show that a more responsible way of dealing with these issues is necessary.

We will see that computers, despite of consuming power themselves, can aid humans in conserving this precious good by either providing feedback by visualizing electric current in its invisible and intangible nature, or supplying instructions for optimized and efficient use.

This article introduces informational facts about why it is necessary to think twice about wasting energy and what consequences arise from generation of electricity, what effects can be achieved by providing highly resolved usage data or giving other incentives to conserve energy, and introduces some interesting projects aiming at the visualization of electricity with the goal of increasing awareness and thus conserving energy.

Index Terms—Energy, Power, Awareness, Consumption, Carbon dioxide, Savings, HCI, Interface

1 Introduction

Since the earthquake of March 2011 in Japan and the resulting damage to the nuclear power plants Fukushima I, Fukushima II, Onagawa and Tokai, many people question nuclear power as a safe source of energy.

Combined with global warming, which is a result of burning fossil fuels and thus releasing carbon dioxide, the power sources which contribute by far the largest part to energy production bear risks and consequences, which to minimize is seen as one of the big challenges of the third millennium.

Recently, the German administration decided to shut down all commercial nuclear power plants until 2022; combined with the self-set goal of reducing CO_2 emissions to 40 percent less than 1990 by 2020 as part of the post Kyoto Protocol process and in accordance with European Union guidelines [5], achieving those goals is not only dependant on expansion of renewable sources of energy, but also on consuming less electrical energy.

A first step in trying to minimize energy consumption is increasing energy awareness by providing data to people about how much energy which device or action consumes: Being well informed about usage significantly reduces it, as [17], [24] and [13] show.

This article focuses on energy consumption in private homes. This is because at home individual persons have the most control, when which device is turned on.

As it is proven, that being informed about consumption increases awareness and thus reduces usage, it seems odd that today nearly no energy consuming device itself saves and displays statistics about power consumption.

For some devices like lighting bulbs or refrigerators, it is mandatory to declare average consumption, so that customers can include this attribute into their purchase decision, but upon purchasing the device, the customer is left alone in the dark, unless he purchases additional devices to show the actual consumption.

The next section analyses statistical data about energy usage in German homes to show the dimensions of potential energy savings.

How much influence detailed information about and incentive to conserving energy can have, is exemplified by describing selected experiments and their results in the section after that; this is followed by a presentation of some very promising projects from human computer interaction and ubiquitous computing research. Those projects aim at increasing awareness, where energy is consumed by either providing feedback data or visualizing the consumption of electric power; only when an individual is aware, where energy is consumed, it can think about if this appliance is necessary or if energy can be saved.

- *Benedikt Zierer is studying Media Informatics at the University of Munich, Germany, E-mail: b.zierer@campus.lmu.de*

- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

2 Energy consumption of households in Germany

In 2008, 27.4% of the total energy consumed in Germany was consumed in private households [4, p. 10]. Because of the rising number of households due declining number of persons per household, it may be assumed this proportion has risen since then; the amount of 40 million households [9] will be used throughout this article.

This chapter will now examine, which devices are present in German households, how much they contribute to energy consumption and energy from which sources is used in what proportion, including the consequences of energy production.

2.1 Areas of consumption

A big survey by the German GfK shows, which power consuming devices are present in German homes [26, p. 69], shown in table 1.

Table 1. Devices in German households 2008 [26, p. 69]

Which device is present in your home?	
TV	97%
Washing machine	94%
Electric kitchen stove	89%
Refrigerator	78%
Microwave oven	66%
Freezer	63%
Dish Washer	60%
Personal Computer	60%
Tumble dryer	40%
Combinated Fridge and Freezer	35%
Small electric heater	17%
Aquarium	6%
Sauna	4%
Water bed	2%
Air conditioner	1%

It also states, that there is a direct, almost linear dependency between how many electric devices are present in a household and how much power is consumed during a year. This ranges from about 1000kWh for household with five devices to 6000kWh with about twenty devices; this result look very reasonable and easy to explain: When owning more devices and also using them, it is consequential that more power is consumed.

There also is a similar dependency between the monthly income and power consumption, ranging from 2000kWh when earning 1500 Euro or less a month to 4000kWh when earning more than 3000 Euro; this can be explained with more money leading to more electric devices as well as bigger homes.

The average energy consumption of a German household is 3340kWh/year, when using statistic regression, each additional lamp adds 2kWh and each additional electric device adds 229kWh.

2.2 Impacts on the environment

To further demonstrate the importance of being energy aware and conserving power, the consequences of electricity generation in Germany are shown. When using data from [4, p. 28], the electricity generation consists of the sources shown in table 2.

Table 2. Electricity generation by source in terawatt hours 2008 [4, p. 28]

Source	tWh	Percent
Total	637.6	100
Brown coal	150.8	23.7
Nuclear power	148.8	23.3
Anthracite	124.6	19.5
Domestic gas	86.0	13.5
Wind power	40.5	6.4
Hydroelectricity	26.7	4.2
Mineral oil products	9.9	1.6
Other power sources	50.3	7.9

Of those sources, only wind, water and sun produce little to none waste during generation. According to [11], the German energy mix results in 563g of carbon dioxide per kWh of generated electricity for 2010. This number will now be used throughout the article to demonstrate consequences of conserving energy.

All energy sources that are based on burning fossil fuels produce large amounts of carbon dioxide and other greenhouse gases, which contribute to global warming. Nuclear power only creates a very modest amount of such gases, on the other hand, there are nuclear substances left over after the fuel rods cannot be used to generate power anymore, some of which with radioactive half lives of some hundred years. For German nuclear power plants, the amount of radioactive waste produced is stated as 0.0027 g/kWh by [34, p. 28], or 0.0007 g/kWh [33] for the energy mix, resulting in 2.3 grams a year per average household.

Another topic not yet mentioned are the costs of electricity. Assuming a price of 0.25 Euro per kilowatt hour using data from German BDEW [7], the average consumption of 3340kWh results in 835 Euro a year for domestic electricity. This does not include hot water and heating, unless those use electricity. This number will also be used for demonstration.

Deutsch [8] conducted an experiment, testing whether consumers make the decision for buying a new washing machine dependent on low life cycle costs, including energy efficiency. He came to the conclusion that, when directly confronted with life cycle costs, the test group would buy devices with only "0.83% less specific energy consumption" than the control group.

Thus it seems that the display of energy consumption and consequences alone are not enough to motivate users to make energy aware decisions in device selection and usage, but the display must be tailored to suit the needs and interests of the users.

3 Results of Experiments

In this chapter, different experiments, which try to influence energy consumption by providing information about it, are shown. Sadly, most of those experiments do not use state of the art interface solutions as presented in the following chapter; however, the following experiments put great effort in measuring the energy savings over longer periods of time and thus provide valuable data, which dimension of savings can be achieved in real world deployments.

3.1 Real-time visual feedback and incentives for dormitory residents

Petersen et al. [24] designed both an automated "high resolution" real time feedback display and a "low resolution" meter, whose data was only supplied once a week with the purpose of measuring differences in energy and water consumption between groups supplied with one of the devices; the study was conducted in college dormitories.

For the high resolution feedback, two dormitory buildings were equipped with automated meters for energy and water, whose data was transmitted wirelessly to a web server each 20 seconds and saved in a database. The data was graphically presented on a website using Flash animations, those websites were shown on large displays in the lobbies of the dormitories and each student could access the data on their own computers.

In 20 other dormitory buildings, data from utility meters was presented on a weekly base for low resolution feedback. The experiment lasted from February 1st 2005 to April 20th 2005.

The initial measurement was divided in three distinct periods: First, a three week baseline measurement, second a two week competition period, in which students were given incentive to win by being the building with the lowest consumption, and last a two week post-competition period, which should show if energy awareness would prevail without incentive.

Since the buildings are not identical in number ob habitants, devices installed and other factors, the collected data was not compared in total, but in percent changes to the baseline. Because of measurement issues, only data from 16 low resolution dormitories and both high resolution dormitories was evaluated for energy savings. The average consumption decreased by 32 percent from 367 to 250Watt per capita between baseline and competition period. The average reduction of energy use was 55 percent for the two high resolution buildings and 31 percent for the low resolution ones.

Another interesting finding was that decrease in consumption averaged 46 percent in the two low resolution buildings inhabited by freshmen only, whereas the two buildings which only house upperclassman resulted in an average decrease of 2 percent.

The consumption continued to decrease in the post-competition period to 241 Watt per capita, but outside temperature and brightness significantly increased during this period by about 73 percent, it can not be said which part of the decrease is due to energy awareness and which to less need for lighting and heating. On the other hand, the ice cream party, which was the prize for the winning dormitory, was only attended by about 10 percent of the eligible students. Thus, either the competition itself or the pure display of consumption data seems to have been a stronger incentive than the prize.

The decrease in water consumption averaged to only 3 percent. When asked, what measures were taken to conserve energy and water, most students stated in descending order "turning bathroom lights off when unoccupied (71 percent), keeping lights off when dormitory rooms were unoccupied (70 percent), using natural lighting during the day (59 percent), shutting off computer monitors while not in use (50 percent), using less hot water in showers and clothes washing (45 percent), turning off hall lights (42 percent), and turning computers off when not in use (39 percent)" [24, p. 14]; this may lead to the conclusion, that electric power is wasted on far more situations than water and thus explain why much more energy can be conserved that easily without lifestyle sacrifices.

If this experiment had taken place in Germany of 2010, the total energy savings of 68,300kWh in the competition period compared to the baseline would result in 38,453kg CO_2 and 58g nuclear waste less produced as well as savings of 17,075 Euro.

3.2 Dynamic energy-consumption indicators for domestic appliances

Wood and Newborough [32] state that not only average power consumption of a household is of importance, but also minimizing peak demand in order to increase efficiency of power production.

Further, they research how interfaces between people and their home appliances may create energy awareness and lower consumption with a special focus on cooking, since cooking is responsible for most of the power spikes in private homes.

There are two general effects on related studies mentioned: First, there is the Fallback effect defined by Wilhite and Ling [31, p. 3] as "the phenomenon in which newness of a change causes people to react, but then that reaction diminishes as the newness wears off", second the Hawthorne effect described by Gortz and Döring [2, p. 472] that is more or less present in every user study, meaning that subjects behave differently because they know that their actions are watched.

Wood and Newborough refer to previous studies, in which houses were equipped with visible feedback on their energy consumption by either installing the electric meter in a more prominent place or using computer programs by McClelland and Cook [20], Dobson and Griffin [10] and Brandon and Lewis [3]. The energy savings compared to control groups or previous data ranged from 12 to 15 percent.

A field study was designed for four groups, measuring how information about power consumption during cooking influences it. The four groups were first the control group, second a group which was given seventeen printed pages of information material, third a group which was supplied with energy-consumption indicators for electric cookers and fourth a group which was equipped with the information material as well as the energy-consumption indicators, totaling 44 households, 12 for the first and second and 10 for the third and fourth group.

The whole monitoring lasted 18 months, for groups two to four baseline data was acquired for 2 months prior to 2 months of monitoring them with supplied materials, the control group was monitored for a whole year.

The energy-consumption indicator (ECI) consisted of a current transformer between the power outlet and the cooker, which resistance was measured beforehand, connected to a laptop computer which stored the data for evaluation as well as supplied it to the visible display shown in figure 1. The display shows kilowatt hours of consumed energy for the current tasks, today, yesterday, this week and last week.

Surprisingly, some of the households supplied with either the information pack, the ECI or both increased their consumption compared to the baseline period.

For group two (information only), 66 percent of the households achieved savings after all, averaging 6.4 percent for those or 3 percent for whole group two; one household achieved savings of 13 percent.

In group three (ECI only), 80 percent of the households achieved savings at all, averaging 20 percent, or 15.2 percent for the whole group; the highest savings by a single household were 39 percent.

Group four (information material plus ECI) contained 75 percent households saving energy at all, averaging 14 percent or 8.9 percent total with the highest savings of one household being 27 percent.

The authors conclude, that the ECI display is a promising approach and should be further optimized in terms of usability, interaction and including the energy saving tips from the paper information pack.

When using the number of 9 percent of domestic energy consumption being used for cooking from Nipkow and Gasser [21], a decrease of 15 percent in energy used for cooking would amount to savings of 45 kilowatt hours per year and household in Germany, resulting in 25kg CO_2 and 0.032g of radioactive waste less produces as well as 11 Euro saved. This might not seem overwhelming, but when assuming about 40 millions of households in Germany, this is quite something - about one third of a percent of Germany's total energy consumption.

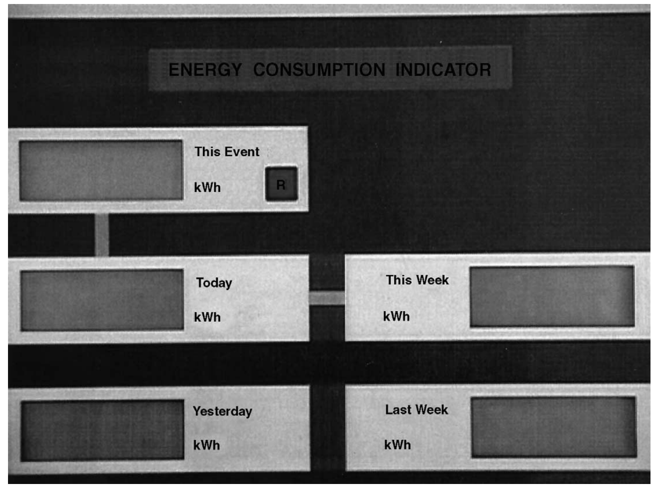


Fig. 1. The energy-consumption indicator [32]

3.3 Energy conservation through product-integrated feedback

McCalley and Midden conducted a similar experiment [19], giving feedback about washing machine energy consumption.

As they dedicate their work to economic psychology, they focus on the effects of feedback and goal setting, not on designing the feedback interface. The experiment was conducted amongst one hundred persons, each performing ten washing trials on a simulator as baseline.

After the global baseline measurement, four groups were formed: The first as control with neither feedback nor goal, the second with feedback only, the third with feedback and a self-set goal and the fourth with feedback and an assigned goal.

The simulator referred to a Miele Novotronic Super washing machine each participant was familiarized with first. For groups two to four, the simulator also included a feedback display, showing how much energy their chosen washing options would consume, group one was informed about their consumption during the ten baseline measurements.

Before actual measurements, group one and two were displayed a message, at least 20 percent of energy consumed by washing machines could be saved by using lower temperatures. Group four was assigned with the goal of saving 20 percent, group three could chose if their goal should be savings of zero, 5, 10, 15 or 20 percent; the average self set goal was 15 percent, with the majority choosing 20 percent savings.

The ten baseline measurements averaged 0.90, 0.91, 0.92, and 0.86 kWh for groups one to four, the twenty actual measurements averaged 0.80, 0.81 kWh, 0.73 and 0.68 kWh for the respective groups. As there was virtually no difference between groups one and two, the data from group two was removed from further evaluation. The results show that goal-setting has a positive effect on savings, reducing the consumption by 21.9 percent for self-set and 19.5 percent for an assigned goal, and, interestingly, that the value of the self-set goal had virtually no influence on the outcome.

Thus, the authors recommend enabling the microprocessors and displays already present in modern washing machines for setting savings goals.

Assuming 7 percent of domestic energy consumption being used for washing machines [21], reducing this by 22 percent would amount to savings of 51 kilowatt hours per year and household in Germany, resulting in 29kg CO_2 and 0.036g of radioactive waste less produces as well as nearly 13 Euro saved. This is a similar number as the possible savings for cooking mentioned before.

3.4 The Benefits of Information on the Efficient Usage of Consumer Durables

Matsukawa [17] analyses an experiment from 1996, where displays were installed in 194 private homes, displaying tips about how to use air conditioners, refrigerators, TV sets, washing machines, clothes driers, and microwave ovens more efficiently as shown in table 3 [17, p. 25]; the energy consumption of those households, excluding the display itself, was recorded for three months.

Table 3. Examples of Information on the Efficient Usage of Electrical Appliances [17]

Appliance	Suggestion
AC Filters	Clean air conditioner filters at least once in two weeks.
AC Timers	[Use AC timers] only when heating or cooling is necessary.
Fridge Food storage	Do not store too much food in a refrigerator.
Fridge cleaning	Keep the door seals of a refrigerator clean.
TV Brightness	Do not make the screen too bright.
TV Operation	Turn off a TV set when you are not watching it.
TV Standby power	Unplug a TV set to save standby power.

Evaluation of the data shows that households using the display at all save 0.141 kilowatt hours in average each day compared to the ones not using it. Multiplying this number with 365.25 days per year (factoring in leap years), one household saves about 51.5 kilowatt hours when being informed about efficient use; this is result is very similar to the ones of the experiments mentioned earlier, thus savings would also be about 30kg CO_2 less produced and about 13 Euro in Germany.

The author concludes that savings can be achieved by supplying people with basic information, how common devices can be used efficiently and mentions the internet as cheap source for such information.

4 Projects

In this chapter, different approaches for interfaces, which inform about amount or quality of energy consumption, are presented. There is a great diversity of design approaches, from detailed information on a portable device to subtle feedback through varying light quality.

All those projects have in common, that they are relatively new and thus the available evalution is limited to small user studies for improving the design, but no long term studies have been conducted.

4.1 WattBot

Dane Petersen et al. [23] introduced the WattBot at CHI 2009. In accordance with [6], the main design goal was to provide real time consumption feedback to induce behavioral changes. Working from the premises, that people would conserve energy on their own, when only being presented with accurate data, they decided to design a program for iPhone and iPod called "WattBot".

Key argument for choosing Apple's mobile products was to have the information available everywhere in the house without needing an additional device. The display of energy consumption is broken down into the different rooms as seen in figure 4.1, the data is provided by sensors on the circuit breaker box, each transmitting its data wireless. The advantages of measuring the energy consumptions directly on the fuses are that devices with high power demand such as stoves or air conditioners often have a fuse of their own, that fewer sensors are needed than putting one on each power outlet and that the data is more accurate, since leakage in the wiring is accounted for, too. The drawback of using this central array of sensors is, no device that is not fused separately can be measured individually; this would significantly aid in detecting devices with high standby power consumption.

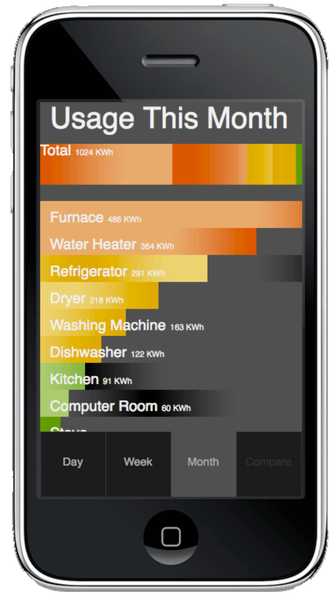


Fig. 2. Concept rendering of WattBot application [23]

The authors conducted user studies with A/V recordings and given tasks to execute, but only evaluated the interface in terms of usability and suggested features, not energy awareness and energy savings.

4.2 eMeter System

Mattern et al. [18] also used the iPod/Phone for display of their eMeter. Despite the similarities in display, the eMeter represents a different approach in visualizing energy consumption: It is based on a single sensor, collecting and analyzing "total domestic load"[18, p. 6].

The eMeter was developed as part of a paper discussing how information and communication technology can help conserving power and "inducing behavioral change by providing direct feedback on household electricity consumption" [18, p. 1] whilst consuming power itself.

Referring to [6], the authors propose using ICT based visualization of energy consumption for detecting energy sinks that provide little to none advantage for the user, such as devices that do not use proper stand by modes when not used or coffee makers keeping the pitcher or even the cups warm with a terrible degree of power efficiency.

After analyzing experiments with smart meters installed in private homes, it seems the financial gain of conserving energy is not attractive enough to keep people committed to saving energy; a solution for display of energy consumption with the goal of reducing it has to be both appealing and involve users. The authors introduce three psychological methods for further involving users: Goal setting, energy budgets and comparison with peers. Considering all those aspects, it is concluded that "effective energy feedback has to

- feature a low usage barrier,
- be presented on a device that is already integrated into users' daily life,
- be given frequently, in real time, and be available when needed, and
- provide the ability to apportion total electricity consumption."

[18, p. 6]

The eMeter incorporates this features, feeding a single sensor's data into a lightweight web server with a SQL database and a Java frontend. The user interface is implemented in Objective-C, accessing data from the frontend by calling URLs, the GUI can be seen in figure 3.

Although the eMeter only uses one sensor, the interface can be calibrated to single devices by switching them off and on while measuring the total energy consumption. Although multiple sensors would provide better data on device level, this approach is very interesting, but fails measuring devices that cannot be switched off and on that easily such as washing machines. The approach of using

Fig. 3. eMeter user interface (from left to right): current consumption view, history view (aggreg. consumption), history view (budgeting), device inventory view, measurement view [18]

characteristics of single devices to distinguish the load using the data of only a single sensor was previously proposed by Weiss[30], who also works on the eMeter.

Mattern et al. conclude that concepts from behavioral psychology should be applied both for designing and evaluating user interfaces with display of energy consumption with the goal of reducing it, but do not supply a user study for their eMeter.

4.3 Web-enabled Power Outlets

Weiss and Guinard [29] moved away from the single sensor approach and provide a mobile web based application for monitoring and managing multiple power meters, that are plugged in between power outlets and devices.

The Plogg power meters used are available on the market [16] and already include a wireless IEEE 802.15.4 adapter; they are mainly designed for controlling and automating powering on and off attached devices in corporate environments.

Weiss and Guinard focused on "easy deployment", which is given with the Plogg meters as they can be bought ready to use, and "fine-grained" aggregation and feedback of energy consumption data. As shown in Figure 4 the architecture consists of four layers, with the devices to measure on the bottom, the mobile client on the top and meters as well as a web server collecting the meter data and providing the web interface in between.

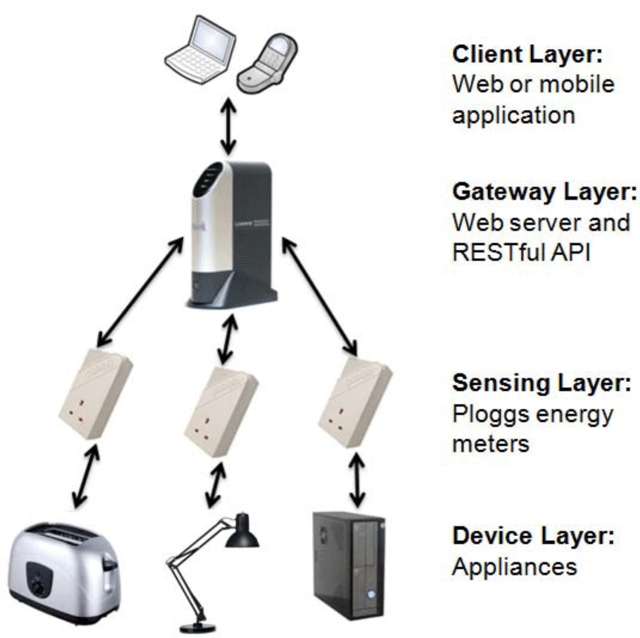


Fig. 4. Appliances connected to Ploggs communicate with a Smart Gateway offering the Ploggs' functionality as a RESTful API. Two client applications with user interfaces are then built on top of the Gateway Layer [29]

The web server connects to single sensors via Bluetooth or Zigbee and loops the data of each one through to the frontend as well as stores it in a database; the mobile client can also start and shut down devices. Previously, Weiss and his colleagues treated this possibility with retention, given the possible dangers of remote controlled complete control mentioned in [18, p. 8] such as computer viruses and denial of service attacks.

As the client only needs to log in to a wireless network and be able to POST and GET HTTP commands, the client is not limited to any particular device or programming language. Weiss shows two different implementations: A detailed JavaScript web interface to be accessed by a browser (Figure 5) and an iPhone app (Figure 6). The first is meant to be used with a bigger screen like the one of a desktop computer or notebook, the latter interface is optimized for mobile usage.

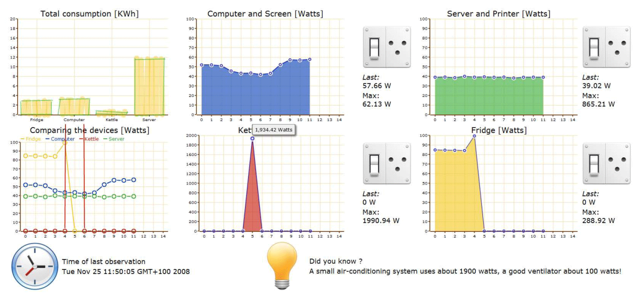


Fig. 5. The monitoring and control web user interface for the Ploggs shows the consumption of each connected appliance. The switch icons can be used to power on / off the devices [29]

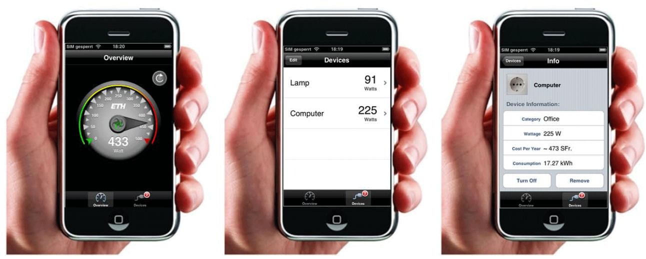


Fig. 6. The mobile user interface shows the current entire consumption of all devices attached (left), a list of discovered Ploggs including their name (middle), and more detailed information per Plogg (right) [29]

The software for the web server and web UI can be downloaded as "Energie Visible" package at no cost at [28]. The prototype is evaluated in a constant field test at the offices of Cudrefin 02, a Swiss foundation for ecological awareness, since December of 2008 [29, p. 8].

The system is used there to monitor power consumption of various devices such as printers, computers and a fridge and provide employees as well as visitors with real time data and thus continuous feedback of how their actions affect energy consumption.

As a result, after familiarizing themselves with the system, employees began to develop awareness about how much energy which device needs for what tasks. Consequently they tried minimalizing wasted power by turning stand by devices off completely in the evening and even shutting their fileserver down over night. However it was also noted that the system was used less after the first period of curiosity, which lasted for about a month, and thus Energie Visible and similar projects need ways and means for motivating users on a long term scale.

Weiss concludes it is mandatory for long term motivation to not only show feedback of current usage but also accumulate the data and provide consumption statistics as well as showing the monetary value of energy consumed or saved.

The mobile UI for Apple devices was evaluated separately in a focus group consisting of experts from industry, academia and consumer organizations.

One main point of critique about the mobile interface was, users knowing little about physics can not relate to units like kilowatt or CO_2 and do not know if a given value is good or bad. Possible solutions are displaying values to compare the one measured with, either from previous usage data, friends' usage data or intuitive real world examples like how far a car can drive until the given amount CO_2 is produced. Long term user motivation was another question, as it was with the web UI. Possible motivations could be energy alerts or consumption related video games; one possible game design mentioned is showing a wattage (e.g. 900W) from this household an having the residents guess which device consumes this much power when turned on (in this example it could be the microwave oven).

To further improve the design and functionality of the mobile UI, a short survey with 185 participants was conducted, asking them which feature they would find useful for themselves in the UI. The most prominent answers were in descending order of occurrence: device yearly cost, last month consumption, energy guzzlers, energy efficiency grade, comparison to average household and consumption of friends.

Sadly, no experiment was conducted how much total savings this project would induce in a household.

4.4 ACme

Jiang et al. [15] describe a quite similar system in great technical detail:

ACme consists of nodes, wireless capable power meters to install between power outlets and devices, a wireless IPv6 network connecting the nodes to the internet and an application with frontend, database and daemon service as shown in figure 7.

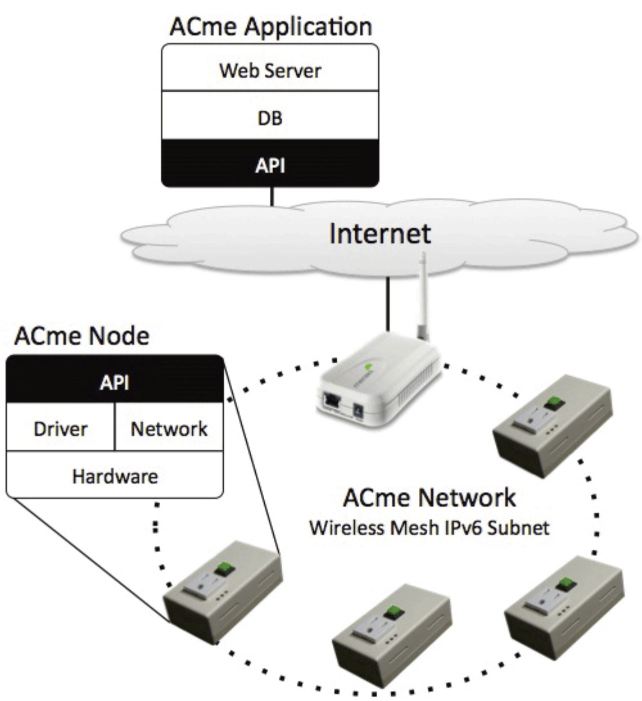


Fig. 7. Three tier ACme system [15]

All parts of the ACme prototype were built especially for this task, so users cannot just buy off-the-shelf products like the Ploggs and only need to install custom software, if they want to try this at home.

Otherwise, due the very similar general design to the Web-enabled Power Outlets by Weiss and Guinard, the same up- and downsides already mentioned also apply here, at least for the hardware; the software is only designed to save and provide the data collected and does not contain a graphical user interface.

As the system was still under development when the paper was written, no user studies if this appliances can aid users in managing and decreasing energy consumption were conducted until then.

4.5 The Power-Aware Cord

The Power-Aware Cord by Gustafsson and Gyllenswärd [14] uses a different approach in visualizing energy consumption: Their are no detailed statistics that can be archived and analyzed, but an ambient display included in a power cord as seen in figure 4.5.

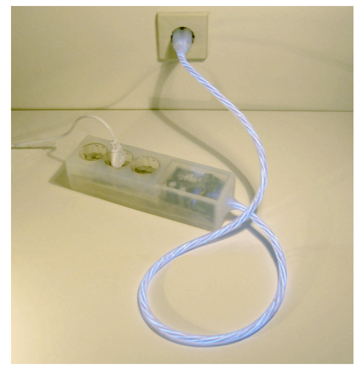


Fig. 8. The Power-Aware Cord prototype [14]

The glow is emitted by three electro-luminescent wires, that are twisted around the electric wires and are controlled by a micro processor, that uses data from a small integrated power meter. One goal of designing this cord was to visualize small loads from stand by devices of about 10 Watts as well as the maximal load of about 2200 Watts distinguishable.

Conducting a user study with 15 participants, only two did not see the glow of the power cord as representation of energy consumption; those two were not confronted with changing load. The evaluation of the user study led to the conclusion that most customers would prefer a different intensity of constant glow over flowing and pulsing glow, although the constant glow was evaluated as the least informative.

This cord can be seen both as conceptual design that encourages further research in ambient power displays and, if it was commercially available, an energy aware piece of furniture with appealing looks. Although detailed information about energy consumption is considered to be a stronger incentive for conserving it, this cord could target especially the group of people, that are not particularly interested in technical details and appliances.

In [1, p. 13] it is mentioned, that several prototypes of the Power-Aware Cord are deployed in Stockholm's private houses as part of a long term study, no results are yet available.

4.6 The Local Energy Lamp

The Local or Seasonal Energy Lamp described by Pierce [25, p. 8] and Paulos [22, p. 5] tries to visualize not power consumption but "metadata". In analogy to HCI metadata, which is information included in pictures, songs and videos about the author, title and other attributes, Paulos suggests to implement metadata, in this case the source, into electric power.

Fig. 9. The Seasonal Lamp [25]

The lamp shines in a different color, depending on what energy source is used at the moment to power it: White for the power outlet, yellow for local solar power, blue for local wind power and red for locally human generated power, as shown in Figure 9. The prototype requires wiring to a power outlet as well as the other energy sources used like solar panels on the roof and a wind turbine in the garden.

This is an interesting prototype for greatly enhancing energy awareness, but only suits homes were there is at least one kind of electric power generated locally.

4.7 The Heat-Sensitive Lamp

Another project utilizing a lamp aims at illustrating the energy used in a very radical display: The Heat-Sensitive Lamp introduced in the Static! project [1, p. 15] is not meant to be an alternative to a power meter but fails not to display the pure force of electricity.

When turned on, the thermal energy emitted by the light bulb causes the material of the lampshade, which is made of a "sensitive paper-like material", to deform. The more thermal energy is emitted, the more drastic is the effect on the lampshade, like seen in Figure 4.7.

Fig. 10. The Heat-Sensitive Lamp [1]

While this project neither uses computers nor displays exact usage statistics in real time, it is nonetheless suitable to demonstrate how much of the energy consumed by an ordinary light bulb does not transform into light intensity but heat and thus might be an impulse to reconsider one's energy usage and develop a degree of energy-awareness.

There are, sadly, no evaluations that would quantify those possible effects.

5 Starting points for further reading

As already mentioned, the academic research in the disciples of visualizing power and inducing conserving behavioral changes receives quite much attention.

The work of Markus Weiss and his colleagues at ETH Zurich shows, how power meters and devices for evaluation of consumption can be integrated in private households, and refers to a lot of important publications in this field.

For the aesthetic aspects of design of devices related to human computer interaction and ubiquitous computing, James Pierce and Eric Paulos from the Carnegie Mellon University in Pittsburgh, USA, as well as Anton Gustafsson and Magnus Gyllenswärd from the Mälardalen University in Sweden contribute a lot of work on this topics and are good starting points for further research.

6 Conclusion

Conserving natural resources and minimizing CO_2 emission as well as production of nuclear waste by reducing consumption of electric energy is not optional, it is mandatory for preserving our world as good as we can for following generations.

Computers, although needing energy themselves, can be of valuable assistance for avoiding unnecessary waste of energy, as we have seen. When for example looking at the experiment by Petersen et al. [24], in which energy consumption of a whole dormitory building could be reduced by over 50 percent, when given incentive and high resolution feedback by turning the lights off when leaving a room and taking similar, simple measures, there is a big deal of potential for conserving energy without sacrificing our lifestyles.

The other experiments also show, that feedback itself is quite an incentive, when the intangible, invisible electric current is illustrated to the user, he may yet realize what devices use how much energy and take countermeasures himself. When 15 percent savings for cooking and 22 percent savings for washing can be achieved in experimental situations by showing real time data and maybe setting goals, it may be assumed that similar results could be achieved for other devices.

Assuming a 10 percent decrease in domestic energy consumption would mean a 2.7 percent decrease of total energy consumed in Germany, resulting in 17 terawatt hours of power, 9,692,157,600kg of carbon dioxide and 12,051kg of radioactive waste saved. This seems to be an unimaginable amount, but also a reachable goal when looking at the outcome of the experiments.

Both HCI and psychology conduct studies about how information about energy consumption affects energy saving.

Froehlich et al. [12] evaluated different studies from both HCI/Ubicomp and environmental psychology fields of research and came to the conclusion, that nearly all of the studies from environmental psychology exclusively focused on reception and effect on behavior, but not on the design of the interface, which should aid the test persons in increasing energy awareness and reducing consumption. None of those studies referred to HCI designs and approaches; field studies are conducted with an average of about 200 participants and for over half a year.

On the other hand, HCI and Ubicomp studies emphasize on technical properties and design of the interfaces. For explanation of their findings, about a half of the evaluated studies refer to environmental and/or behavioral psychology. The user studies average to only about 10 participants, the studies are often part of an iterative process for improving the interface.

When looking at this, collaboration between those disciples could greatly improve both development and real world examination of interfaces for energy awareness.

A topic not boldly covered in this article is the power consumption of devices turned off and thus being in a stand by state, a bad habit of many devices. Because this state provides little to no advantage to the user while using up to 10 Watt and thus might cost 22 Euro a year per device, this should be avoided by physically disabling access to power (i.e. by using switchable power cords or multi-plugs) for those devices where ever possible.

Speaking of power cords, the Power-Aware Cord was one example of exciting projects which aim at a more energy aware and conservative handling of electric power by visualizing it.

A maybe less aesthetic, but highly informative design is the wireless connection of one or more power meters, accumulation of data and evaluation via software, as WattBot, eMeter, Web-enabled Power Outlets and ACme offer. Comparing single- and multi-sensor approaches, the multi-sensor seems to be the better choice for fine-grained, distinguishable analysis of consumption data from different devices:

Knowing that the whole system or one room related to a fuse contains an energy guzzler and searching for it seems a lot less fun than just looking at one's smartphone and seeing, that the BluRay player is in stand by and consumes a ridiculous high amount of energy for doing nothing, before just cutting its power remotely.

In his Ph.D. thesis [13], Gustafsson introduces, next to projects as the Power-Aware Cord, video games with the goal of increasing energy awareness in a fun way. He underlines the importance of social interaction within this games, for example in the game Power Agent "two teams of teenagers compete together with their families at saving energy in their homes".

Comparing the energy consumption to the one of friends or neighbors might not always have a positive effect on conserving behavior, Schultz et al. [27] show that "for households that were initially low in their base rates of energy consumption, the same descriptive message produced a destructive boomerang effect, leading to increased levels of energy consumption", however, the experiment could be changed in a way so that this effect was no longer present.

All of the projects shown aim at increasing energy awareness. I do not think different approaches as for example the eMeter and the

Power-Aware Cord can or should be compared only in terms of pure energy reduction during an user study; they target different users and can each for itself or in combined use be of great assistance in achieving what is needed: Energy aware citizens to face the challenges of the twenty-first century.

REFERENCES

[1] S. Backlund, M. Gyllenswärd, A. Gustafsson, S. Ilstedt Hjelm, R. Mazé, and J. Redström. Static! the aesthetics of energy in everyday things. Technical report, SICS publications database [http://eprints.sics.se/perl/oai2] (Sweden), 2007.

[2] J. Bortz and N. Döring. *Forschungsmethoden und Evaluation*. Springer Verlag Berlin Heidelberg New York, 2. edition, 1995.

[3] G. Brandon and A. Lewis. Reducing household energy consumption: a qualitative and quantitative field study. *Journal of Environmental Psychology*, 19(1):75–85, 1999.

[4] S. Bundesamt. Energie auf einen blick 2009. `http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Publikationen/Fachveroeffentlichungen/Produzierendes_20Gewerbe/EnergieWasserversorgung/EnergieBlick,property=file.pdf`, 2009. visited 21.06.2011.

[5] Bundesumweltministerium. Klimaagenda 2020: Der umbau der industriegesellschaft. `http://www.bmu.de/files/pdfs/allgemein/application/pdf/hintergrund_klimaagenda.pdf`, 2007. visited 21.06.2011.

[6] M. Chetty, D. Tran, and R. E. Grinter. Getting to green: understanding resource consumption in the home. In *Proceedings of the 10th international conference on Ubiquitous computing*, UbiComp '08, pages 242–251, New York, NY, USA, 2008. ACM.

[7] B. der Energie-und Wasserwirtschaft (BDEW). Bdew-musterhaushalt für strom 2011. `http://bdew.de/internet.nsf/id/DE_20100311_PM_46_Prozent_des_Strompreises_sind_Steuern_und_Abgaben`, 2011. visited 21.06.2011.

[8] M. Deutsch. Life cycle cost disclosure, consumer behavior, and business implications. *Journal of Industrial Ecology*, 14(1):103–120, 2010.

[9] S. B. Deutschland. Haushalte nach haushaltsgrößen. `http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Statistiken/Zeitreihen/LangeReihen/Bevoelkerung/Content100/lrbev05ga,templateId=renderPrint.psml`, 2010. visited 21.06.2011.

[10] J. K. Dobson and A. J. D. Griffin. Conservation effect of immediate electricity cost feedback on residential consumption behaviour. In *American Council for an Energy-Efficient Economy: Summer Study on Energy Efficiency in Buildings*, 1992.

[11] A. Energiebilanzen. Entwicklung der spezifischen kohlendioxid-emissionen des deutschen strommix 1990-2009 und erste schätzung 2010. Technical report, Umweltbundesamt (UBA), 2010. visited 21.06.2011.

[12] J. Froehlich, L. Findlater, and J. Landay. The design of eco-feedback technology. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 1999–2008, New York, NY, USA, 2010. ACM.

[13] A. Gustafsson. *Positive Persuasion - Designing enjoyable energy feedback experiences in the home [Elektronisk resurs]*. PhD thesis, University of Gothenburg, 2010.

[14] A. Gustafsson and M. Gyllenswärd. The power-aware cord: energy awareness through ambient information display. In *CHI '05 extended abstracts on Human factors in computing systems*, CHI EA '05, pages 1423–1426, New York, NY, USA, 2005. ACM.

[15] X. Jiang, S. Dawson-Haggerty, P. Dutta, and D. Culler. Design and implementation of a high-fidelity ac metering network. In *Proceedings of the 2009 International Conference on Information Processing in Sensor Networks*, IPSN '09, pages 253–264, Washington, DC, USA, 2009. IEEE Computer Society.

[16] E. O. Ltd. Plogg. `http://www.plogginternational.com/`, 2011. visited 21.06.2011.

[17] I. Matsukawa. The benefits of information on the efficient usage of consumer durables. Others, EconWPA, 2005.

[18] F. Mattern, T. Staake, and M. Weiss. Ict for green: how computers can help us to conserve energy. In *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking*, e-Energy '10, pages 1–10, New York, NY, USA, 2010. ACM.

[19] L. T. McCalley and C. J. H. Midden. Energy conservation through product-integrated feedback: The roles of goal-setting and social orientation. *Journal of Economic Psychology*, 23(5):589–603, 2002.

[20] L. McClelland and S. Cook. Energy conservation effects of continuous in-home feedback in all-electric homes. In *Journal of Environmental Systems 9*, page 169Ű173, (1979Ű1980.

[21] J. Nipkow and S. Gasser. Der typische haushalt-stromverbrauch, 2007.

[22] E. Paulos and J. Pierce. Citizen energy: Towards populist interactive micro-energy production. In *Proceedings of the 2011 44th Hawaii International Conference on System Sciences*, HICSS '11, pages 1–10, Washington, DC, USA, 2011. IEEE Computer Society.

[23] D. Petersen, J. Steele, and J. Wilkerson. Wattbot: a residential electricity monitoring and feedback system. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, CHI EA '09, pages 2847–2852, New York, NY, USA, 2009. ACM.

[24] J. E. Petersen, V. Shunturov, K. Janda, G. Platt, and K. Weinberger. Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives. *International Journal of Sustainability in Higher Education*, 8:16–33, 2007.

[25] J. Pierce and E. Paulos. Materializing energy. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, DIS '10, pages 113–122, New York, NY, USA, 2010. ACM.

[26] B. Schlomann, E. Gruber, W. Eichhammer, N. Kling, J. Diekmann, H.-J. Ziesing, and H. Rieke. Energieverbrauch der privaten haushalte und des sektors gewerbe, handel, dienstleistungen (ghd). abschlussbericht des fraunhofer-isi karlsruhe, diw berlin, gfk nürnberg, institut für energetik und umwelt leipzig, und tu münchen an das bundesministerium für wirtschaft und arbeit[energy consumption of private households and the tertiary sector (trade, commerce and services). final report to the federal ministry of economics and labour]. Technical Report Projektnummer 17/02, Fraunhofer-ISI Karlsruhe, DIW Berlin, GfK Nürnberg, Institut für Energetik und Umwelt Leipzig, und TU München, 2004. visited 21.06.2011.

[27] P. W. Schultz, J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius. The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18:429–434(6), May 2007.

[28] R. Uneregger and D. Guinard. Energyvisible. `http://webofthings.com/energievisible/`, 2010. visited 21.06.2011.

[29] M. Weiss and D. Guinard. Increasing energy awareness through web-enabled power outlets. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*, MUM '10, pages 20:1–20:10, New York, NY, USA, 2010. ACM.

[30] M. Weiss, F. Mattern, T. Graml, T. Staake, and E. Fleisch. Handy feedback: connecting smart meters with mobile phones. In *Proceedings of the 8th International Conference on Mobile and Ubiquitous Multimedia*, MUM '09, pages 15:1–15:4, New York, NY, USA, 2009. ACM.

[31] H. Wilhite and R. Ling. Measured energy savings from a more informative energy bill. *Energy and Buildings*, 22(2):145 – 155, 1995.

[32] G. Wood and M. Newborough. Dynamic energy-consumption indicators for domestic appliances: environment, behaviour and design. *Energy and Buildings*, 35(8):821–841, 2003.

[33] www.bdew.de. Datenerhebung 2009 bundesmix 2009 (stand 14.09.2010). Technical report, BDEW Bundesverband der Energie- und Wasserwirtschaft e.V., 2010. visited 21.06.2011.

[34] www.bdew.de. Leitfaden stromkennzeichnung. Technical report, BDEW Bundesverband der Energie- und Wasserwirtschaft e.V., 2010. visited 21.06.2011.

Programming the mechanical turk

Bernhard Hering

Abstract— This paper gives an overview about crowd-sourcing services. It presents programing models and tools, especially turkit is mentioned and explained. Furthermore the importance of a good user-interface is described. Crowd-sourcing is a possibility to fight poverty in emerging countries. In this paper mostly the example Amazon mechanical turk is used. We look at the demographics of the turkers as well as on their motivation and the quality of the solutions they make. There are several different kinds of crowd-sourcing services, which are listed and classified in this paper. Each of them has its special characteristics. Some are specialized on mobile task distributing, others are only a human intelligence driven question-and-answer machine. Furthermore crowd-sourcing services are a good possibility to get data for surveys. Obviously crowd-sourcing services have to deal with some limits like motivations problems of turkers, or a lack of demographic representativeness.

Index Terms—crowd-sourcing, mechanical turk, turkit, outsourcing, human computing

1 Introduction

In the late 18th century, Wolfgang von Kempelen invented his 'mechanical turk', a automatic chess machine. (See figure 1) This machine was a desk with a chess field on it. Behind the desk there was a statue of a turkish man. The machine won nearly every chess game it played. This circumstances lead to the name mechanical turk. But in fact the Machine was a fake. Behind some gear wheels a chess master was hiding inside. [28]

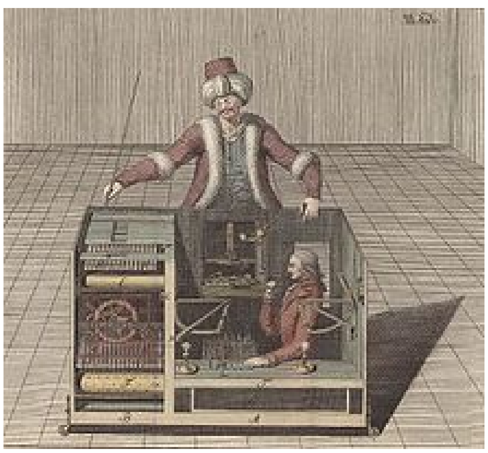


Fig. 1. Wolfgang von Kempelens 'mechanical turk. [28]

Nowadays there are still tasks left a computers cannot do for humans. Examples are to label images, to recognize bad handwriting, or to categorize articles. These are all tasks where content, like pictures or the sense of an article has to be compared with background knowledge, for example "How does a horse look like?'.' For humans it might be easy to tell the difference between a cow and a horse, but it is still hard to teach a computer.

For example, for humans it is easy to detect a red Ferrari on a picture even if there a many shadows which make some parts of the car look dark red and other parts bright red. For computers such a task is difficult if it just has the information that a red Ferrari is red and every part of the Ferrari is red. The computer could calculate the shadow, if it makes assumptions on the light situation when the picture was made . But to determine such assumptions is not easy.

Computer scientists invented a model they call 'mechanical turk'. In such a system, micro tasks, called HITs (Human Interaction Tasks), are outsourced to human workers. These workers, we call 'turkers' in the following, have access to tasks over a webpage or another interface. By that, people from all over the world can participate and each HIT is payed with a specific amount of money, typically a few cents. Hence, the work could be distributed to people who need money and work in the moment the task has to be fulfilled. Sometimes people even have fun doing these tasks. Later in chapter 3.2 there is more about the motivation of the turkers.

This short paper gives an overview about programming models to implement HITs as well as interfaces for turkers. Furthermore it takes a view on social aspects of the work as a turker. Who is doing such a job? Who works for how long on how many HITs? The demographic background will be taken into account to give more insights. Another important aspect in this research is how to motivate the workers. In chapter 3 social background, the motivation and the quality of the turkers' work is considered. Furthermore this paper presents some algorithm how to check and improve the quality of the turkers' work.

Amazon's mechanical turk is the most popular crowd-sourcing service but there are others to mention. For example, non-profit services and services specialized on detecting new galaxies. Many services use the huge man power of emerging countries and even provide their tasks on mobile phones. Chapter 4 describes these services and their characteristics.

Finding participants for surveys is always difficult. So people try to use the mechanical turk service to get fast and cheap results. The advantages and disadvantages of mechanical turk as a user study platform is looked in chapter 5.

2 Tools and Models

There are many possibilities to implement crowd sourcing computation. On the one hand there are services like Amazon's mechanical turk which provide solutions for short independent tasks. On the other hand there are wikis which are platforms where different people try to solve more complex depending tasks like the wikipedia encyclopedia. [29].

Amazon Mturk is mainly used for independent Tasks. This is easy to implement because you can publish all tasks at once and wait for their completion.[19]

Zuang et al. [29] describe basic algorithms which are good for human computation. First they mention divide-and-conquer. This method is roughly described by dividing a problem into subproblems which can be solved independently. Such algorithms are very good for parallel processing and are ideal for applications using human computation. [17]

- *Bernhard Hering is studying Media Informatics at the University of Munich, Germany, E-mail: —@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

One example for a divide-and-conquer algorithm is Quicksort. Quicksort is a sort algorithm that use the divide and conquer pattern. Still, the performance of the human component has to be evaluated in detail. For example, Quicksort would most likely perform poor when sorting restaurants. Thats why Zuang et al. [29] suggest to look for a good searching algorithms depending on people's performance on different tasks. So as conclusion they judge the algorithm efficiency with "respect to the availability and cost of human oracles of varying complexity" [29].

2.1 Turkit

Little et al. present Turkit. It is a Toolkit that allows users to use the mechanical turk service for depending tasks. Figure 2 shows an example of depending tasks. First the question "What's fun to see in New York City?" has to be answered, afterwards the answers could be sorted.

```
ideas = []
for (var i = 0; i < 5; i++) {
    idea = mturk.prompt(
        "What's fun to see in New York City?
         Ideas so far: " + ideas.join(", "))
    ideas.push(idea)
}

ideas.sort(function (a, b) {
    v = mturk.vote("Which is better?", [a, b])
    return v == a ? -1 : 1
})
```

Fig. 2. Example of depending Tasks. [18]

Implementing algorithms on Mturk is difficult. Human Intelligence Tasks (HIT) may take some time to complete and cost money. So programmers have to think on some special cases, like what about the results if the machine running the program crashes. Or what about throwing exceptions after completion of some HITs? [18].

So facing these questions Little et al.[18] introduce the crash-and rerun programming model. The main idea behind is to rerun a program without the necessity of costly turkers' work.

How does Turkit work?

A javascript program is sent to the Turkit platform. Turkit runs the program and creates the HITs. While running it stores information about the completed HITs in a database. With information in the database the program can crash and rerun without the necessity to repeat costly work. This is implemented with the *once* function. That means it is possible to pass *once* as an argument to every function that should be run only once. After completing a function with the once argument the result is stored in the database for later use. E. g. if the program crashes and has to be rerun.

2.2 The Find-Fix-Verify Pattern

Implementations of crowd-sourcing algorithms have to face various problems. One of them is the danger of bad results. (See chapter 3.3) Bernstein et al. describes the Find-Fix-Verify Pattern which is especially designed to improve the reliability of the results. Although the pattern is for a special task, it is worth mentioned here.

This pattern separates tasks into three stages (compare Figure 3 which is applicable to the special task of Soylent(see chapter 4)). In the first stage the turkers identify patches of the task on which they think further turkers have to look at with more attention. In Soylent for example it is to mark the parts of the phrases or the sentences that needs editing. Afterwards a predefined threshold is considered. This Value, for example 20%, means that if 20% of the workers agree with one patch it will be hold for further review.

The next stage is to fix values. Here only a small number of workers is needed. They see the whole task but can only edit the marked patches from stage one.

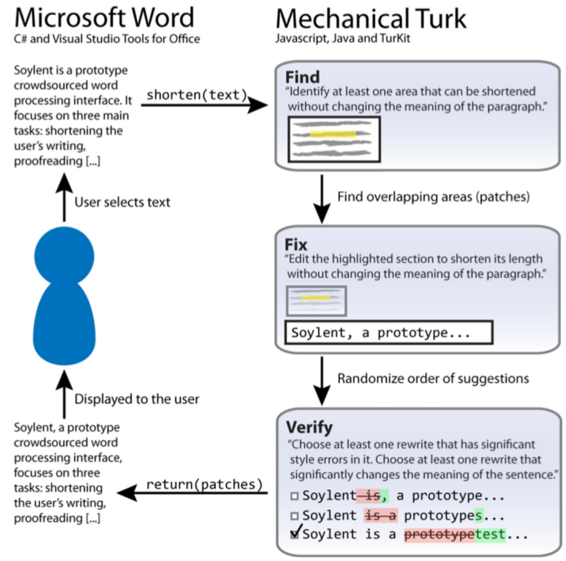


Fig. 3. The tree stages of the Find-Fix-Verify Pattern in Soylent. [2]

For stage three there are also only few workers necessary. Here they get a random subset of the results produced in stage two to verify. It is important that you ban the workers who did stage two from participation in this stage.

2.3 User Interfaces for turkers

To find many people who want to use Amazon's mechanical turk, it is important to keep entry barriers low. This means that the interface has to be easy to understand. People with very low digital literacy skills as well as 'Digital Natives' have to be capable of using the system. Khanna et al. [16] made a survey about the usability of a mechanical turk for low income workers. 13% of all Indian workers who have finished 10-12 years of schooling, earn less then $1700. But only 3% of all Indian turkers are from this group. So why do not more of them use such a system?

Khann et al. identified five difficulties users are facing using Amazon's mechanical turk. [16]

- **Complexity of instructions.** The problem is that most of the Indians are no good english speakers. Primarily, they had problems in understanding the instructions. For example they associated the word keyword with something that was typed with a keyboard.

- **User interface complexity.** The interface of mechanical turk has too much functionality for the turkers. There are always buttons for controlling the specific task as well as for general preferences. The participators were confused.

- **Navigation difficulties.** People had problems with the navigation in the tasks. The interface has more than one nested scrollbar. The turkers sometimes didn't see parts of the tasks because they used the wrong scrollbar. The back button of the browser also caused problems, because the participants used it and lost data.

- **Sequencing problems.** Some people had problems with the workflow in the tasks. For example they forgot to press the "accept" button that starts the processing of the tasks. Without that the turker can't earn money.

- **Cultural context.** Differences between cultures are a big thing when you are working with low income workers. For example, one participant had problems with a CAPTCHA because he read the letter from top-to-bottom. Another had problems to recognize a Western-syle kitchen because it has no gas stove. He labeled it as a bedroom.

Based on these results Khann et al. [16] suggest four guidelines for an interface.

- Use simple, illustrated instructions for each task
- Minimize visual complexity
- Streamline navigation
- Anticipate sequencing of steps

With these guidelines they are not only facing one aspect of improving the usability. They also determine space for development in language localization and video tutorials.

3 Social aspects of mechanical turk

Amazon's mechanical turk was launched 2005. It is the most popular micro-task service in the world. In this chapter we take a closer look at the turkers working for Amazon's Mturk. Everybody in the world who has access to a computer with internet connection can participate in this network. Independent of time and place people can do micro-tasks and earn money. This chapter gives an overview which people use this opportunity and why.

3.1 Demographics of Turker

The majority of the turkers is from the US or from India. This has many reasons. At the beginning of Amazons Mturk, only workers with a bank account in the US got payed in money. Turkers without a US bank account were payed with Amazon.com gift cards. [14] So in 2008, 76% of all workers were from the US [13]. The data used in this chapter is based on surveys described in "Who are the Crowdworkers? Shifting Demographics in Mechanical Turk", Ross et all.[22] and "Demographics of mechanical Turk", Ipeirotis. [14]

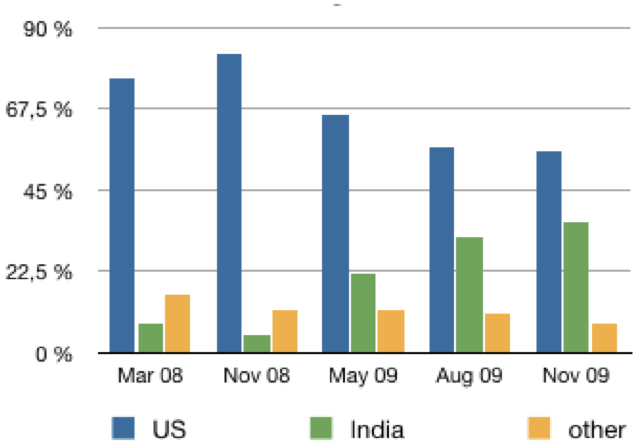


Fig. 4. Nationality of MTurk workers over time. You see the increasing percentage of Indian workers. Countries with more than 1% of respondents include Canada, the U.K. and the Philippines. There is no country data for Feb. 2009. [22]

Since 2008 the proportion between US turkers, India turkers and turkers from other countries in the world has changed. See figure 4. Ross at al. got the same data as Ipeirotis for 2008. So after nearly 2 years, in autumn 2009 only 56% of turkers are US citizens. Dates from Ipeirotis even say only 46% in February 2010.

There is also a change in the average age. 2008 the average age of turkers was 32.9. This changed to 31.6 in 2009 with an increase of workers int the range of 18 to 24 years of age. See Figure 5.

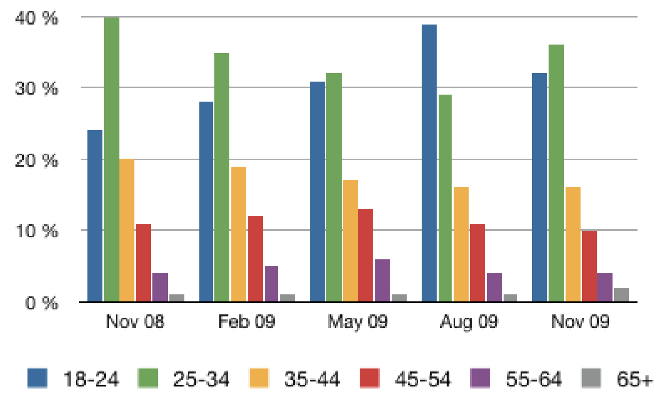


Fig. 5. Gender of MTurk workers over time. Average is decreasing because the percentage of young workers is increasing. [22]

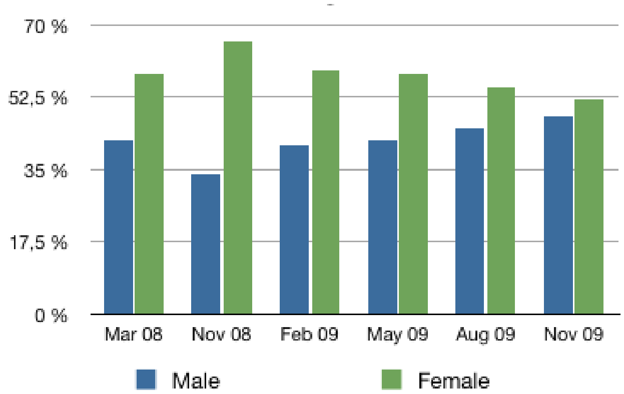


Fig. 6. Age of MTurk workers over time. percentage of male workers increased steadily. [22]

The proportion between women and men also changed since 2008. The percentage of male workers increase fairly steady. See Figure 6.

Ross et al. as well as Ipeirotis also discuss the income level and material status of workers. First, Ipeirotis assert that significantly more workers from India participate on mechanical turk because the online marketplace is their primary income source, while in the US most workers consider mechanical turk a secondary income. [14] You can see this also in figure 7. In the US people with various income level participate in the network, whereas in India mostly people with low income work as turkers. Overall the percentage of workers with low income is increasing. However Indian turkers are consistently younger than workers from the US.

The data of Ipeirotis and Ross et al. is based on surveys made with Amazons MTurk. Heymann and Garcia-Molina [8] also made a survey about demographics data from MTurk. They developed a tool to store IP addresses workers. In 2011 they state 44% workers to be from the US and 38% from India. This is very similar to data of other surveys.

3.2 Motivate the turkers

Ipeirots [14] collected data regarding the reason of the motivation of turkers. He tries to answer the following questions:

Why do you complete tasks in mechanical turk?

1. Fruitful way to spend free time and get some cash (e.g., instead of watching TV)
2. I participate on mechanical turk because the tasks are fun
3. I participate on mechanical turk to kill time

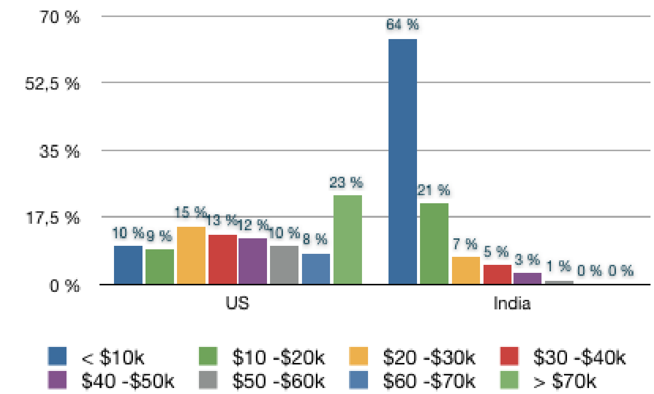


Fig. 7. reported annual household income by country, from Nov. 2009.. [22]

4. Mechanical turk is my primary source of income (paying bills, gas, groceries, etc)
5. Mechanical turk is my secondary source of income, pocket change (for hobbies, gadgets, going out)
6. I am currently unemployed, or have only a part time job

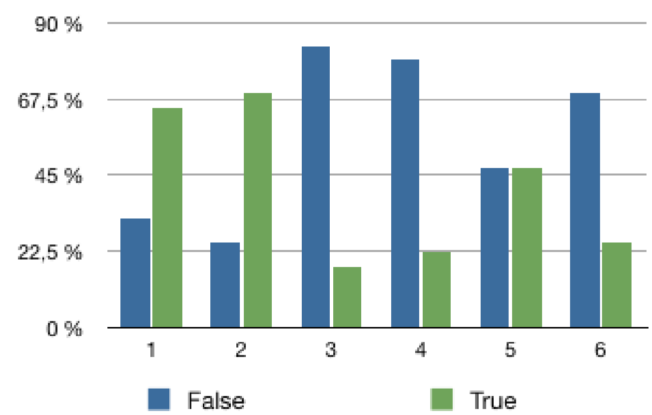


Fig. 8. Answers to the question above. Summarization of Ipeirotis results.[14]

Most of respondents agree with the first question but only 30% of the probands say that tasks are fun. In this question opinions are different between the US and India. US workers have more fun with the tasks than Indian workers. This might explain the circumstance that also people with high income form the US participate in the network. With question three it can be seen that most of the people use the platform as possibility to earn money. But only half of them have fun to complete tasks. (See question 5)

This means that the main reason for participating in this network is to earn money. But how could you motivate turkers apart from money? Von Ahn and Dabbish [27] created their ESP game which can be played `http://www.espgame.org`. A game that is fun and can be used to create valuable output. Their game helps to label images. So if this game would be played as often as other online games, like FarmVille most images of the web could be labeled in just a few month. [5]

Horton and Chilton [9] introduce a workers reservation wage in "The labor economics of Paid Crowd-sourcing". This is 'the smallest wage a worker is willing to accept for a task ..." [9]. They show that the reservation wage of a sample of workers from mechanical turk are willingly to work at $1.38 per hour.

3.3 Quality of results

The turkers on Amazon's mechanical turk are anonymous and they get less money when completed a task compared to other platforms. This makes the people try to complete tasks as fast as possible to earn money with the network. When turkers are working fast there is the question how engaged they complete the tasks and of which quality of the results might be.

Bernstein et al. [2] define two types of workers. The 'lazy turker', as mentioned just before, does as little work as possible to get paid. For example if this type of turker has to proofread a error-filed text, he might only insert one character to correct only one word.

But there is also another type of turker which produce unusable work. 'Eager Beaver' [2] do more as they should and go beyond the task requirements. With this behavior they are no longer helpful but produce only more work for the owner of the task.

Snow et al. [24] published a survey in "Cheap and Fast - But is it Good? Evaluating non-expert annotations for Natural Language Tasks", about the quality of Amazon's mechanical turk output. They compared the results of five tasks with results from experts. These tasks covered recognition, word similarity, recognizing textual entailment, event temporal ordering, and word sense disambiguation.

The results are very interesting. In all five tasks there is a huge agreement between the turkers annotations and the gold standard annotations. But in the results you can also see that workers tried to finish the tasks as fast as possible. There are some who did a very large amount of tasks, but the result produced by them were often very bad. These are the same workers Bernstein et al. define as 'Eager Beaver' [2]. On Figure 9 you can see some circuits downright. This are the workers who produced very much low quality results.

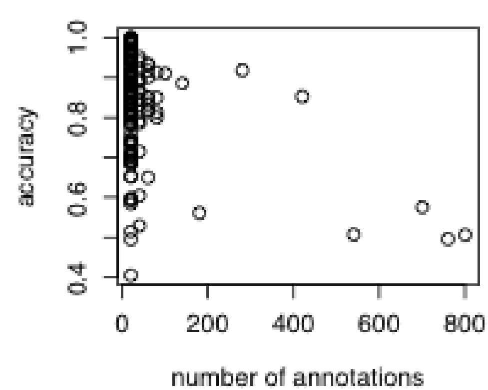


Fig. 9. Relation between accuracy and number of annotation. [24]

Callison-Burch [4] also evaluated the quality of the results. They came to a similar finding: "Non-expert annotators produce judgments that are very similar to experts and that have a stronger correlation than Bleu", Callison-Burch [4]. Bleu is a method for automatic machine translation evaluation, for further reading see Papineni et al. [21].

So Snow et al. suggest three ways to enhance the quality of the work. First you could use more workers, which improve the reliability of the data. This is the standard way. It is suggested also from Ipeirotis in "Quality Management on Amazon mechanical turk[15]. The problem is that massive redundancy is expensive. If you use too much workers the costs increase significantly. [15]. Amazon also provides a mechanism. The Amazon compensation mechanism that gives monetary bonuses to highly performing workers and deny payments to unreliable ones. The third possibility is to 'recalibrate' the workers. You train them with expert-labeled training data to correct the individual biases. [24]

A similar approach is described by Ipeirotis et al. in "Quality Management on Amazon mechanical turk" [15]. They suggest to use redundancy not only to evaluate the results but also to measure the labeling quality of the workers. They first identify a correct solution for one task by getting answers from multiple turkers. Afterwards they compare the solution with the answers of other workers to estimate the quality of their answers.

This technique is very similar to the Fix-and-Verify Pattern.

4 Currently running web-services

Most of papers referenced in this papers refer to Amazon's mechanical turk. This is the biggest and most popular crowd sourcing service in the web. The Service was founded in 2005, and there are now 123180 HITs available (14.6.2011). [1] Already in 2007, over 100000 worker from over 100 countries earn micro payments with this service. [20]

At Amazon's mechanical turk the tasks are very different. The users can add tasks like labeling pictures, or let them bring in categories. Another task is to let turkers sum up the content of a text. These and many other tasks are possible.

There are many other micro task services beside Amazon. Some services specialize on questions like "What is the percentage of professional football league players who do not have a college degree?" [10]. That means question that need some research to answer. Such services solve tasks similar to algorithmic search engines like Google or ask.com but no computers are answering the questions, but real people. One of the services was google's web-service called 'Google Answers' which was closed in 2006. Maybe thats because only 800 people were participating. [7]. Although this service was closed, some former researchers from Google Answers founded 'uclue'. [10]. In uclue many of the former Google Answer researcher signed up.

There are more similar web-services, for example: 'Mahalo.com', 'Answer.com' and 'Yahoo answers'.

Through the years also scientist discovered the power of crowed sourcing. There are services like 'Stardust@home' or 'galaxy zoo' were people classify pictures. Galaxy zoo asks users for the form and rotation of galaxies. The process is totally automatic. Huge telescopes make the pictures and send them as open tasks to 'Galaxy Zoo'. Then users can watch and classify them. So for the users, which are mostly hobby astronomers it is exciting, because they are the first ones who ever look at the galaxies. Figure 10 and figure 11 shows a sample of a task in Galaxy Zoo.

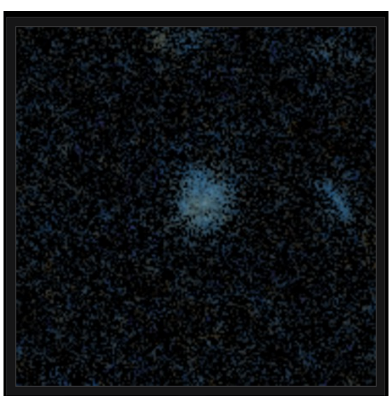


Fig. 10. Galaxy Zoo: What kind of galaxy is this? [11]

In chapter 3.2 the ESP game is mentioned. People also participate voluntary, like in the galaxy zoo. The difference of the ESP game is that the labeling task is integrated in a game: Every time a user logs in, he gets a partner which he does not know or see. So both players get the same picture to label. They put in words that describe the picture. When both enter the same word the picture is labeled and they get a new one. Aim is to label as many pictures as possible in 15 minutes.

But there are also further services which are very similar to Amazon's mechanical turk. First there is the non-profit organization 'Samasource'. It's aim is to "bring dignified, computer-based work to women, youth, and refugees living in poverty" [23]. Samasource tries to divide tasks into micro-tasks. These tasks will be fulfilled by low

Fig. 11. Galaxy Zoo: Choose one of the possibilities.[11]

income workers from all over the world. When the micro-tasks are answered Somasource puts them together and ensures their quality.

Texteagle [6] is also specialized on completing micro-tasks in emerging countries. But for this service workers do not need a computer, they only need a mobile phone. By collaborations with cell phone service providers the operating company has a very easy way to pay the workers. They pay in airtime that means the workers get some money on their prepaid cards or in MPESA. MPESA is a electric kind of currency used in Kenya for money transactions with mobile phones. 90% of all people living in East Afrika, have access to a GSM network. And many of them has one or more mobile phones, whereas many people are unemployed. Hence, there is very high potential in micro-task workers, which could not be reached if the service would only run on normal personal computers.

Another service is Soylent [2], which is not an independent service but is build on Amazon's mechanical turk. It is built into a Microsoft Word interface and provides help with shorten, proofread and various other ways of edit parts of documents on demand. [2]. This is unique because customers can select the service of Soylent in a very similar way like the normal artificial driven spell check in MS Word.

5 Mechanical turk as a user study platform

User studies are very important in prototyping and design processes. With user studies the interaction design as well as the usability can be improved significantly.

For every empirical study subjects are needed. In the best case, these people should reflect the target group and they are of significantly large number. This is usually not easy to reach and sometimes very exhausting. Common methods are sending letters directly to households or trying to convince people in the streets to answer questions.

The design of the study highly depends on money and time that should be invested. In any case it is a trade-off between number of participants and money respectively time costs. [29]

Spool and Schroeder [25] showed that in a survey already five participants are necessary to discover considerable 85% of the problems which occur with the questions.

These problems have led to the development of survey tools like surveymonkey.com or vividence.com. Here the potential target group is very big. But still, users have to be recruited for the single survey.

A different approach is to use micro-task markets like Amazon's mechanical turk to complete surveys quickly and cheaply. But here you have to keep some specialties in mind. Typically micro-task markets rely on simple and short tasks, done by many people whereas user evaluation is normally done with fewer users and more complex tasks.

Another problem is the quality of the result. In chapter 3.3 there is a discussion about how to improve the quality of the results.

Downs et al. discuss a screening process to use together with a survey to identify those workers who do not answer seriously. [5]. They included two test questions that should uncover bad workers.

With this method they determined 764 of 1962 people that did not answer consistently in their survey.

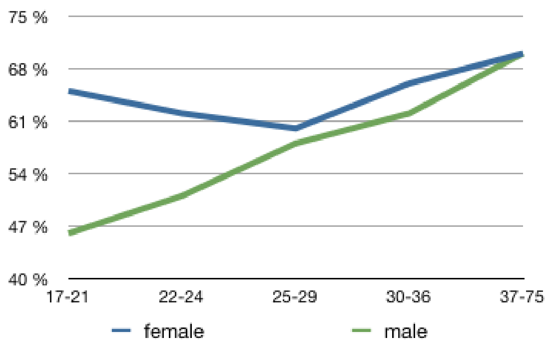


Fig. 12. Gender by Age. Who did a good job?.[5]

Figure 12 shows the dependency between age and percentage of failure in the qualification task. A big difference between young girls and young boys can be seen. Boys in young age are apparently not very reliable workers whereas girls reach worst level at the age of 25 to 29.

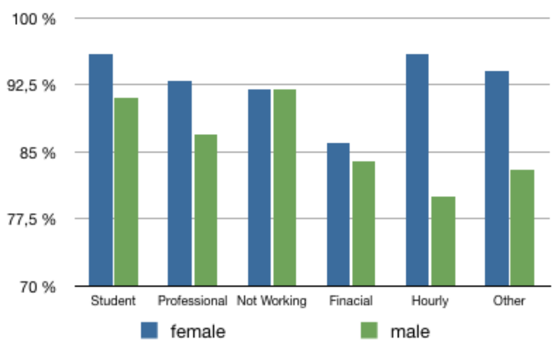


Fig. 13. Gender by Occupation. Who did a good job?.[5]

In any case, such answer schemes are highly depended on the first occupation of the workers. Students, professionals and non-workers did the tasks better than financial, hourly and other workers. But the difference between female and male has to be considered here too.

Doing a survey with Amazon's mechanical turk another problem occurs. In the demographics chapter it was stated that the average age in 2009 is 31.6. In the US people have a median age of 36.5 [3] and in India 26.2 [12]. But in Figure 5 you see that the age band of the 65+ is very low. So you can not estimate a equal image of the population in the mechanical turk population. Besides, the income range of the turkers doesn't mirror the overall population.

6 Conclusion

"Give Work" [23] and "We bring dignified, computer-based work to women, youth, and refugees living in poverty" [23] are the slogans of samasource. So what do they mean by that?

But first, why do some people have no work? This has various reasons, maybe they do not have any qualification or maybe they are just living in the wrong place. So for example many african countries have a very high unemployment rate. This is not because the people do not want to work or have no qualifications. In such countries there is simple no work for the people.

This is the chance of crowd-sourcing services to reach people that are willing to work for small wages. They only need a computer and a connection to the internet. To participate in txteagle they even only need a mobile phone. One might think, there aren't enough computers or mobile phones in emerging countries. But looking on figure 14 it is clear that even in LDCs, (least developed countries) more than 60 percent of the population has access to a mobile signal. So the potential users are numerous.

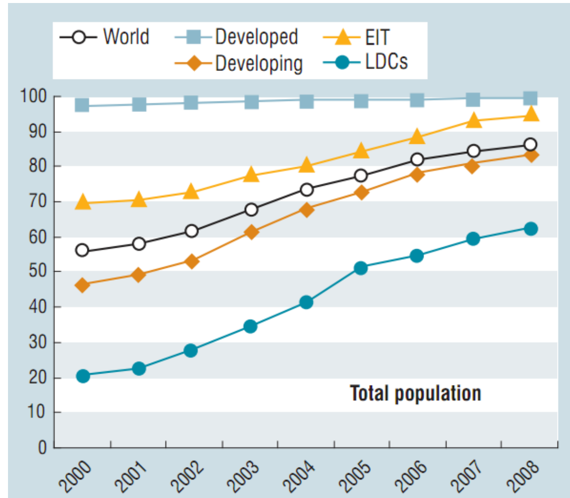


Fig. 14. Population covered by a mobile signal by country group 2000-2008 [26]

This is a reason why crowd-sourcing services are interesting in a social way. But there are also economic or scientific reasons. Since the beginning of globalization, outsourcing heavy production to emerging countries is common. Nowadays there are many parts of business primarily settled down in low wage countries. Even high tech services like software coding is often cheaper in such countries.

With crowd-sourcing even knowledge discovery might get outsourced. This has advantages and disadvantages and is a chance for both developed and emerging countries. In the Information economy report [5] it is stated that there is a strong correlation between poverty rate and mobile subscription. So if crowd-sourcing services bring work to mobile phone users, more people can afford such devices and the poverty rate will decrease.

After reading the paper it might be clear that the biggest problem with crowd-sourcing services is, how to check the quality of the results. The Find-Fix-Verify-Pattern faces this problem as well as some mechanism in Amazon's mechanical turk (see chapter 3.3).

Another problem which occurs with use of workers from various cultural backgrounds is interface design. Problems might arise from different reading direction or simply a different technical background.

Hence, there is still lot of work to do. Especially how to deal with cultural divergence should be considered and further be researched on.

References

[1] Amazon. Amazon mechanical turk, www.mturk.com.

[2] M. S. Bernstein, G. Little, R. C. Miller, B. Hartmann, M. S. Ackerman, D. R. Karger, D. Crowell, and K. Panovich. Soylent: a word processor with a crowd inside. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, UIST '10, pages 313–322, New York, NY, USA, 2010. ACM.

[3] U. S. C. Bureau. 2005-2009 american community survey 5-year estimates, http://factfinder.census.gov/servlet/ACSSAFFFacts.

[4] C. Callison-Burch. Fast, cheap, and creative: evaluating translation quality using amazon's mechanical turk. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, EMNLP '09, pages 286–295, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[5] J. S. Downs, M. B. Holbrook, S. Sheng, and L. F. Cranor. Are your participants gaming the system?: screening mechanical turk workers. In *Proceedings of the 28th international conference on Human factors in*

computing systems, CHI '10, pages 2399–2402, New York, NY, USA, 2010. ACM.

[6] N. Eagle. *txteagle: Mobile Crowdsourcing*, volume 5623, pages 447–456–456. Springer Berlin Heidelberg, 2009.

[7] A. Fikes and L. Baugher. Adieu to google answers, `http://googleblog.blogspot.com/2006/11/adieu-to-google-answers.html`, 2006.

[8] P. Heymann and H. Garcia-Molina. Turkalytics: analytics for human computation. In *Proceedings of the 20th international conference on World wide web*, WWW '11, pages 477–486, New York, NY, USA, 2011. ACM.

[9] J. J. Horton and L. B. Chilton. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM conference on Electronic commerce*, EC '10, pages 209–218, New York, NY, USA, 2010. ACM.

[10] http://uclue.com/. http://uclue.com/, 2011.

[11] http://www.galaxyzoo.org/classify. http://www.galaxyzoo.org/classify.

[12] index mundi. India median age, `http://www.indexmundi.com/india/median_age.html`.

[13] P. Ipeirotis. Mechanical turk: The demographics, `http://behind-the-enemy-lines.blogspot.com/2008/03/mechanical-turk-demographics.html`, 2008.

[14] P. Ipeirotis. Demographics of mechanical turk. *CeDER Working Papers*, 10, 2010.

[15] P. G. Ipeirotis, F. Provost, and J. Wang. Quality management on amazon mechanical turk. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 64–67, New York, NY, USA, 2010. ACM.

[16] S. Khanna, A. Ratan, J. Davis, and W. Thies. Evaluating and improving the usability of mechanical turk for low-income workers in india. In *Proceedings of the First ACM Symposium on Computing for Development*, ACM DEV '10, pages 12:1–12:10, New York, NY, USA, 2010. ACM.

[17] G. Little, L. B. Chilton, M. Goldman, and R. C. Miller. Exploring iterative and parallel human computation processes. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '10, pages 68–76, New York, NY, USA, 2010. ACM.

[18] G. Little, L. B. Chilton, M. Goldman, and R. C. Miller. Turkit: human computation algorithms on mechanical turk. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, UIST '10, pages 57–66, New York, NY, USA, 2010. ACM.

[19] G. Little, L. B. Chilton, R. Miller, and M. Goldman. Turkit: Tools for iterative tasks on mechanical turk. In *In Human Computation Workshop (HComp2009*, 2009.

[20] J. P. New York Times. Artificial intelligence, with help from the humans, `http://www.nytimes.com/2007/03/25/business/yourmoney/25Stream.html`, 2007.

[21] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 311–318, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

[22] J. Ross, L. Irani, M. S. Silberman, A. Zaldivar, and B. Tomlinson. Who are the crowdworkers?: shifting demographics in mechanical turk. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, CHI EA '10, pages 2863–2872, New York, NY, USA, 2010. ACM.

[23] samasource.org. samasource.org.

[24] R. Snow, B. O'Connor, D. Jurafsky, and A. Y. Ng. Cheap and fast—but is it good?: evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 254–263, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.

[25] J. Spool and W. Schroeder. Testing web sites: five users is nowhere near enough. In *CHI '01 extended abstracts on Human factors in computing systems*, CHI EA '01, pages 285–286, New York, NY, USA, 2001. ACM.

[26] UNCTAD. Information economy report 2010. ict, enterprises and poverty alleviation. 2010.

[27] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '04, pages 319–326, New York, NY, USA, 2004. ACM.

[28] K. Windisch and C. Mechel. *Karl Gottlieb von Windisch's Briefe "uber den Schachspieler des Hrn. von Kempelen: nebst drey Kupferstichen, die diese ber "uhmte Maschine vorstellen*. Mechel, 1783.

[29] H. Zhang, E. Horvitz, and R. C. Miller. Crowdsourcing general computation.

Design principles for participatory sensing

Thomas Bornschlegel

Abstract— Participatory sensing is an interesting concept of collective data acquisition. It uses available devices, like smartphones, to gather information. Potential participants for such networks can be recruited from the growing number of smartphone users. Thus the efforts of many individuals can be seen as the basis of such systems. This work gives an introduction to the basic concepts of participatory sensing. Existing applications are summarized in a separate section. Special interest is paid to the incentives that were chosen to engage participants to contribute to the systems. This gave inspiration for a catalog of design principles towards a user centric approach for participatory sensing. The catalog presents four main issues that should be considered. First, possibilities are shown to form an interesting concept that provides incentives for participation. Second, user's resentments against sensing capabilities are discussed. Third, the importance of adapting input mechanisms is described. And finally new ideas on recruiting participants are addressed.

Index Terms—smartphones, sensor networks, data acquisition, participatory sensing, mobiscope, design principles

1 Introduction

The growing popularity of smartphones creates new possibilities for mobile applications. One interesting area is participatory sensing: the combination of a large number of smartphones into a network, which is used for collective data acquisition. This was made possible due to distinctive properties of smartphones. First, they combine a large amount of sensors to measure e.g. the position, acceleration, or orientation of the phone. External devices that are connected to the smartphone (e.g. via bluetooth) can enhance this spectrum even more[1]. As a second point connectivity to the internet is a standard feature of smartphones. By that the collected data can be sent to a central server, where it can be evaluated. And finally central directories to distribute mobile applications (apps) are a basic part of the big smartphone platforms. In this way clients for data collection can be deployed quickly to a large amount of users.

This article presents the topic from two perspectives. First the basics are discussed in a *top-down manner*. Section 2 gives an overview on related work. Afterwards the general structure of such systems is explained in section 3. These two sections clarify terms and concepts, so that a more detailed, *bottom-up*, look can be taken at existing systems in the following. Section 4 shows applications and investigates how users are induced to participate in collective data acquisition. Six different categories are defined and are explained with existing projects. Building on this overview, design principles were extracted and explored further. They are described in detail and summed up in section 5. Up to my knowledge such a classification that centers on incentives for participation has not yet been created so far. Thus this work can give inspiration for the creation of more intrusive systems. Finally section 6 sums up the paper and gives an outlook on future systems.

2 Related work

Two articles, [7] and [2], laid the foundation for this paper and are summed up in the following paragraphs briefly. Important terms that are used in the following are also introduced here.

One of the earliest approaches that addresses collective data acquisition using mobile phones is [7]. The term used for this idea is **participatory sensing**. It is defined as a system that "tasks everyday mobile devices, such as cellular phones, to form interactive, participatory sensor networks that enable public and professional users to gather, analyze and share local knowledge" [7]. The authors see participatory sensing as a tool that can be used by professionals as well as community groups to create campaigns about various topics like traffic patterns or pollution exposure of school buses. They classify four groups of people that are involved in the sensor network: *initiators, gatherers, evaluators, and analysts*. Initiators set up the system and specify the data that should be collected. Gatherers are the ones that provide the data, e.g. by installing and using an app that sends data to a server[2]. Evaluators and analysts verify, classify, process and interpret the data, so that it can be presented to the public. As potential applications the areas urban planning, public health, cultural identity, creative expression, and natural resource management are mentioned.

The second important term is **mobiscope**. Abdelzaher2007 et al. [2] define a mobiscope as "a federation of distributed mobile sensors into a taskable sensing system that achieves highdensity sampling coverage over a wide area through mobility". This definition is not solely focused on cell phones. The authors distinguish between vehicular mobiscopes and handheld mobiscopes. Vehicular scopes are formed by equipped vehicles and are used for mapping, or for measuring road conditions. However, they can not be quickly adapted to new tasks and sensors are attached to vehicles, which limits their application area. Handheld mobiscopes describe the class of systems that this paper is focused on. Design issues that should be considered when setting up a new handheld mobiscope are discussed in detail. Important points are e.g. transferring the data efficiently, heterogeneity of devices, and data privacy. They gave inspiration for the description of the general system design in section 3.

In the following the term sensor network is used to describe a federated network of mobile phones that enables collective data acquisition.

3 General system design

Before a participatory sensor network can be put into operation it has to be made clear which data should be collected, for which reasons, and who should collect the data. After these considerations are made, the actual design of the system can be created. A good structure for issues that have to be addressed in the system design can be extracted from the steps that are run through when collecting, processing and evaluating data. Five steps are identified and are briefly summed up in

- *Thomas Bornschlegel is studying Media Informatics at the University of Munich, Germany, E-mail: thomas.bornschlegel@campus.lmu.de*

- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

[1] Just recently Google took a step in this direction by acquiring Arduino (`http://www.arduino.cc`), "an open-source electronics prototyping platform based on flexible, easy-to-use hardware and software" [3]. Devices built on Arduino can be connected to Android phones and thus enhance their functionality.

[2] This is the main difference to *opportunistic sensing* [13], which does not require the user to be aware of a running application. An example for this is the number of users logged in at a certain cell phone tower. This number can be obtained without any interaction or knowledge of the users.

Figure 1. The first two steps are carried out on the device that is used for data collection. The last two steps are implemented on the server, and the third step takes place in between. In the following each point is addressed separately and is discussed in detail.

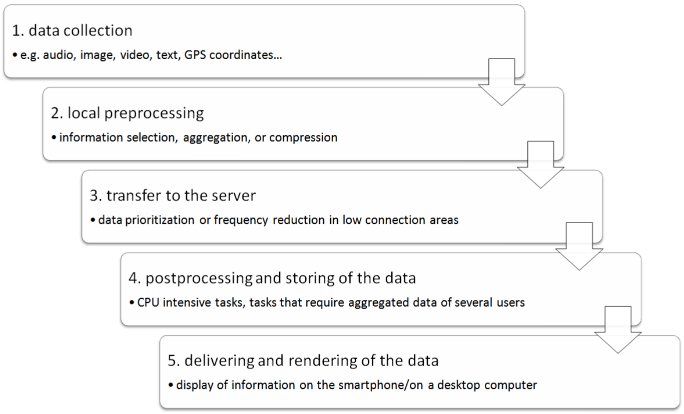


Fig. 1. General system design

Data collection or gathering is the first step. Data can be collected implicitly or explicitly by a gatherer. If a gatherer agrees to share her location, the GPS sensor automatically reads her position in determined intervals and the device sends the data implicitly to the server. Other data could require the user to interact with the device before data is collected. For example, sharing a photo of a statue in a sightseeing network needs the user to explicitly take a photo. Further one can distinguish between data collection about the user and about the users environment. The two mentioned examples illustrate this again. In the first case the user shares her location, while the second case does not describe a property of the user but her surroundings[3].

Local preprocessing describes the process of selecting, aggregating and compressing the data locally on the collecting device. This has two purposes. Firstly, the amount of data that is sent over the network is minimized, which leads to faster data transfers. Secondly the data that is being sent can be controlled. Mun et al. [21] introduce "Personal Data Vaults (PDVs)" for this purpose. This concept suggests privacy policies to the user which result in different constraints for the transferred data. In addition the granularity of data can be defined, so that only a subset of the collected data is shared e.g. continuous measurements or solely aggregated measurements for a single day. Before any data is send to the server the user receives a summary of the information to be shared. Furthermore every transaction is logged and it is made transparent who has access to the data and for which purpose.

The transfer of the data to the server is executed afterwards. The frequency and the amount can be adapted to the available network rate. In locations with low or no connectivity, the sending could be postponed until the user reaches a place with better connectivity (e.g. a wifi-network near her home). If the connectivity is not getting better, data can be prioritized, so that the most important part of the data is send when possible. Less important data can be omitted or aggregated and send at a later point in time.

Postprocessing and storing of the data happens on the server. As server processors are usually faster than mobile CPUs, elaborated computations should be performed in this step. Some processing requires the combined data of all users and can therefore only be performed on the server side. Further issues to consider here are the modeling of the database tables to use, the selection of hard- and software, and security considerations.

[3]However, when evaluating which sights are being photographed by the user we could create a profile that shows which kind of sights, e.g. ancient or modern, she prefers. In this case we are describing a property of the user.

Delivering and rendering of the aggregated data is the final step. The rendering can be carried out on the device that collected the data, or on other devices. Desktop computers with larger screens have more space to arrange informations and are therefore better suited for displaying complicated sets of data. A good example of visualizing data comprehensively can be found at [19]. The authors created a tool to visualize human mobility patterns in a city. The necessary data for the project was extracted from cell phone traces. Figure 2 shows the user interface which consists of a combination of 3D and 2D panels. It enables analysts to browse and evaluate "population density distribution over space and time" [19].

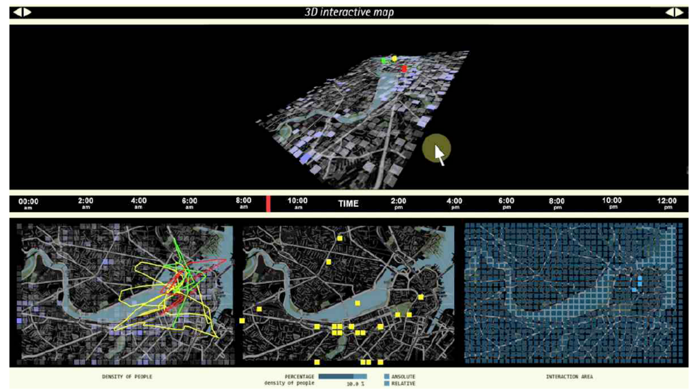


Fig. 2. Visualizing sensor data, image taken from [19, p.359]

3.1 Frameworks

EpiCollect [1] and Nokia Data Gathering [26] are examples of frameworks that implement the generic system design of section 3. The first is an open-source software which is available for Android and Iphone devices. It enables the collection of text, images and GPS coordinates. Projects for EpiCollect are set up on the respective homepage[4] and results can be visualized using Google Maps. The whole procedure does not require any programming skills, so that data gathering campaigns can also be initiated by non computer-scientists.

The second tool, Nokia's Data Gathering Tool, can also be downloaded[5] for free and is compatible with Java ME capable phones. It is used in the fields of health, agriculture, census, and child welfare by different organizations, as for example the World Wide Fund For Nature (WWF).

4 Existing Applications

Sensor networks are deployed in a variety of application areas. Of special interest for this work were the incentives that were chosen to gain participants. The following gives an overview of existing projects and how they try to attract possible users.

4.1 Information sharing

Receiving useful information from others can encourage people to share their knowledge with them. An example of this is "LiveCompare" [11], which enables collective bargain hunting. Users contribute to the database by capturing and uploading price tags to a central database. In return they get informations about the lowest price of the respective product in stores nearby. Similar services are deployed commercially, as e.g. Google Goggles[6] or ShopSavvy[7]. However, they only allow the receiving of information, while the information for the database is provided by the vendors. In contrast to that LiveCompare allows a democratic data gathering by the users.

[4]http://www.epicollect.net/
[5]http://www.nokia.com/corporate-responsibility/society/nokia-data-gathering/english
[6]http://www.google.com/mobile/goggles/
[7]http://shopsavvy.mobi/

A similar project is "Hapori" [14] that personalizes search results to the contexts of different users. This can be achieved by matching the search query, the user's profile and the position of the user with data from similar persons. The authors state that the system can "improve the relevance of local search results by up to ten times when compared to the results currently provided by commercially available search engine technology".

4.2 Data collection as a game

The term "game with a purpose" [28], describes a system "in which people, as a side effect of playing, perform tasks computers are unable to perform". On a website[8], people can participate in such games. The following two applications convey this concept to sensor networks.

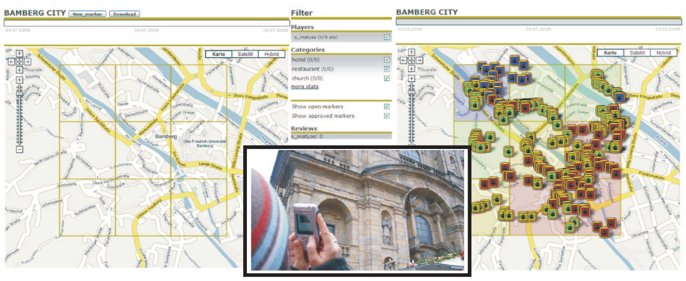


Fig. 3. Data collection and representation in CityExplorer, image taken from [20, p.245]

In the team game CityExplorer [20] players take geotagged photos and classify these photos into categories like "bar" or "church". The classification can be peer-reviewed by another, randomly chosen, player of the opposing team and is thus checked for correctness. The game ends after the predefined time has expired and teams are rated in terms of properly tagged pictures. As a result every round of the game creates new sets of semantically annotated images. Figure 3 gives an impression of the gathering process and of the representation of the collect images on a map.

Interesting in this context is the "gopher game" [9] which allows users to create missions that can be picked up by other persons. A mission description is called "gopher" and is very generic. It could e.g. require persons to describe a specific location, or to take a photo of a certain object. Gophers can be further delegated to another person, which enhances interaction between users. A similar idea can be found in [2] with the concept of "actuators". Actuators can be tasked to gather data at a defined place. This can help to assimilate unbalanced resolution of data. Spots that are not well covered can be explored directed with the help of actuators.

4.3 Motivating users with micropayments

Micro-payments are an interesting way to motivate users to take part in data collection. Reddy et al. [24] conducted a study with 55 users to improve the distribution of recycling bins around the campus. Each participant used an Android phone to collected geotagged photos of the contents of outdoor waste bins. For this purpose the app "GarbageWatch" was designed, that automated the tagging and uploading procedure. The user interface of this app can be seen in Figure 4. Users were divided into different payment groups: single payment, micropayment and competition based micro-payment. The first group was rewarded with a one time payment of 50 dollars that was not bound to a certain ratio of collected photos. The second group got a fixed reward for each image and the last group got a varying price per photo, which was higher for persons that were more active than others. The authors concluded that the best payment scheme was the second, as the first scheme did not provide incentives to collect more than necessary and the last strategy withdrew some participants that felt stressed by the competitive nature.

[8]http://www.gwap.com/

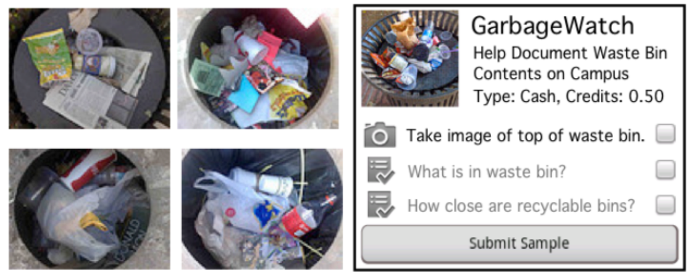


Fig. 4. User interface of GarbageWatch, image taken from [24, p. 34]

Potential money sources could be commercial companies, academical institutions, the state, or the users. To my knowledge the last named group has never been explored in an existing application. It could be useful to transfer ideas of existing browser games like FarmVille[9] to sensor networks. A sensor network could provide gaming functions that include the trading of virtual items for real money. The earnings could then be redistributed to the most active users to valuate their contribution to the network.

4.4 Altruistic reasons

Various projects highlight the benefits for society. Maisonneuve et al. [17] created "NoiseTube" which involves citizens in gathering data about noise levels in a city. They state that politicians could be convinced to take countermeasures against noisy environments, if enough data was collected by the public. Another instance of this pattern is proposed by Bayir et al. [5], who describe the framework "MobilityProfiler" that generates mobility profiles of cellphone users. The results "can be used for city wide sensing applications like air pollutant exposure estimation". Other altruistic projects were put into practice with the EpiCollect and the Nokia Data Gathering frameworks that are explained in section 3.1.

4.5 Health

Of course cell phones can not substitute the doctor. However they can lower the need for consultation with a medical practitioner. Leguay et al. [15] showed how elderly people can be supervised from the distance by generating reports of their current state of health that are send to a doctor. Sensor networks with a high participation could also help the detection of epidemics in their early stage. Madan et al. [16] try to predict the health status for persons without having actual health measurements. While Leguay et al. [15] relied on information of external devices that were connected via bluetooth (like pulse oximeters or blood pressure monitors), Madan et al. [16] evaluate implicit information. The number of calls or SMSs per day can e.g. be seen as an indicator for social interaction. Also the time and duration of calls can be taken as an indicator. Real contact with other persons is extracted from bluetooth proximity: If the bluetooth signal of another person's phone can be measured, that person has to be nearby. An ill person could thus be indicated by low social interaction and few contacts to other persons. Rachuri et al. [23] conducted a similar experiment called "Emotionsense" and try to predict emotions by evaluating voice samples. As the information for health projects tells a lot about the individual person it remains to be seen if enough people can be convinced to participate.

4.6 Socializing with others

Possibly the most successful application of sensor networks is location sharing. It has moved mainstream with new social networks like Foursquare and Gowalla which inspired other networks to jump on the bandwagon, as e.g. Facebook did with Facebook Places. The collected data is a valuable source for data mining, especially in combination with further information about the user. Location sharing in social

[9]http://www.farmville.com/

networks is a novelty because it enables the broadcasting of ones position to multiple people very quickly. There does not have to be a practical reason for this, like meeting somebody. Tang et al. [27] describe this as *social-driven location sharing*. In contrast, they term the former practice of telling only few, selected persons about ones location *purpose driven*. Social-driven location sharing can create new services like friendship prediction based on location trails [10]. However broadcasting ones location permanently can also lead to tensions between individuals, as Mancini et al. [18] showed. A solution for this problem could be the request-reply-manner proposed by Wagner et al. [29]. It allows persons to answer requests about their location with different accuracy for different people or situations. A request from a member of the family could be answered with the exact location, while a request from somebody less well known could be answered by the current city or state.

5 Design principles

Some design principles are summarized in this section to simplify the design process of future systems. Because sensor networks rely on a broad and stable user base, the starting point for all succeeding concepts is the idea to design for the user's needs. The foundation for the principles is derived from the existing applications that were shown in section 4. For a better overview, the applications are summarized in Table 1 on page 5.

5.1 Provide incentives for participation

Finding incentives for participation to contribute to the network is an important point. A classification of patterns is introduced, that builds on the categorization of existing applications in Table 1. Six classes are distinguished: *altruism, game, health, micropayments, sharing, and social*. They are defined as follows:

- *Altruism* addresses the inherent nature of people to help for the sake of a greater good.
- *Game* shows sensor networks from a playful dimension. The actual tasks are part of a game concept, that makes tedious tasks interesting.
- *Health* provides functions to sustain or improve the health status of individuals or to detect changes that could have impacts to the health of many, as e.g. epidemics.
- *Micropayments* compensate user's time and efforts with small amounts of money. A competitive element could reward the most valuable committers of information with higher payments. Thus other, less busy participants, could be encouraged to be more active.
- *Sharing* encourages users to contribute data to a system where they can access similar data of other users. The aggregated data is used to either provide access to a set of information, or to create a personalized service for the user.
- *Social* combines the collection of information with social networks. Participants can communicate their data with friends and thus enhance their social life.

Each of these concepts can be used on its own, or in combination with others. A game can e.g. provide micropayments for successful players, or highlight the altruistic nature of a task. This can give way for completely new designs that are more intriguing for the user. Parts could also be weighted according to how important they are for the whole concept. Figure 5 shows a concept which relies heavily on social and sharing elements. Others like health are not that important.

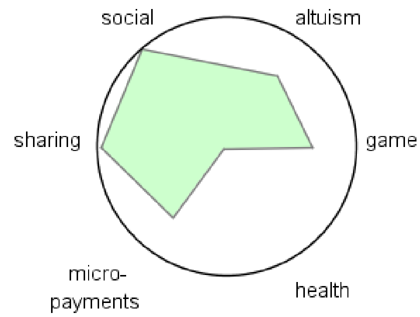


Fig. 5. Sample concept consisting of the combination of basic patterns

5.2 Meet users privacy resentments

To the developer of a sensor network, the sensing capabilities look very interesting and yield new possibilities. To the end-user, however, the possibility to upload data to a network might look disquieting or even scare them off. Participants needs should be carefully studied and addressed.

5.2.1 Location obfuscation

Location technology is a driving force for participatory sensing systems. This can be seen from the column “collected data” in Table 1. All of the presented works collect location data. This data is especially critical, as it can reveal the user's home and give hints about her daily activities. Therefore this information should be blurred with appropriate mechanisms. Brush et al. [6] describe how locations can be obfuscated according to different schemes to pay attention to privacy concerns of users. Possible options include the randomization of positions and the automatic deletion of locations near the users house. Further schemes enable the combination and mixing of data of multiple users, so that no conclusions about the individual user can be made. This way users could be convinced to participate in the network that would not without this possibility.

5.2.2 Possibilities to control and audit data

The sharing of additional data should also be defined and controlled by the user. Personal Data Vaults [21], as introduced in section 3, are a good way to address this topic. Still the mechanisms related to this concept are solely implemented on a software basis. The user might not recognize malware that corrupted the original program, or might take an imitation as the real program. Additional security could be provided by hardware mechanisms [12], or “Virtual Individual Servers” [8] that allow users to store data on machines that they own and control.

5.2.3 Transparency of data usage

The usage of the data should be clearly defined in a document to which the network provider should adhere to. A counterexample of this is the following case. In spring 2011 the GPS manufacturer TomTom sold anonymous data of his navigation system users to a trader. The data included anonymized tracks of vehicles, including the speed at track segment. Further the data was sold by the trader to the Dutch police. As a result the police put up speed controls at the locations at which speeding occurred most frequently [4]. This was of course not in the interest of TomTom and neither of its customers. Nevertheless it was made possible due to imprecise contracts for the sold data.

5.3 Adaptable input mechanisms

Quality of data is linked to the familiarity of the user with the system. Consequently the input mechanisms should be adopted to the audience. For example, data collection in developing countries can have different requirements than data collection in the western hemisphere. People might not be able to read or write which makes it necessary to use different methods. Patnaik et al. [22] conducted a study about the collection of health data with mobile phones in rural India. They conclude: “within the context of our study, the error rates for electronic forms (4.2% of entries wrong) and SMS (4.5% of entries wrong) may be too high to deploy these solutions in a critical application. In contrast, the accuracy of the voice interface was an order of magnitude better (0.45% of entries wrong), with only a single error observed across all trials.”. As voice input is harder to process, other options

work	application or framework	main incentive	collected data	year
[11]	LiveCompare	sharing	location, bar codes	2009
[14]	Hapori	sharing	location, search query	2010
[20]	CityExplorer	game	location, images	2008
[9]	Gopher	game	location, images, text	2007
[24]	GarbageWatch	micropayments	location, images	2010
[17]	NoiseTube	altruism	location, audio statistics	2009
[5]	MobilityProfiler	altruism	location trails	2009
[15]	health and activity monitoring framework	health	location, health data from external devices	2008
[16]	implicit data and Survey Launcher	health	location, phone usage statistics, bluetooth and wifi proximity	2010
[23]	Emotionsense	health	location, audio statistics	2010
[10]	Locaccino	social	location	2010
[18]	BuddyTracker	social	location	2011

Table 1. Summary and classification of existing applications

should be considered as well. Users with minor technical affiliation could be prompted with a verbose introduction wizard. And elderly persons could e.g. choose a design scheme with bigger text. Conducting studies with users can help to identify typical activities during data collection, so that the interface can be adjusted to these requirements.

5.4 Recruit participants by cultivating user interaction

For the recruitment of participants there exist two possibilities. The first is to select persons manually. Frameworks such as created by Reddy et al. [25] can support this process. They are well suited for relatively small campaigns and could also help during the initial phase for larger networks. In such a network, the first users could be hand selected. Additional users could be gained by self-energizing techniques which are introduced as the second possibility for creating a user base here. The following suggest four patterns:

- Provide options to share the system in common social networks. This enables the broadcasting of the sensor network in a viral way.
- Enable users to solicit other participants by returning rewards. Rewards could be additional income (in case of micropayments) or points, virtual awards, or extra features (for games).
- Foster user interaction to create a community-like feeling. Collected data could be retrieved, annotated and commented by other user. These functions could also help to improve the quality of the data. Peer checking of the data could be used for data validation.
- Steer the direction of the data collection, by e.g. rewarding data collection in regions that were not covered so far.

All of these suggestions can increase the number of participants, as the sensor network is promoted from user to user. To anticipate misuse of the necessary reward system, further methods have to be developed that detect and prohibit exploits. Peer checking could play an important role here, as well as manual reviewing of automatically detected suspicious actions.

6 Conclusion

This work shows how smartphones can be used for collective data acquisition. Important basic literature and existing applications were summarized in the beginning for the orientation of the reader. The presented applications were used to derive a catalog of design principles towards a user centric approach for participatory sensing. The catalog presents four main issues that should be considered. First, possibilities are shown to form an interesting concept that provides incentives for participation. These incentives are namely: altruism, game, health, micropayments, sharing, and social. The possibility to combine several items to form a new concept was highlighted. Second, user's resentments against sensing capabilities are discussed. Special interest was payed to the topics location obfuscation, user control of data, and the transparency of data usage. Third, the importance of adapting input mechanisms was described. And finally new ideas on recruiting participants are addressed. They could transmit new impulses on the creation of participatory sensor networks and shift the recruiting from a central authority to the users.

Further research could include the actual implementation of a sensor network which would be built on these principles. Particularly the combination of different incentives should be explored in depth. This is of interest because until now participatory sensing networks were developed that mainly relied on a single incentive. Additionally the concept to let users recruit users should be pursued. It could speed up the initial phase to built up a broad user base. The new system could be evaluated by asking participants to rate each of the realized design principles separately. Results could identify further topics that were not yet identified. An additional topic of interest is how participants can be encouraged to contribute to a system on a long term basis. This is especially important for projects, that are designed to collect data for several years.

References

[1] D. M. Aanensen, D. M. Huntley, E. J. Feil, F. Al-Own, and B. G. Spratt. Epicollect: Linking smartphones to web applications for epidemiology, ecology and community data collection. *PLoS ONE*, 4(9):7, 2009.

[2] T. Abdelzaher, Y. Anokwa, P. Boda, J. Burke, D. Estrin, L. Guibas, A. Kansal, S. Madden, and J. Reich. Mobiscopes for human spaces. *IEEE Pervasive Computing*, 6:20–29, 2007.

[3] Arduino Team. Arduino. `http://www.arduino.cc`, 2011. visited 10.07.2011.

[4] C. Arthur. Tomtom satnav data used to set police speed traps. `http://www.guardian.co.uk/technology/2011/apr/28/tomtom-satnav-data-police-speed-traps`, 2010. visited 10.07.2011.

[5] M. A. Bayir, N. Eagle, and M. Demirbas. Discovering spatiotemporal mobility profiles of cellphone users. In *in Proceedings of the 10th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2009*, pages 1–9, 2009.

[6] A. B. Brush, J. Krumm, and J. Scott. Exploring end user preferences for location obfuscation, location-based services, and the value of location. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 95–104, New York, NY, USA, 2010. ACM.

[7] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava. Participatory sensing. In *In: Workshop on World-Sensor-Web (WSW06): Mobile Device Centric Sensor Networks and Applications*, pages 117–134, 2006.

[8] R. Cáceres, L. Cox, H. Lim, A. Shakimov, and A. Varshavsky. Virtual individual servers as privacy-preserving proxies for mobile devices. In *Proceedings of the 1st ACM workshop on Networking, systems, and applications for mobile handhelds*, MobiHeld '09, pages 37–42, New York, NY, USA, 2009. ACM.

[9] S. Casey, B. Kirman, and D. Rowland. The gopher game: a social, mobile, locative game with user generated content and peer review. In *ACE*

'07: *Proceedings of the international conference on Advances in computer entertainment technology*, pages 9–16, New York, NY, USA, 2007. ACM.
[10] J. Cranshaw, E. Toch, J. Hong, A. Kittur, and N. Sadeh. Bridging the gap between physical location and online social networks. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 119–128, New York, NY, USA, 2010. ACM.
[11] L. Deng and L. P. Cox. Livecompare: grocery bargain hunting through participatory sensing. In *Proceedings of the 10th workshop on Mobile Computing Systems and Applications*, HotMobile '09, pages 4:1–4:6, New York, NY, USA, 2009. ACM.
[12] P. Gilbert, L. P. Cox, J. Jung, and D. Wetherall. Toward trustworthy mobile sensing. In *Proceedings of the Eleventh Workshop on Mobile Computing Systems 2010; Applications*, HotMobile '10, pages 31–36, New York, NY, USA, 2010. ACM.
[13] N. D. Lane, S. B. Eisenman, M. Musolesi, E. Miluzzo, and A. T. Campbell. Urban sensing systems: opportunistic or participatory? In *Proceedings of the 9th workshop on Mobile computing systems and applications*, HotMobile '08, pages 11–16, New York, NY, USA, 2008. ACM.
[14] N. D. Lane, D. Lymberopoulos, F. Zhao, and A. T. Campbell. Hapori: context-based local search for mobile phones using community behavioral modeling and similarity. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 109–118, New York, NY, USA, 2010. ACM.
[15] J. Leguay, B. Ravera, M. Lopez-Ramos, V. Gay, and E. Robert. Multimodal health and activity monitoring framework for elderly people at home. In *Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments*, PETRA '08, pages 84:1–84:4, New York, NY, USA, 2008. ACM.
[16] A. Madan, M. Cebrian, D. Lazer, and A. Pentland. Social sensing for epidemiological behavior change. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 291–300, New York, NY, USA, 2010. ACM.
[17] N. Maisonneuve, M. Stevens, M. E. Niessen, and L. Steels. Noisetube: Measuring and mapping noise pollution with mobile phones. In *ITEE 2009 - Information Technologies in Environmental Engineering*, pages 215–228. Springer Berlin Heidelberg, May 2009. Proceedings of the 4th International ICSC Symposium in Thessaloniki, Greece, May 28-29, 2009.
[18] C. Mancini, Y. Rogers, K. Thomas, A. Joinson, B. Price, A. Bandara, L. Jedrzejczyk, and B. Nuseibeh. In the best families: tracking and relationships. In *29th International ACM Conference on Human Factors in Computing Systems, ACM CHI 2011*, 2011.
[19] M. Martino, F. Calabrese, G. Di Lorenzo, C. Andris, L. Liang, and C. Ratti. Ocean of information: fusing aggregate & individual dynamics for metropolitan analysis. In *Proceedings of the 15th international conference on Intelligent user interfaces*, IUI '10, pages 357–360, New York, NY, USA, 2010. ACM.
[20] S. Matyas, C. Matyas, C. Schlieder, P. Kiefer, H. Mitarai, and M. Kamata. Designing location-based mobile games with a purpose: collecting geospatial data with cityexplorer. In *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology*, ACE '08, pages 244–247, New York, NY, USA, 2008. ACM.
[21] M. Mun, S. Hao, N. Mishra, K. Shilton, J. Burke, D. Estrin, M. Hansen, and R. Govindan. *Personal Data Vaults : A Locus of Control for Personal Data Streams*, page 1. ACM Press, 2010.
[22] S. Patnaik, E. Brunskill, and W. Thies. Evaluating the accuracy of data collection on mobile phones: A study of forms. In *SMS, and Voice, Proc. Intl Conf. Information and Comm. Technologies and Development (ICTD 09), ICTD, 2009*, pages 74–84, 2009.
[23] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas. Emotionsense: a mobile phones based adaptive platform for experimental social psychology research. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 281–290, New York, NY, USA, 2010. ACM.
[24] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava. Examining micropayments for participatory sensing data collections. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 33–36, New York, NY, USA, 2010. ACM.
[25] S. Reddy, D. Estrin, and M. B. Srivastava. Recruitment framework for participatory sensing data collections. In *Pervasive Computing*, LoCA '09, pages 138–155, Berlin, Heidelberg, 2010. Springer-Verlag.
[26] B. Ren, J. Ma, and C. Chen. The hybrid mobile wireless sensor networks for data gathering. In *Proceedings of the 2006 international conference on Wireless communications and mobile computing*, IWCMC '06, pages 1085–1090, New York, NY, USA, 2006. ACM.
[27] K. P. Tang, J. Lin, J. I. Hong, D. P. Siewiorek, and N. Sadeh. Rethinking location sharing: exploring the implications of social-driven vs. purpose-driven location sharing. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, Ubicomp '10, pages 85–94, New York, NY, USA, 2010. ACM.
[28] L. von Ahn and L. Dabbish. Designing games with a purpose. *Commun. ACM*, 51(8):58–67, 2008.
[29] D. Wagner, M. Lopez, A. Doria, I. Pavlyshak, V. Kostakos, I. Oakley, and T. Spiliotopoulos. Hide and seek: location sharing practices with social media. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, MobileHCI '10, pages 55–58, New York, NY, USA, 2010. ACM.

Actuator Technologies of Active Haptic Environments

Beatrice Lamche

Abstract— This article is the first of a two-part series intended to be an introduction to *active haptic environments*, their technology aspects and underlying tactile actuator systems, in contrast to the applications and scenarios of active haptic environments, which will be covered in part II of this series. *Active haptic environments* are able to facilitate our everyday life enormously. These technologies are computer systems, which are invisibly embedded in our everyday environment and are able to automatically adapt to our necessities. *Active haptic environments* contribute to overcome the problem of information overload. Due to the fact that information is mainly transmitted through one sensory channel and due to shrinking displays, humans visual senses are overtaxed more and more frequently. Five different types of actuator technologies are described, which can be used for the realization of *active haptic environments*. The findings indicate that although several actuator technologies which provide different types of feedback have already been developed, they can not be universally implemented and therefore are nowadays hardly used in the form of *active haptic environments*. This paper concludes with the identification of several research gaps.

Index Terms—Active Haptic Environments, Ubiquitous Computing, Multimodality, Actuator Technologies, Haptic Feedback, Tactile Feedback

1 Introduction

Technological advance is not always a linear outgrowth of past technology, meaning that future computing environments can be expected as comprising more processing power, more memory and better color resolution. Time has taught, that information and communication technologies do not always evolve linearly. Due to the human striving for liberty, unobtrusive technologies are gaining more and more importance. Therefore, the desire for self-effacing user interfaces, or rather interfaces you do not even notice, is quite high. Since information and communication technologies became progressively cheaper, they nowadays effortlessly and unobtrusively surround us. In the long run, computing access will be everywhere, and therefore, the personal computer and the workstation will become practically obsolete [26]. The trend will move towards information and communication technology systems, that enable information to be available everywhere. In brief *ubiquitous computing* [19].

What ubiquitous computing is and what it does. The area of *ubiquitous computing* was introduced by Mark Weiser in 1993. Mark Weiser foresaw the development towards *ubiquitous computing* and described this term in his influential article 'The Computer for the 21st Century' [23]. In this paper, Weiser describes *ubiquitous computing* as a possible next-generation computing environment. In this environment, the individuals are continually interacting with a huge amount of nearby wirelessly interconnected computers. These technologies are most effective if they manage to work invisibly in the background. Weiser's aim was to create a new kind of relationship of people to computers, one in which the computer fades into the background, allowing people to be freed from tedious routine tasks [25]. *Ubiquitous computing* helps to overcome the problem of information overload. These machines have to fit the human environment, instead of forcing humans to enter theirs, thus they can be considered as the opposite of *virtual reality* systems, which focus on simulating a world inside the computer, instead of enhancing the world that already exists [24].

Active haptic environments. Having explained the term *ubiquitous computing* leads to a better understanding of this paper's topic: *Active haptic environments*. *Ubiquitous computing* describes computing as disappearing into the physical environment [1]. Thus, *active haptic environments* can be defined as computer systems, that are invisibly embedded in our everyday environment, without forcing the user to wear special goggles, gloves or even body suits [24]. They facilitate the everyday life of the user by using haptic sensors, so that they can automatically adapt to his necessities. Therefore, *active haptic environments* can be considered as being a part of *ubiquitous computing*.

Multimodality. Early research of *active haptic environments* mainly tried to tackle the challenge of *sensory substitution*. Systems have been developed, which were able to convert imagery or speech information into electric or vibratory simulation patterns on the skin. These results of research have been a tremendous enrichment especially for visually- and also hearing-impaired users. However, *active haptic environments* also contribute to overcome the problem of information overload. Due to the fact that information is mainly transmitted through one sensory channel and due to shrinking displays, humans visual senses are overtaxed more and more frequently. Therefore, a solution has to be found to offload visually presented information. *Multimodality* represents one solution to this problem [27]. *Multimodality* allows to communicate data to users more effectively by reducing information overload placed on any one sense [28]. For this transmission of information, multiple senses are used, especially the use of sound and touch, such as through the use of vibrating alerts on mobile phones [5]. *Haptic design* can be considered as a form of *multimodal design*, using the touch sense in conjunction with other sensory modalities [14]. The combination of vision and haptics leads to additional benefits. The process of learning to recognize haptic feedback is quite short and reaction time is very fast [9]. However, functional haptic interfaces depend on the human physiology. Ensuring effective design of *active haptic environments*, a greater understanding of how best to convey haptic information is required. Therefore, for an efficient development process, physiological and psychophysical knowledge is needed, which will be discussed in chapter two.

Note: This paper deals with the technology aspects of active haptic environments, in contrast to the applications and scenarios of active haptic environments, which will be covered in another paper within the scope of the *Media Informatics Advanced Seminar on Ubiquitous Computing*.

2 Haptic System Physiology and Psychophysics

A high amount of perceptions can be activated through a haptic system. Haptic system designers have to consider which haptic stimulation suits a given task. It can be distinguished between tactile and kinesthetic stimulation. Therefore, for an effective development of *ac-*

- *Beatrice Lamche is studying Media Informatics at the University of Munich, Germany, E-mail: beatrice.lamche@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

tive haptic environments, a greater understanding of how best to convey haptic information, as well as a physiological and psychophysical knowledge, is required. Haptic interface designers can derive tactile interaction design guidelines from such physiological and psychophysical studies [9]. Haptic perception can be considered as the sum of tactile mechanoreceptors and kinesthetic receptors [4]. The location and characteristic features of all types of cutaneous receptors are shown in *Figure 1*.

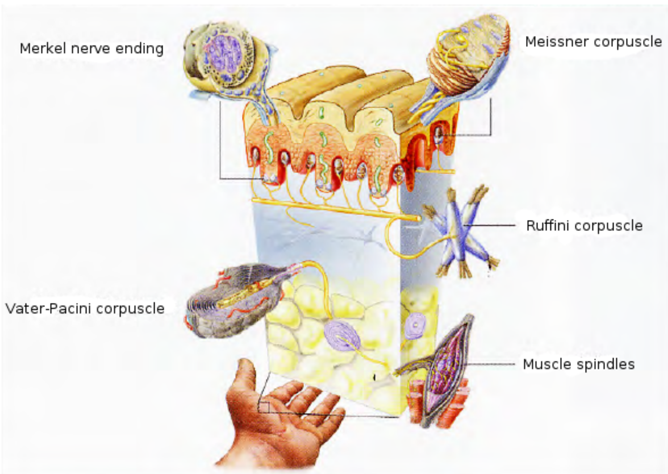


Fig. 1. Schematic drawing of different types of receptors [8].

2.1 Tactile Mechanoreceptors

For the development of efficient active haptic environments, designers have to consider which haptic sensors best support human-system interaction [9]. *Table 1* provides an overview of haptic tactile skin mechanoreceptor characteristics. These mechanoreceptors are located at a depth of several millimeters from the skin surface [20]. The total number of tactile mechanoreceptors per hand is estimated at 17,000. Merkel mechanoreceptors are found close to the surface of the skin, the others are located in the dermis [8]. Due to the fact, that tactile meachnoreceptors are adapted to the structure of the skin in fingers and palms, they mainly react to mechanical stimuli. Human skin can be hairy, on the one hand and glabrous on the other. Glabrous or hairless skin is found on the palms, fingertips and soles of the feet. However, most of the human body is covered by hairy skin. A schematic drawing of tactile mechanoreceptors is shown in *Figure 2*. This chapter also deals with a more specific type of sensory receptors: Thermoreceptors, which have to be considered for the development of thermal actuator systems.

2.1.1 Mechanoreceptors in glabrous skin

Hairless skin has four significant mechanoreceptors:

- Meissner Corpuscles
- Merkel Disks
- Pacinian Corpuscles
- Ruffini Endings

These receptors can be stimulated by physical parameters. Their sensitivity can depend on their density, size, nerve fiber branching and frequency range. Mechanoreceptors can be activated by skin motion, on the one hand, and sustained pressure, on the other. These two stimulation types affect the degree of activation. The mechanoreceptors in glabrous skin should be stimulated, if effective transmission of detailed tactile information is intended. Therefore, for this purpose, haptic interface designers should try to activate the skin on the palms, fingertips, or soles of the feet [9]. Meissner corpuscles and Merkel disks are directly located in the finger tips. Pacinian corpuscles and Ruffini endings are mainly found on the inner palm of the hand and on the finger limbs [4].

Table 1. Haptic Tactile Skin Mechanoreceptor Characteristics [9].

Haptic Features	**Pacinian Corpuscles**	**Ruffini Endings**	**Meissner Corpuscles**	**Merkel Disks**	**Hair Follicles**
Skin Type	Glabrous and hairy	Glabrous and hairy	Glabrous	Glabrous	Hairy
Stimulation objective	Vibration, acceleration, roughness	Skin stretch, lateral force, motion direction, static force	Velocity, flutter, slip, grip, control	Skin curvature, pressure, form, texture, edges	Touch
Spatial Resolution	2 cm	1 cm	3 -5 mm	0.5 mm	
Stimulation Frequency Range (Hz)	100 - 1,000	0.4 -100	2 - 40	0.4 -10	

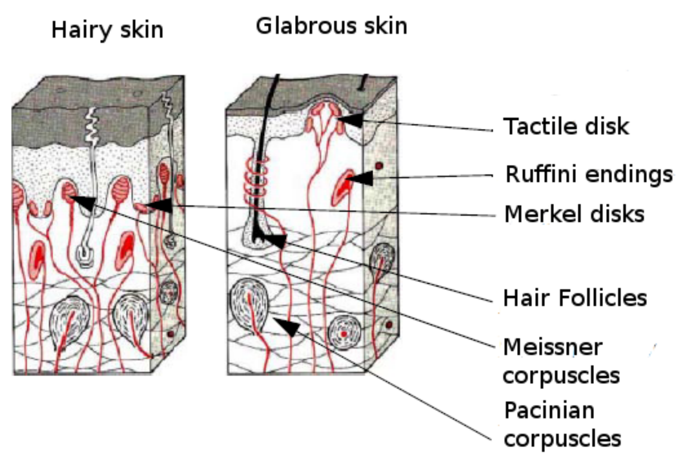


Fig. 2. Schematic drawing of tactile mechanoreceptors [4].

2.1.2 Mechanoreceptors in hairy skin

Three distinct mechanoreceptors exist in hairy skin:

- Pacinian Corpuscles
- Ruffini Endings
- Hair Follicles

Table 1 shows, that the spatial resolution of these receptors is quite low and therefore, they do not effectively perceive the specific geometric structure or form of an object. Hairy skin is most effective for the presentation of spatial tactile cues at various skin locations and detects in particular vibration and static force. If haptic interface designers intend to transmit vibratory information, the tactile actuators can be placed anywhere on the body [9].

2.1.3 Thermoreceptors

It is important to understand the properties of thermal receptors to design efficient thermal actuator technologies. Even small changes in temperature are sensed by the skin. There are different factors, which influence this ability. One of them is the amplitude and rate of temperature change. Other factors are the baseline temperature of the skin and also the stimulated site on the body [13]. The skin has two different kinds of receptors, which provide the ability to perceive variations in temperature. One of them is called *cold receptor*, the other one *warm receptor*. The skin has more cold receptors than warm receptors by a ratio of up to 30:1. Cold receptors react to decreases in temperature in the range of 5 - 43 ° C. The discharge of warm receptors is caused by increases of skin temperature up to 45 ° C [12]. Thermal stimuli need to be carefully designed due to the existence of the thermally sensitive *nociceptors*, which respond to innocuous thermal stimuli. Identification of objects is assisted by the thermal signals deriving from changes in skin temperature which are caused when the object is held in hand. The receptors pass on the signals, which are received by the integrative centers in the spinal cord and the brain, thus supporting identification of objects [13].

2.1.4 Combination of different tactile mechanoreceptors

Although distinct mechanoreceptors with different impacts exist, *active haptic environment* designers have to consider, that all skin sensors are simultaneously stimulated and therefore, tactile perception results from a combination of inputs of all receptors in a given skin area [9].

2.2 Kinesthetic receptors

Besides tactile skin mechanoreceptors, another form of stimulation exists: Kinesthetic receptors. *Table 2* provides an overview of haptic kinesthetic receptor characteristics. Kinesthetic senses make humans aware whether the movement is voluntarily or externally imposed and also indicate how fast and in which direction their limbs are moving and which static position they have. Four significant receptors are associated with kinesthetic sense: Golgi endings, Ruffini endings, Muscle spindles and Golgi tendon organs [9]. Kinesthetic sensation is acquired by the proprioceptors that are located in the muscle and joint [20]. The feedback loop for kinesthetic receptors is slow (between 0.5 and 1.7 Hz). Nevertheless, humans can execute complex movements very fast. This occurs due to the fact, that the motor and cognitive haptic systems (responsible for the direction of active movement) create a memory motor trace of a specific movement [9].

2.2.1 Receptors located in joints

Two significant receptors are located in joints:

- Golgi Endings
- Ruffini Endings

Joint receptors mainly react to vibration stimuli (optimum vibration amplitude should be below 0.1 µm, optimum frequency range about 200 Hz) as well as to stretching of the collagen fibres [8]. Receptors located in joints have an important protective function for the human body. Ruffini endings can be activated either by static position or dynamic motion.

2.2.2 Receptors located in muscles

Two significant receptors are located in muscles:

- Muscle Spindles
- Golgi Tendon Organs

Information about supported limb weight is provided by Muscle spindles. More precisely, these receptors detect the subjective weight of an object in a persons hand. Moreover, they are also responsible for conscious body movement [9]. Golgi tendon organs are receptors mainly located in the juncture between skeletal muscle and tendon. Golgi tendon organs provide the awareness of active positioning and static limb position [8].

Table 2. Haptic Kinesthetic Receptor Characteristics [9].

Haptic Features	**Golgi Endings**	**Ruffini Endings**	**Golgi Tendon Organs**	**Muscle Spindles**
Location	Joint ligaments	Joint capsules	Tendons	Muscles
Stimulation Objective	Joint movement at end range of motion Extreme flexion/extension	Joint movement, particularly at end range of motion Static and dynamic	Active position sense Link to limb position Force	Active movement of muscles Conscious experience of body movement and position Weight supported by limb
Stimulation Type	Joint tension at extreme positions	Capsule stretch	Muscle tension and force	Muscle stretch/rate of change Vibration

2.2.3 Combination of different kinesthetic receptors

The combination of sensory information from the kinesthetic receptors with information from the motor and cognitive systems, provides awareness of limb position and movement. The joint moving, the movements velocity, and the contractile state of the muscles controlling the joint, are responsible for a persons ability to perceive limb movement [9].

3 Actuator Technologies

Different actuator technologies can be used for the realization of *active haptic environments*. Before turning into details of different actuator technologies, the term *active haptic environments* has to be specified more exactly. Again, it should be mentioned, that this paper excludes *virtual reality* technologies. This is due to the fact, that this paper focuses on *ubiquitous computing* and instead of simulating a world inside the computer, *active haptic environments* have the opposite aim: Fitting the real environment [24]. Moreover, this paper also does not include dedicated systems. These are systems that have to be strapped on as belts, gloves or helmets and can not be embedded into the human environment. Therefore, this paper also does not refer to so-called *PHANTOM-based* systems. The *PHANTOM* haptic interface is a hand-held-device which measures the finger tip position of the user. Therefore *PHANTOM-based* systems can not be inconspicuously integrated into the environment [16]. A huge amount of haptic interfaces has been developed for the purpose of *sensory substitution*. These are systems which are able to convert imagery or speech information into electric or vibratory simulation patterns on the skin. Therefore they serve as a tremendous enrichment especially for visually- and also hearing-impaired users [27]. Nevertheless, the term *active haptic environments* is more general, and *sensory substitution* is only a subarea of this topic.

3.1 Classification of Active Haptic Environments

Having specified the topic more exactly, *active haptic environment* systems can now be classified. *Figure 3* shows this classification. First of all, as already shown above, *active haptic environment* systems can be distinguished between technologies with kinesthetic feedback and technologies with haptic feedback. *Active haptic environment* system designers have to consider which haptic stimulation suits a given task. As shown in *Figure 3*, haptic feedback systems can again be classified in three different types of feedback:

- Thermal Actuator
- Electro- or Vibrotactile Stimulation

- Free Space Feedback

First of all, there exist *thermal actuators*. These technologies use different temperatures for giving the user a certain feedback. Moreover, a number of technologies have been developed which are based on *electro- or vibrotactile stimulation*. Newer active haptic environment technologies enable feedback in *free space*. These systems provide feedback by radiating airborne ultrasound or through air-jets.
Concerning kinesthetic feedback systems, it can be distinguished between two different types of feedback:

- Force Feedback
- Motion Feedback

Motion feedback systems provide proprioceptive motion feedback. Furthermore, kinesthetic feedback can be based on *force*. In this case, the feedback is provided by movement resistance. Examples of different actuator technologies, each of them using different types of feedback, will be presented below.

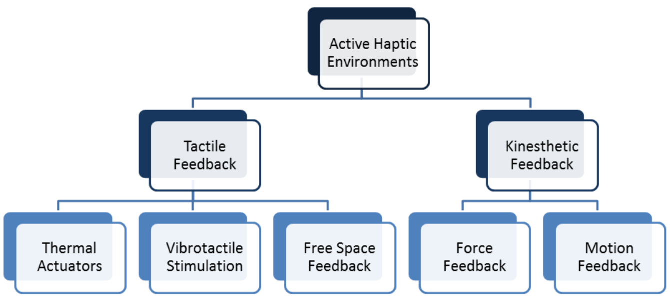


Fig. 3. Classification of *active haptic environment* technologies.

3.2 Actuator Technologies and Application Examples

The following chapter gives an overview of different actuator technologies, each of them giving different types of feedback (e.g. vibrations, force, temperatures). Moreover, certain technical actuator characteristics (e.g. frequencies, amplitude ranges) will be mentioned. A specific prototypical application scenario will be given for each actuator technology as well. It will be distinguished between actuator technologies providing *kinesthetic feedback* and actuator technologies providing *tactile feedback*.

3.2.1 Actuator Technologies Providing Tactile Feedback

This section discusses three different actuator technologies, all of them using different types of feedback.

- Electro- and vibrotactile stimulation
- Thermal actuators
- Systems providing free space feedback

Electro- and vibrotactile stimulation. For the attempt to optimize a tactile actuator design several characteristics have to be taken into account. For the signal itself, it is important to consider its strength/amplitude, the amplitude range, power consumption and the frequency range. For tactile actuator technologies the classes *electrical tactile systems* and *vibro-mechanical tactile systems* exist. In the following section the two classes of actuator technologies are described in greater detail [17].
Electrotactile stimulation systems use currents which are passed through the skin via a source. By these currents, the afferent nerves are excited directly instead of the tactile receptors. The supply of current is provided by electrodes of different types, or by insertion of fine wires into the skin [21]. The neighboring afferent nerve fibers are excited by the generated electric field. These nerve fibers are responsible for normal mechanical touch sensations. Depending on the electrode and waveform properties, the elicited sensation can be experienced as a tingle, pressure, vibration, or pain. Typically the system produces voltage-based pulses and allows control of several signal characteristics including voltage amplitude, frequency, duty cycle, and polarity.
Electrical tactors are not in widespread use in the scientific and operational community. Presumably this is caused by the more common use of vibro-mechanical systems [17]. *Figure 4* shows examples for electrical tactile systems.

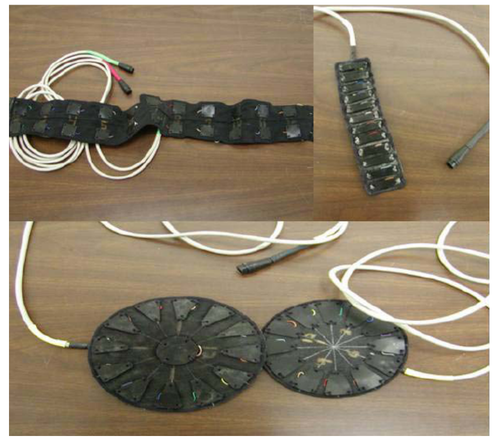


Fig. 4. Examples of electrical tactile systems [17].

Vibro-mechanical tactile systems contain objects which vibrate against the skin or other body surface. To maximize *Pacinian corpuscles* receptor sensitivity typical frequencies are selected. Due to the high sensitivity of make-and-break contact, vibrations can be conveyed effectively through an air gap [21]. The electronic components which are needed are vibrotactile transducers, amplifiers and signal generation circuitry (typically some form of computer with associated digital to analog converter) [22]. The *rotary motion tactor* is an omnipresent vibro-mechanical actuator. It is also found inside several mobile phones. A housing which incorporates a motor with an eccentric mass is the base for this actuator. Stimulation to the skin is effected by the rotation of the motor which makes the housing vibrate. It is not necessary that the housing has direct contact to the skin; clothing between the skin and the tactor is no obstacle, although the intensity and frequency of the specific motor and the type of housing that are used are of influence. Typically the tactile frequency stimulus is in the area of 70 - 150Hz [17].
An example for vibro-mechanical actuators are so-called *floor surfaces*, which are a newly developed application. Several computer-supported activities which involve movement on foot in everyday environments are enhanced or enabled by haptic communication via floor surfaces. A linear voice coil motor is used by these floor surfaces. [22]. Linear motors consist of miniature, coil based actuators which are optimized for use against the skin. It is possible that linear actuator tactors incorporate a moving *contactor* which is lightly preloaded against the skin or which can be embedded in a closed housing. In case of application of an electrical signal, the *contactor* or housing oscillates perpendicular to the skin [17]. *Figure 5* demonstrates a floor surface system which is based on linear actuators.

Thermal actuators. Thermal actuator technologies are part of haptic interfaces which, either in real or in simulated environments, deliver information about the properties of encountered objects [13]. As force and tactile feedback still have a long way to go, it is advisable for active haptic environment designers to consider the development of tactile feedback models which can simulate local geometry and texture reproduction as well as heat flux and temperature. These systems are especially advantageous if short reaction times are desired [7]. Thermal actuator technology, in its basic components, consists of thermal stimulators, temperature sensors and a controller of temperature [13].

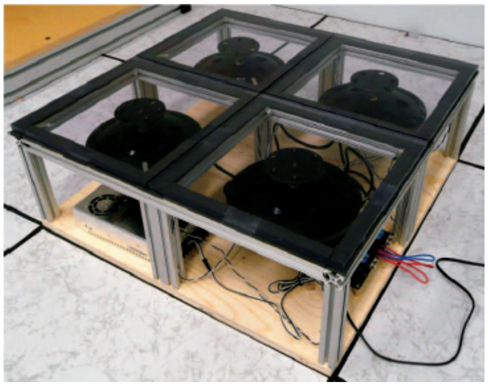


Fig. 5. Floor surfaces based on vibro-mechanical actuators [22].

Thermal characteristics of an object, as well as the thermal sensations associated with contact are reproduced by these actuators in a display [7]. For any application, the particular components chosen are depending on the projected use of the display and the implemented thermal model. A thermal actuator technology should be able to provide realistic thermal sensations. This can be achieved by the integration of appropriate hardware components and the application of a thermal model which delivers appropriate predictions of the thermal responses of the user [13].

The skin can experience thermal stimulation by using radiation (infrared and microwave), convection (air and liquid) or conduction (thermo-electric heat pumps) and by some combination of these [12]. The most widely used thermal stimulators in thermal displays have been Peltier devices, also known as thermoelectric modules. The base for these devices is the Peltier effect. This effect refers to the creation of a difference in temperature at the junctions of two dissimilar conductors in contact when a DC current flows through the circuit [13]. See *Figure 6* for a schematic diagram of a thermal feedback display.

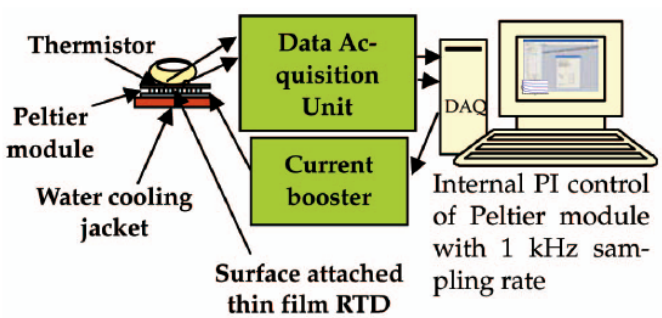


Fig. 6. Schematic diagram of a thermal feedback display [13].

Another thermal feedback device was developed by Zerkus et al. [29]. This comprises eight thermodes and a control interface. To measure skin temperature a thin film RTD (Resistance Temperature Detector) is used. This temperature is then compared to the programmed temperature. The measured difference is transferred to a software controller, which controls the current amplifiers that drive the thermoelectric heat pump [12].

Typically a closed-loop circuit with a proportionalintegral (PI) *(see figure 6)* or a Proportional-Integral-Derivative (PID) control loop is used to control the cooling or heating of thermal displays. For temperature feedback to the controllers, contact temperature sensors such as thermocouples, RTDs, or thermistors are commonly used. A contact temperature sensor has to be in thermal equilibrium with the measured object to give accurate measurements. Therefore sensors with small dimensions and low thermal mass are preferable for thermal displays to enable high-frequency temperature control [13].

The thermal kit developed by Dionisio et. al [7] is an application scenario for thermal actuator technologies. It can be used to explore how thermal cues could be used effectively in environments. This thermal kit also applies a Peltier element and gives feedback to the user via different temperatures. Cold is generated by ventilators and warmth by infrared lamps in this application [7].

Systems providing free space feedback For tactile feedback in free space two conventional strategies are used. One of them is to attach tactile displays on the users hands. It has to be noted that this strategy inherently degrades tactile feelings due to the contact between the skin and the device, which occurs even when it is not needed to provide tactile sensation. The second strategy is to control the position of tactile displays. Here they get into contact with the skin only when tactile feedback is necessary.
Iwamato et. al [11] have proposed a method to produce tactile sensation with ultrasound to facilitate handling of 3D graphic objects with the hands of the users. In previous prototypes, water as medium was used for sound propagation. The use of airborne ultrasound makes it possible that the method is applied to tactile feedback to bare hands in free space with high spatial and temporal resolution. Wave field synthesis [11] is used to control the spatial distribution of pressure. The airborne ultrasound tactile feedback system, which has been developed by Iwamato et. al provides tactile feedback in 3D free space. Airborne ultrasound is radiated by the display. It produces high-fidelity pressure fields onto the users hands and no gloves or mechanical attachments are needed. Acoustic radiation pressure, a nonlinear phenomenon of ultrasound, is the base for the method. A pressure field is exerted on the surface of an object, when the propagation of ultrasound is interrupted by the object. This effected pressure is named acoustic radiation pressure [10]. Elements of this free space feedback system are an annular array of airborne ultrasound transducers, a 12 channels amplifier circuit and a PC. The annular array of airborne ultrasound transducers is shown in *Figure 7*. The arrangement of the airborne ultrasound transducers is in a hexagonal structure. Each transducer has a diameter of 10 mm. The resonant frequency of the transducer is 40 kHz. The transducer emits a sound pressure of 20 Pa at 300 mm from the radiation surface. These six transducers, which are electrically connected, are driven at the same time [11].

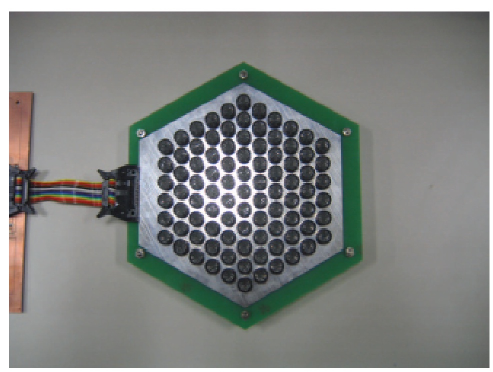


Fig. 7. Airborne ultrasound transducers [11].

Another possible candidate to deliver free space feedback is air-jet [10]. *Apple* for example filed a patent which describes a very small keyboard with a pneumatic air-jet system packed underneath [18]. The small jets of air would outline the shape of each key. Therefore, feedback before actual contact with the key can be provided. The functionality of *Apple's* keyboard using air-jets is shown in *Figure 8*. This approach is useful for really tiny keyboards where the individual keys are too close together and where the user is unable to distinguish just by the touch of his finger between the individual keys [30].

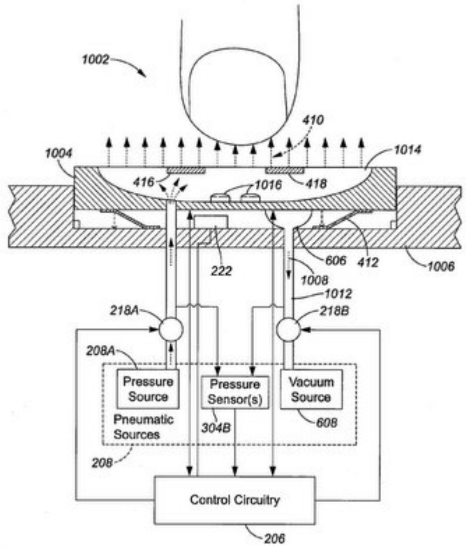


Fig. 8. Illustration of *Apple's* new keyboard approach [30].

3.2.2 Actuator Technologies Using Kinesthetic Feedback

This section discusses two actuator technologies, both of them providing different types of kinesthetic feedback:

- Force Feedback
- Motion Feedback

Technologies providing force feedback. The idea of force feedback technologies is to access primarily the kinesthetic haptic channel. However, when friction phenomena, vibration, or contact transients are inevitably generated, they interact with the cutaneous tactile sense [21]. By this, classification of these technologies is problematic. Therefore, it is difficult to classify these technologies and they can be considered as being based on kinesthetic feedback as well as being based on tactile feedback stimulation.

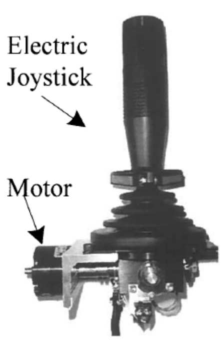


Fig. 9. *A* force reflecting joystick using a rotary actuator [2].

A common field of application where force feedback is provided is the *joystick*. Direct-drive joysticks are used for many commercial applications. These include industrial assembly, underwater, ground, and space vehicular control, medical applications, virtual reality interfaces, as well as the entertainment industry. The advantage of adding an extra-sensory perception or force input to the operator is, that the time and expense required to achieve the desired results can be reduced and final results can be improved [6]. These technologies are often based on a master/slave system. A master/slave system is a model where one component (master) has the control over the others (slaves). In this example, the electrical joystick is the master and the slave system is composed of a hydraulically actuator cylinder with a linear position sensor. In order that the joystick can reflect force feedback, a rotary actuator is used, as shown in *Figure 9*. The information of the pressure of the hydraulic cylinder is measured, so that a force feedback signal can be generated [2]. An assembly drawing of a suitable actuator (a spherical motor) for force reflecting joysticks is shown in *Figure 10*. A permanent magnet (shown in green) with 2 degrees-of-freedom drives the spherical motor. The permanent magnets rotate within a nominally spherical stator. The spherical stator (colored in brown) is provided with curved laminations. The actuator is magnetically coupled. Therefore, the electromagnetic torque can act on the handle gripped by the user [6]. For the purpose of providing a clear picture, the copper drive coils are not shown in the figure.

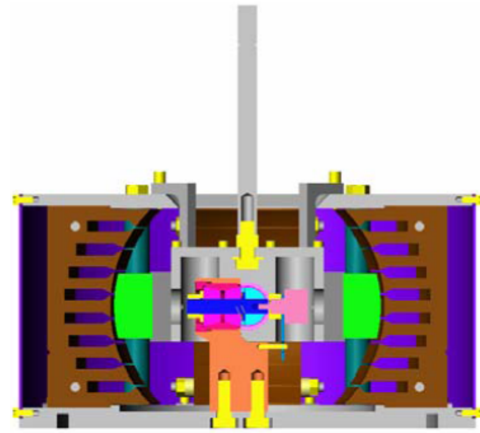


Fig. 10. *A*n assembly drawing of a spherical motor [6].

Technologies providing motion feedback. Motion feedback actuators are primarily used for the realization of prosthesis. The virtual hand prosthesis, which has been developed by Blank et. al [3], is an application example for a system which provides motion feedback. This prosthesis enables the user to control the motion of a virtual finger by the use of force input. Proprioceptive feedback is provided by allowing the index finger to move so it matches the movement of the virtual finger. In contrast, proprioceptive feedback is removed by holding the finger still [3]. Force sensors are applied by most current prosthesis. Researchers experiment with various methods of delivering sensory feedback to the user to decrease the need for visual control. Primarily this is aimed at through the use of small actuators placed in contact with the users skin. *Figure 11* shows a custom finger apparatus. It measures the force applied by the users fingertip and thus determines the movement of the virtual finger. The elements of this device are an aluminum base structure, a 128:1 geared Maxon A-max 22 DC motor with attached encoder, a clear acrylic plate affixed to the motor shaft, and an ATI Nano17 force sensor [15].

4 Conclusion

This paper provided an overview of tactile actuator technologies that are used for the realization of *active haptic environments*. The desire for *active haptic environments*, computer systems, that are invisibly embedded in our everyday environment, is getting stronger. Due to the fact that they can automatically adapt to our necessities by using haptic sensors, they are able to facilitate our everyday life enormously. Five different types of actuator technologies have been described, which can be used for the realization of *active haptic environments*. Three actuators providing tactile feedback: *Thermal actuators*, *vibrotactile stimulation* and *free space feedback*, and two actuator technologies providing kinesthetic feedback: *Force feedback* and *motion feedback*.

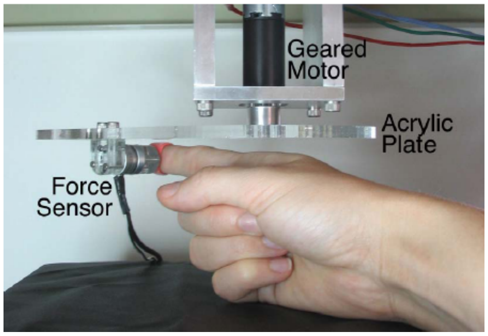


Fig. 11. A custom finger apparatus [15].

An important fact that emerges from this overview is that several actuator technologies providing different types of feedback have already been developed, but they are nowadays hardly used in our everyday life. A reason for that might be, that most of the actuator technologies are not commercially available, so they can not be universally implemented and mainly have to be build individually. Moreover, research has shown, that *active haptic environments* are mainly used for the purpose of *sensory substitution* in contrast to overcome the current problem of information overload. Nevertheless, due to the omnipresence of information and communication technologies, there exist many possibilities where *active haptic environments* could be implemented. In the long run, researchers will probably focus on the development of actuator-technologies which will be smaller, multimodal, more robust, less expensive and commercially available so that they can be implemented more frequently. Moreover, some more haptic feedback systems which provide other types of feedback could be imagined, e.g. the use of liquids, odors and different temperatures. Further development of more haptic systems which act contactless is also quite probable. These and other findings are motivating future research on *active haptic environments*.

REFERENCES

[1] G. Abowd and E. Mynatt. Charting past, present, and future research in ubiquitous computing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7(1):29–58, 2000.

[2] K. Ahn. Development of force reflecting joystick for hydraulic excavator. *JSME International Journal Series C*, 47(3):858–863, 2004.

[3] A. Blank, A. Okamura, and K. Kuchenbecker. Effects of proprioceptive motion feedback on sighted and non-sighted control of a virtual hand prosthesis. In *Haptic interfaces for virtual environment and teleoperator systems, 2008. haptics 2008. symposium on*, pages 141–142. IEEE, 2008.

[4] J. Blaurock. *Haptische Wahrnehmung rechnerinterner Freiformflaechen*. PhD thesis, Universitaetsbibliothek, 2004.

[5] S. Brewster. The impact of haptic touchingtechnology on cultural applications. *Digital Applications for Cultural Heritage Institutions*, pages 273–284, 2005.

[6] C. Corcoran, D. Cope, A. Wright, and R. Thomas. Direct-drive force-feedback joystick. *Engineering Matters, Inc. www. engineeringmatters. com. Vai*.

[7] J. Dionisio. Virtual hell: A trip through the flames. *Computer Graphics and Applications, IEEE*, 17(3):11–14, 1997.

[8] M. Gruenwald. *Human haptic perception: basics and applications*. Birkhauser, 2008.

[9] K. Hale and K. Stanney. Deriving haptic design guidelines from human physiological, psychophysical, and neurological foundations. *Computer Graphics and Applications, IEEE*, 24(2):33–39, 2004.

[10] T. Iwamoto, M. Tatezono, T. Hoshi, and H. Shinoda. Airborne ultrasound tactile display. In *ACM SIGGRAPH 2008 new tech demos*, pages 1–1. ACM, 2008.

[11] T. Iwamoto, M. Tatezono, and H. Shinoda. Non-contact method for producing tactile sensation using airborne ultrasound. *Haptics: Perception, Devices and Scenarios*, pages 504–513, 2008.

[12] L. Jones and M. Berris. The psychophysics of temperature perception and thermal-interface design. In *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002. HAPTICS 2002. Proceedings. 10th Symposium on*, pages 137–142. IEEE, 2002.

[13] L. Jones and H. Ho. Warm or cool, large or small? the challenge of thermal displays. *IEEE Transactions on Haptics*, pages 53–70, 2008.

[14] A. Karon Maclean et al. Do-it-yourself haptics, part ii: Interaction design. 2008.

[15] K. Kuchenbecker, N. Gurari, and A. Okamura. Effects of visual and proprioceptive motion feedback on human control of targeted movement. In *Rehabilitation Robotics, 2007. ICORR 2007. IEEE 10th International Conference on*, pages 513–524. IEEE, 2007.

[16] T. Massie and J. Salisbury. The phantom haptic interface: A device for probing virtual objects. In *Proceedings of the ASME winter annual meeting, symposium on haptic interfaces for virtual environment and teleoperator systems*, volume 55, pages 295–300, 1994.

[17] B. McGrath, A. McKinley, M. Duistermaat, O. Carlander, C. Brill, G. Zets, and J. van Erp. –tactile actuator technology. *Tactile Displays for Orientation, Navigation and Communication in Air, Sea and Land Environments*.

[18] A. Pance, M. Sinclair, and B. Bilbrey. Input devices and methods of operation, May 12 2011. US Patent 20,110,107,958.

[19] S. Poslad. *Ubiquitous computing: smart devices, environments and interactions*. John Wiley & Sons Inc, 2009.

[20] K. Sato, H. Kajimoto, N. Kawakami, and S. Tachi. Improvement of shape distinction by kinesthetic-tactile integration. In *EuroHaptics Conference, 2007 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. World Haptics 2007. Second Joint*, pages 391–396. IEEE, 2007.

[21] Y. Visell. Tactile sensory substitution: Models for enaction in hci. *Interacting with Computers*, 21(1-2):38–53, 2009.

[22] Y. Visell, A. Law, and J. Cooperstock. Touch is everywhere: Floor surfaces as ambient haptic interfaces. *IEEE Transactions on Haptics*, pages 148–159, 2009.

[23] W. Weber, J. Rabaey, and E. Aarts. *Ambient intelligence*. Springer Verlag, 2005.

[24] M. Weiser. The computer for the 21st century. *Scientific American*, 265(3):94–104, 1991.

[25] M. Weiser. Some computer science issues in ubiquitous computing. *Communications of the ACM*, 36(7):75–84, 1993.

[26] M. Weiser. Ubiquitous computing. *Computer*, 26(10):71–72, 1993.

[27] A. Wright. The touchy subject of haptics. *Communications of the ACM*, 54(1):20–22, 2011.

[28] N. Yannier, C. Basdogan, S. Tasiran, and O. Sen. Using haptics to convey cause-and-effect relations in climate visualization. *IEEE Transactions on Haptics*, pages 130–141, 2008.

[29] M. Zerkus, B. Becker, J. Ward, and L. Halvorsen. Temperature sensing in virtual reality and telerobotics. *Virtual Reality Systems*, 1(2):88–90, 1993.

[30] C. Zibreg. Apple dreaming up a tiny keyboard with pneumatic system, air-expelling keys for tactile sensation. `http://9to5mac.com`, 2011. visited 18.06.2011.

Scenarios of Active Haptic Environments

Annika Tonch

Abstract— Technology has meanwhile overtaxed our visual senses, especially since the size of computer screens has become smaller. Haptic interfaces can be a good opportunity to offload the visual channel and transmit data to the human brain trough tactile or kinesthetic feedback like vibrations. The advantages of haptics are the fast reaction time on haptic input and that the haptic language is easy to learn. In the terms of Ubiquitous Computing Active Haptic Environments can be defined as integrating haptic feedback into everyday objects so that it fits seamlessly into the peoples environment and disappears into the background. This paper outlines several scenarios of Active Haptic Environments. These scenarios are classified into the user groups "people with disabilities", "people in special environments", "people who need training" and "people who want entertainment" The user groups can be further divided into the semantics of haptics "warning", "information" and "entertainment". There are two systems that are used in several scenarios: The haptic chair and the haptic floor. Both systems provide the basis of different scenarios of Active Haptic Environments with different user groups and different semantics.

Index Terms—Ubiquitous Computing, Haptic Feedback, Active Haptic Environments, Haptic stimulation

1 Introduction

According to Wright [28] integrating haptics into technology is today common in buzzing cell phones or force-feedback joysticks in computer games. Nevertheless, it was not yet successful to integrate haptic technology into the mass consumer market. But haptic could be a good solution to one of the defining challenges of our age: information overload.
Technology has meanwhile overtaxed our visual senses, especially since the size of computer screens has become smaller. Haptic interfaces could be a good opportunity to offload the visual channel and transmit data to the human brain trough vibrations or other tactile or kinesthetic feedback. In Addition, haptic has two important advantages. First, people can react on haptic input very fast, because it is a background sense and in the real world a lot of things happen on the periphery in an intuitive way. Second, the haptic language is easy to learn. The reason for that is that people already learned to rely on haptic feedback in their everyday lives, for example when they judge someone based on the firmness of his handshake or when they enjoying patting a dog [28].

Even so the integration of haptic feedback into technological devices is not a common procedure, according to Gallace et al. [8], researches have been made on this topic for about 50 years. The first step to use the body surface as a communication system was made in 1960 as a group of investigators discussed about the attempt to communicate through the skin. Gallace et al. describe this event as the "symbolic birth of the first extensive research in this field".
Gallace et al. point out that the interest in this subject started this early because of the widespread belief that presenting information through the body surface can be of advantage when other senses cannot be used or are overloaded. First researches dealt with electrocutaneous forms of bodily communication like electrodes that were placed directly on the skin surface. Some earlier studies also focused on vibrotactile stimulation where tactors can be placed on the normal clothing of the user, what is called "wearable computing". This is today a lot more common than the electrocutaneous communication.
During the 1970s the main goal of the researches was to assist individuals with severe visual and/or hearing impairments through sensory substitution. At this time a first tactile vision substitution systems (TVSS) was developed. Over the years researchers have investigated the use of haptic feedback presented to the torso, had, hands, wrists, buttocks, and feet.
Since the 1990s and the substantial progress in technology, new application fields were discovered like teleoperation and telepresence, 3D surface generation and games [4]. Applications for robotic and flight simulation have immerged as well. Like mentioned before researches of today deal with the problem of information overload especially because of screen-based computers [28]. In addition, haptic feedback is today applied in Ubiquitous Computing.

Mark Weiser can be named "the founder of Ubiquitous Computing". He was the chief technology officer at Xeroxs Palo Alto Research Center and had already in 1991 visions about Ubiquitous Computing in the 21th century. In his opinion "the most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it" [27]. Weiser [27] states that the computers should be integrated seamlessly into the world so that they vanish into the background. The importance of this disappearance can be explained through human psychology. If people have learned something sufficiently well, they cease to be aware of it and can use the information without thinking about it. They can process information in the background and therefore concentrate on foreground activities.
In contrast to Ubiquitous Computing Weiser mentions the notion of Virtual Reality. In cases of Virtual Reality people have to wear gloves or even bodysuits to move in a virtual world and manipulate virtual objects. Thus Virtual Reality needs enormous apparatus on simulating a new world, where Ubiquitous Computing uses invisible technology to support the real world .
As Weiser only had a vision about Ubiquitous Computing one decade ago, implementing such smart everyday objects is now actually possible because of the enormous technological progress.
Therefore haptics has to be put into the focus. Especially because of its advantage to appeal to the human tactile sense in the background it can be an important feature for devices in Ubiquitous Computing.

In order to provide appropriate feedback Hale et al. [9] point out that it could be an advantage to combine haptic feedback with feedback that appeals to other human senses. This means multimodal feedback. For example visions and haptics provide completely different types of information. Combining haptics and visual displays can improve task performance in object interaction, way finding, and collaborative environments. Like mentioned before, if the visual system is overloaded, haptic devices can provide information without significantly increasing cognitive load.
In order to develop a haptic interface or to integrate haptic feedback into everyday objects it is not enough to know about the history and

- *Annika Tonch is studying Media Informatics at the University of Munich, Germany, E-mail: annika.tonch@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

advantages of haptics. The effective designing requires a greater understanding of how best to convey haptic information what is related to physiological and psychophysical factors. One has to understand the users sensory, perceptual and cognitive abilities and limitations in order to estimate which haptic stimulation best suits a given task and therefore which kind of haptic feedback should be used [9].
In the next section principles of haptics are discussed as well as the definition and characteristics of the tactile and kinesthetic haptic stimulations.

2 Principles of Haptics

The term "haptic" can be defined as sensing, manipulating and perception through the haptic sense [23]. This sense responds to pressure, movement and temperature changes [28]. The human tactual sensory system has two subsystems: the tactile and kinesthetic sense [23]. These two terms are defined in the following subsections.

2.1 Tactile Mechanoreceptors

The tactile mechanoreceptors can be described as the sense of touch. Tactile cues appealing to this sense are for example vibrations or varying pressure and are best suited for simple alarms. Receptors for this sense exist in hairless skin like palms, fingertips and the soles of the feet and hairy skin that covers most of the body [9]. Examples of tactile stimulation is the sense of a texture or the vibration you feel when you touch loudspeaker [23].
Each receptor is sensitive to distinct physical parameters. According to Hale and Stanney [9] this depends on:

- "their size (large receptors have poor spatial resolution)"
- "density (many receptors in a given area results in high spatial acuity)"
- "frequency range (receptors dont perceive signals outside their range)"
- "nerve fiber branching (higher branching leads to spatial and temporal summation of signals)"

While hairless skin is most effective for detailed tactile information, hairy skin is sensory to vibration and static force and is suitable for spatial tactile cues. To perceive individual cutaneous signals the stimuli must be at least 5.5 ms apart. To perceive stimuli order they have to be 20 ms apart. The pressure sensors are activated by a force greater than 0.06 to 0.2 Newtons per cm. The just-noticeable values of pressure depend on body loci and pressure. It ranges from 5 milligrams on a womans face to 355 mg on a mans big toe. Vibrations must exceed 28 decibels so that humans can receive them [9].

2.2 Kinesthetic Receptors

Kinesthetic receptors should be appealed when tasks involve hand-eye coordination, for example object manipulation [9]. Tan and Pentland [23] describe this sense as the awareness of the limb position, for example how fast and in which direction they are moving. They also include muscle tensions mediated by sensors in the muscles, skin and joints. Examples of using the kinesthetic sense are estimating the weight of an object when you hold it in your hands or touching your nose with closed eyes, because your kinesthetic sense knows how to move the finger relative to the nose. In both cases the tactile sense is also involved because you touch the surface of an object .
According to Hale and Stanney [9] four receptors rely to the kinesthetic sense which can be found in muscles, tendons and joints:

- Receptors found in or near joints are called Golgi and Ruffini endings. They are considered as "protective receptors" because they are most active at extreme joint positions (full flexion or extension).
- Another receptors are called Ruffini endings. They are activated during static position and dynamic motion.
- Receptors in the tendons are called Golgi tendon organs. They sense active positioning (for example, the self-initiated movement of the arm in a given position) and static limb position.
- Muscle spindles give information about supported limb weight. This means the subjective weight of an object in a persons hand. Thats why they are thought to provide the conscious awareness of body movement.

Detecting limb positions depends on the joint moving, the movements velocity and the state of the muscles. The kinesthetic sense bandwidth ranges from 20 to 30 Hz where the sensitivity depends on the proximity of the joint rotations to the center of the body. For example the shoulder is more sensitive than a finger. The sensitivity also depends on the movement direction and speed. Haptic information transfer can most effectively be provided by a surface stiffness of 400 Newtons per meter and end-point forces of 3 to 4 Newtons [9].

3 Scenarios and Systems of Active Haptic Environments

There are many possibilities of different scenarios how to use haptic environments. For a better overview of existing scenarios a classification of the scenarios is described. After that two important systems of haptic environments are introduced, because they are used in many different scenarios so they have to be explained only once. After that a precisely definition of the presented scenarios is given before the different scenarios are introduced considering the classification.

3.1 Classification

The first step of classification is the consideration of different user groups. These can be divided into

- people with disabilities
  (for example blind or deaf)
- people in special environments
  (for example car driver, pilots or people in the cinema)
- people who need training
  (for example students of medicine or rescue teams)
- people who want entertainment
  (for example people who play computer games or watch movies)

Every user group can be further divided into the semantics of haptic, which means the kind of data that is transmitted by the haptic feedback. The semantics is classified in

- alarm
  (for example alarming through vibration)
- information
  (various information are transmitted)
- entertainment
  (haptic feedback is only for entertaining purposes)

The classification is charted in a tree diagram for better understanding, see figure 1
Certainly describing a scenario example for every category path is not possible. The reason is that on the one hand there is not enough literature available. On the other hand not every path makes sense. For example the user group "people who want entertainment" has no reasonable scenario for the semantics "warning" or "information". The same is true for the user group "people who need training" and the semantics "entertainment".

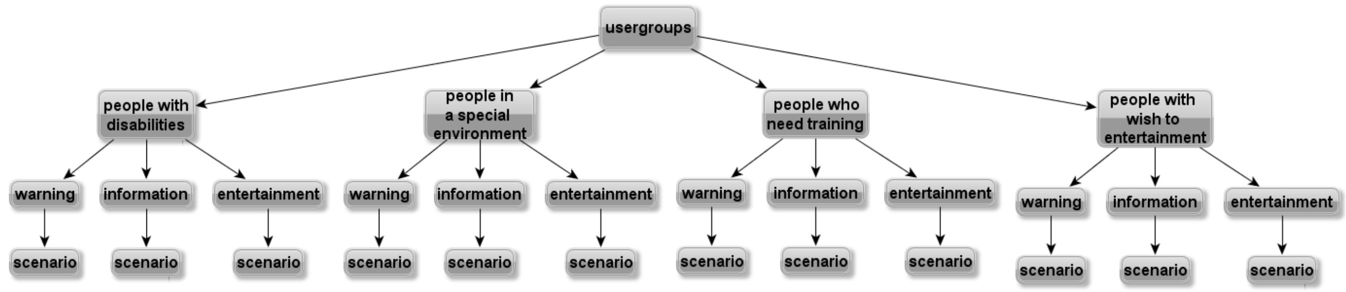


Fig. 1. The scenarios are classifies into different user groups and different semantics of haptic.

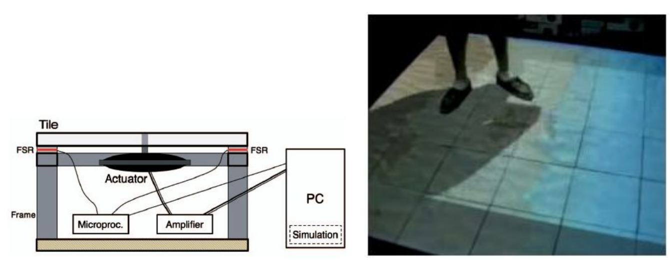
3.2 Important Systems

Before several scenarios of Active Haptic Environments are outlined, the haptic chair and the haptic floor are introduced. These two systems are important because they are used in various scenarios. They can be applied in different fields of user groups and with different semantics. In the next section they both provide the basis for several scenarios of Active Haptic Environments and are thus explained in detail.

A chair can be integrated into human-machine-interaction by making it a haptic user interface. Actuators (or tactors) can be mounted on the back of the chair and the armrest [22].
Actuators stimulate the humans haptic channel and exist in different types. The rotary motion tactor consists of a housing and a motor. The motor causes the housing to vibrate what sends stimulation to the skin. The housing does not have to make direct skin contact so that clothes can be worn. Linear actuators are small coil based actuators that provides a strong, point-like sensation to the skin [17].
Through the actuators the haptic chair can deliver haptic information to the user. This can be simple vibrations or geometric patterns what enables the communication of specific tactile cues [22].

According to Visell et al. [26] a floor can be used as a communication channel by using vibrotactile floor displays, see figure 2. Vibrotactile stimuli are presented via linear actuators, see figure 2(a). Different interaction cases are possible:

- structured vibrotactile signals, for example a sum of sinusoids with a certain temporal envelope
- stimuli that resemble physical objects, for example a hammer, tapping with a rhythmic pattern
- approximation of a natural ground material, for example gravel or snow, see figure 2(b)



(a) Technical components of the haptic floor. The actuator presents vibrotactile stimuli to the user [26].

(b) A haptic floor that simulates a beach. The structure of (wet) sand is approximated so that the user has the feeling of standing on a real beach [26].

Fig. 2. Communication via vibrotactile floor display.

An advantage of the haptic floor is among others the acceptability, because people do not have to touch something with the hands but only with the shoes. Besides the haptic floor is well suited for applications that involve pedestrian navigation.

For both systems several applications are described in section 3.3.

3.3 Scenarios

As mentioned in the introduction haptic devices can be classified in "wearable computing" and devices that fit into the environment. Wearable computing, where the user has to put on a special belt or helmet, is not presented in this paper. Considering the definition of Weiser wearable computing is more connected with Virtual Reality than with Ubiquitous Computing because it is not invisible and not integrated into the real world. The same is true for devices like the Phantom of SensAble Technologies which is a special haptic device people have to hold in his hands that enables the perception of virtual surfaces [5]. According to that only haptic devices are be presented that fit in the environment and are no "extra" devices.

3.3.1 People with disabilities

At first scenarios of the user group "people with disabilities" are presented.

Semantics: information
The haptic floor as described in 3.2 can be used as a communication system for blind or visually impaired people. Visell et al. [26] present a scenario where pedestrians can receive information about their environment via vibrotactile cues. If the pedestrian stands in front of the traffic light he can receive cues near the curb that the crossing signal is red or green, see figure 3. At the subway he gets haptic feedback near the platform edge to know if the train has already arrived. When the train arrives he gets haptic cues about the location of the doors. Another possible field of application is in elevators. Blind or visually impaired people get vibrotactile cues from the elevator floor about the current floor number so they know when to disembark.
The advantage of this scenario is that blind or visually impaired people get important local information via haptic feedback. Certainly this information could be presented via auditive signals. But if the pedestrian is in a noisy environment, haptic feedback is better perceived than auditive feedback. Another advantage of communication through a haptic floor is the tactile sensory physiology and psychophysics of the foot. The foot is one of the most sensitive parts of the body to vibrotactile stimulation [26].
Visell et al. conducted an experiment with 24 people (twelve female and twelve male) aged between 20 and 39 to evaluate the communication through the floor surface. The participants had to identify eight different communication units of the interface after five minutes of self-guided learning. The mean rate of correct identification was 61% with a standard deviation of 21 percent.

Fig. 3. A haptic floor that helps blind people in the road traffic. Vibrotactile cues provide information whether the crossing signal is red or green [26].

Fig. 4. A haptic chair as a sensory substitution system. The music chair enables deaf people to hear music by converting sound into vibrations [24].

Semantics: entertainment
The haptic chair as described in 3.2 can be used to enable deaf people to feel music, see figure 4. Karam et al. [12] introduce a system, called "Emoti-Chair" where "music can be experienced as a tactile modality, revealing vibrations that originate from different instruments and sounds spanning the audio frequency spectrum along multiple points of the body". The model human cochlea (MHC) is used to convert sound into vibrations, which is a sensory substitution system.
Sensory substitution means translating sensory information that is available via one sense to another. Commonly the term describes a technological translation of signals realized by a system. Information about the environment from sensors corresponding to a specific modality is transduced into a set of signals and is digitized. This data is mapped into another set of signals and is presented to another human sensory modality [25].
The MHC translates music into discrete vibration signals. The vibration signals are displayed on the back of the users body by the usage of voice coils mounted on the back of the chair [13].

The Emoti-Chair develops a large part of the skin as a "pseudo-hearing organ"[12]. This is an important approach because it shows the potentials of the skin as the largest organ of the humans body. However in most human-computer-interaction systems only a small part of the skin, for example the hands, is involved [23].

3.3.2 People in a special environment

In the following scenarios of the user group "people in a special environment" are outlined.

Semantics: warning & information & entertainment
Haptic devices integrated in the car equipment can be a useful application for car drivers. Haptic feedback through vibrations of the steering wheel or the seat can provide important information to the driver.
Van Erp et al. [7] present five different types of information that can be provided by tactile interfaces:

1. Spatial information. The visual system is restricted in the field-of-view which often results in accidents because of cars in dead angles or crossing pedestrians.

2. Warning signals. As the visual channel is overloaded the tactile channel is always ready to receive information which is well suited for warnings. Auditive warning signals can be lost in radio or conversation noise.

3. Communication. Tactile displays allow silent communication with the driver so that the passengers will not recognize the communication of the car to the driver.

4. Coded information. Abstractly information like speed or fuel supply can be presented to the driver when he asks for it.

5. General. The driver can be guided to specific locations through tactile information.

In addition they name four categories of applications:

1. Safety. Haptic feedback allows the driver to focus on the road while driving, because the visual workload is released. Actuators in the seat can inform the driver about cars or cyclists in dead angles or crossing pedestrians.

2. Assistance. An example is a small tactile vibration on the handbrake-release-button if there is enough torque on the wheels. Another possibility is a tactile stimulus on the accelerator pedal to indicate when a speed limit was exceeded.

3. Fun. Tactile information enables private and silent communication. For example a tactile signal on the door handle when lights are left on is more pleasant than an annoying auditive signal.

4. Efficiency. Tactile displays integrated in the steering wheel can indicate the moment to shift gears. This can support a more efficient gearshift regime.

Another scenario example is the lane departure warning system by Citroen [20]. Infrared sensors detect a lane change when the driver crosses a line for example because of fatigue or negligence. If this happens the driver is warned by a vibrating mechanism in the seat, see figure 5.
Semantics: information
The Tactile Situation Awareness System (TSAS) introduced by McGrath et al. [18] is a System that presents information to pilots via tactile stimulators. It was developed to counter mishaps because of spatial disorientations during the flight. Spatial disorientation occurs mostly during drift and/or descent in hovering flight. Developed systems like Multi Function Displays or Helmet Mounted Displays which appeal the visual sense could not eliminate this problem. Either they are not interpreted correctly or they are even ignored.
The TSAS provide situation awareness information through haptic tactile cues. The tactors are integrated into the flight garments, see figure

Fig. 5. The Lane Departure Warning System warns the driver by a vibrating mechanism in the seat if he crosses the lane, for example because of fatigue or negligence [20].

6. It can present various flight parameters like acceleration, velocity or the target location. The TSAS helps the pilots to navigate during complex flight stages and allows the pilots to focus on other instruments and systems by reducing operator workload. To display a particular flight parameter indicators are varied. These indicators are the tactor position and the activation or pulse patterns which means the rate of turning the tactor on and off. For example if the helicopter is moving left the tactors on the left side are activated. If the helicopter is drifting in the range of 0.3 to 0.7 m/sec the tactor is activated at 1 pulse per second [18].
Haptic feedback integrated into flight garments could also be interpreted as wearable computing. In this paper it is classified as Active Haptic Environment because it is no "extra device". Pilots would wear their flight garments anyway so that they can be seen as their natural environment.

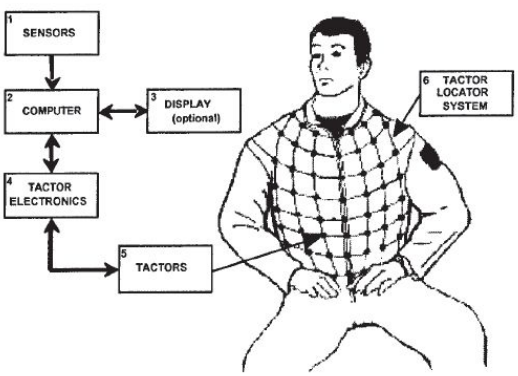


Fig. 6. The Tactile situation Awareness System is integrated into the flight garments and provides situation awareness information to the pilot [18].

Semantics: entertainment
The model of the haptic chair can also be used in the cinema to deeper the sense of immersion and believability of movies. In todays entertainment technologies emphasis is put on extensive visual and audio effects to make the movie more realistic for the audience. Integrating haptic devices into the cinema experience strengthens this effect, see figure 7.
Israr and Poupyrev [10] developed a system that stimulates the skin with complex waveforms. This system can among others be used in cinemas to enrich movie effects. A chair is used as a haptic surface where twelve tactile actuators are mounted on the back of the chair in three rows. These actuators get the power from a control board.
In order to create new haptic special effects, Israr and Poupyrev created a set of haptic morphs that can be combined or interchanged. These morphs are:

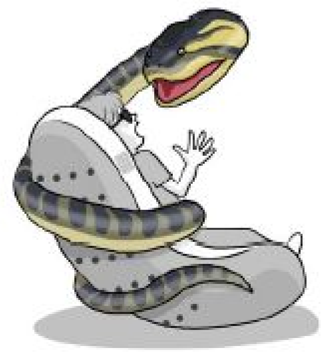
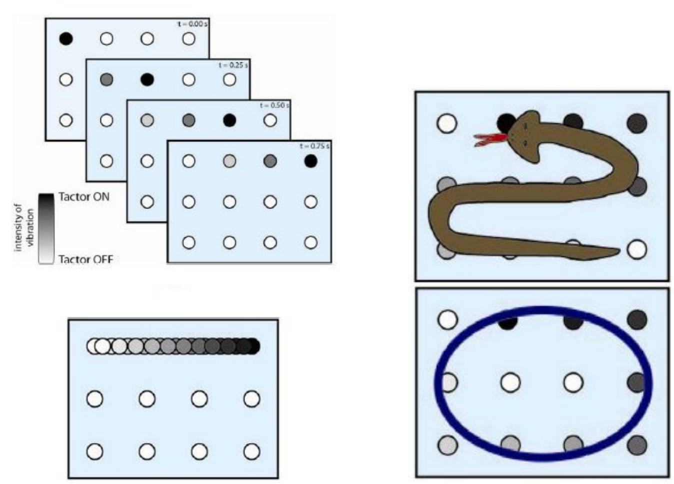


Fig. 7. The haptic chair enables the user to feel what he is seeing in the movie. Creating the feeling of continuous motion the visitor can for example feel the creeping motion of a snake that is shown in the movie scene [10].

- "Onset turn ON the channel abruptly"
- "Reset turn OFF the channel abruptly"
- "Linear rise: rises amplitude linearly"
- "Linear decay: decays amplitude linearly"
- "Exponential rise: rises amplitude exponentially"
- "Exponential decay: decays amplitude exponentially"
- "Frequency modulation: linear modulation of frequency from the start level to the final level"

An important goal of Israr and Poupyrev is to create the feeling of continuous motion. In order to not having to use a large number of actuators they developed a method called "Haptic Blur", see figure 8. The method of the Haptic Blur is a uniform intensity at the point of stimulation. The intensity steadily dies out away from this point. The locus of stimulation is limited in time and the intensity decreases to zero at the temporal reset, see figure 8(a). By combining the haptic blur with amplitude and frequency modulation a feeling of creeping or circular motion can be simulated, see figure 8(b).



(a) The locus of stimulation is limited in time and the intensity decreases.

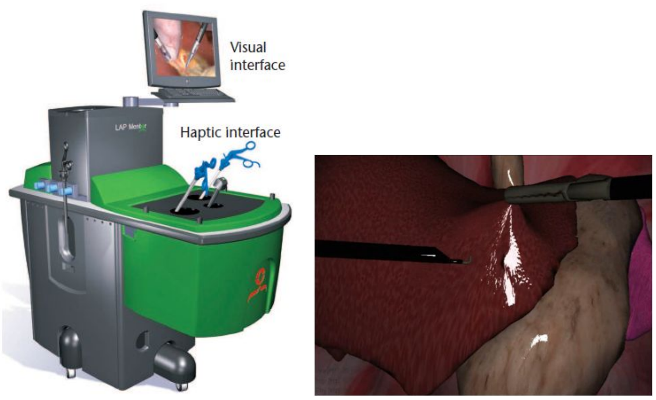
(b) Simulation of a feeling of creeping or circular motion.

Fig. 8. The Haptic Blur simulates the feeling of continuous motion [10].

3.3.3 People who need training

The next scenarios are classified as the user group ”people who need training”.

Semantics: warning & information
Haptic feedback can be an important tool to support minimally invasive surgical simulation and training (MISST). In minimally invasive surgery (MIS) procedures, a small camera is inserted through a small skin incision and the surgeons can view their actions on a video display. Compared to open surgery MIS has reduced the sense of touch because the surgeons have to interact with the organs through surgical instruments attached to long, thin tubes. As a consequence more training with the help of MISST is needed. Haptic feedback can therefore be a good option to improve MISST [1]. A typical MIS Simulator consists of a visual display and surgical instruments fitted with haptic devices, see figure 9(a). When the user moves the surgical instruments a collision detection algorithm checks whether the haptic devices touch the virtual organs. If this is the case force feedback is provided to simulate the interaction with organs in a real surgery. The user can touch and manipulate virtual organs, see figure 9(b). Altogether the MIS Simulator enables Palpation, Grasping, Translocation, Clamping, Cutting, Dissection and Suturing of virtual organs.



(a) A typical MIS Simulator consists of a visual display and surgical instruments fitted with haptic devices [1].

(b) At the display of a MIS Simulator the student can see the virtual surgery. The haptic devices provide feedback according to the operations that are displayed [15].

Fig. 9. Haptic feedback makes virtual surgery training more realistic.

Integrating haptics into MIS simulators can improve the training situation which leads to better educated students and reduces the need for using animals and human cadavers for medical training [2].

Semantics: information
Visell et al. [26] suggest the haptic floor as a supportive haptic system for people who need training. The haptic floor can be integrated in an augmented reality training simulation for rescue teams. Realistic cues provided by a haptic floor can inform the trainees about the material and local stability of a ground surface as well as changing conditions during an emergency. In this way the trainees can learn how to behave during a real emergency. For example they can examine the stability of a frozen lake and learn how to move on it, see figure 10.

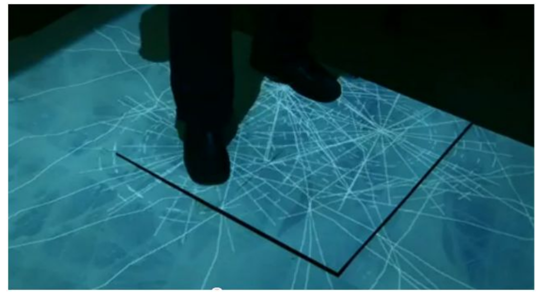


Fig. 10. A haptic floor that simulates a frozen lake. The user not only sees but also feels the cracking of the ice if he steps to hard [14].

3.3.4 People who want entertainment

At last scenarios of the user group ”people who want entertainment” are presented.

Semantics: entertainment
Haptic feedback can also be supportive in the field of entertainment to integrate the sense of touch into computer games. By now gloves or mechanical attachments were required to realize the extension to the tactile channel [3].
Iwamoto et al. [11] developed the Airborne Ultrasound Tactile Display that provides tactile feedback to the users hand in 3D space, see figure 11. To realize this kind of tactile feedback acoustic radiation pressure is used. When the users hand interrupts the propagation of ultrasound a pressure field is exerted on the hand. Furthermore the system consists of a vision based hand tracking system comprised of a single camera. The camera tracks the position of the users hand and a number of ultrasonic transducers emit ultrasound. This creates a feeling of tracing the edge or surface of the virtual object. The advantage of this system is that one can feel it with both hands and several people can use it at the same time.

Fig. 11. Airborne Ultrasound provides tactile feedback to the users hand in 3D space. For example it can enable the user to feel rain drops on his hand that are displayed on a screen [21].

Dionisio [6] developed a technique to add a temperature stimulus within tactile feedback. This VR Thermal Kit consists of ventilators and infrared lamps, sensors that record the surrounding environmental temperature, actuators, a peltier element with direct contact to the users skin and a control system that connects with the graphics workstation, see figure 12. The ventilator and the infrared lamps cool or warm large areas of the body skin and the ventilator also generates perceptible air motion. The Peltier elements cool or warm the skin locally. According to the temperature in the computer game the ventilators or lamps are activated with distinct intensity levels. The user can now feel the same temperature as the virtual temperature in the computer game.

4 Conclusion

This paper demonstrates the big potential of active haptic environments in general but especially in relation to the fact that today the visual channel is overloaded because of small computer screens. Besides the technological process makes active haptic environments possible.
Altogether there is not much literature available in the field of Active Haptic Environments. Many paper focus on wearable computing or

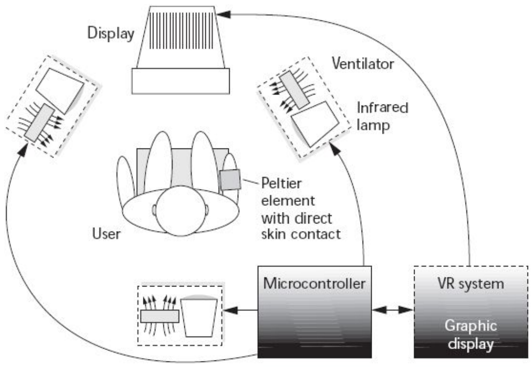


Fig. 12. The thermal display enables the user to feel the temperature of computer games or movies through ventilators, infrared lamps and a peltier element that has direct contact to the skin [6].

”extra devices” like the Phantom instead of haptic devices integrated into the environment. Therefore I could not find any literature dealing with the user group ”people with disabilities” together with the semantics ”warning”.
In relation to the number of papers lots of literature can be found in the field of warning systems for cars. According to Gallace et al. [8] the majority of commercial vehicles will have some kind of vibrotactile stimulation device by 2020. This shows that haptic environments could gain acceptance at least in the automotive sector. Lots of literature can also be found in the field of sensory substitution or in the context of this work the user group ”people with disabilities” together with the semantics ”information”. Therefore not every existing application is presented in 3.3 for example like vibrotactile letter reading. Mitsuhiro et al. [19] developed a method to read letters with the help of the haptic chair. A 3by3 array of vibrating motors mounted on the back of a chair is used to communicate alphanumeric letters to the user. Tests with this application revealed a result of 87% successful letter recognition.

In my opinion the most important field of application for Active Haptic Environment lies inside the user group ”people with disabilities”. This is also clarified by the fact that researches have been made on sensory substitution technologies since the early 1970s like mentioned in the introduction. The haptic sense can act as an important communication channel for blind or deaf people. Information that healthy people see or hear can be delivered through haptic feedback. Like described in 3.3 applications that use sensory substitution can not only be applied in scenarios where important information about the environment have to be transmitted but also in scenarios with the goal to entertain people. The fact that these applications are integrated into the environment makes it feel a lot more natural for the users and more intuitive to use. Sensory substitution applications in active haptic environments can therefore be important and helpful tools in everyday life of disabled people.
Another important field of application is for me the training situations. Haptic feedback can make training simulations much more realistic. Students of medicine or rescue teams can be better and faster trained which results in fewer mistakes for example in surgeries or emergencies.
In addition I think that haptic feedback has big potential concerning warning systems. As described in the introduction and 3.3 the haptic channel is always ready to receive and is characterized by a fast reaction time. This makes the haptic sense a perfect channel to receive simple warning signals for example through vibrations especially when the visual sense is overloaded and auditive warning signals vanish in noise. As this is not as complex as transmitting specific information it can be used in various scenarios. Examples are besides the warning systems in cars that are already mentioned in 3.3 warning signals in the road traffic for example when you stand too near at the street when the crossing signal is red or simply when you forget to lock the door or turn off the stove when you are leaving you could get vibrotactile signals from the floor in front of your front door. Another advantage of haptic warning systems is that they are not disturbing for oneself or other people around because they are silent and personal.
Even so I think that these three fields of applications are the most important ones, I believe that economically speaking the field of entertainment has the biggest potentials. Entertainment systems are continuously advanced to make media like movies or computer games more realistic. Nevertheless visual and auditive effects are already on a high standard which makes it difficult to advance them. Integrating the haptic sense is a new technique that not only has big effects but also can be further developed.

As already mentioned I could not find any literature concerning the user groups ”people with disabilities” combined with the semantics ”warning”. As these fields are in my opinion the most important ones I will describe some possible scenarios focusing on that.
The haptic floor as communication channel for blind people that gives detailed cues about the environment could be reduced to a simple warning system. The floor could only give warning signals in front of a street when the crossing signal is red or in front of other potentially dangerous spots like construction zones or big crossroads. Like mentioned before another example could be a warning signal through the floor in front of the front door or through a haptic chair in the living room when the stove is left on. Such an application would be less complex and expensive and more intuitive like detailed cues through haptic feedback.
In my opinion thermal feedback like the VR Thermal Kit of Dionisio [6] has also big potential though there is only few literature available. When you think of appealing the haptic sense you mainly think of pressure of movement. But temperature changes should be kept in mind as well because it conveys a completely different feeling. Thermal feedback could be integrated into the haptic chair for cinemas what could help making movies more realistic. Thermal feedback could also be applied in training simulations to make them more realistic. For example rescue teams could learn to act in big heat when there is a fire or in bad cold in the winter.

Altogether many additional scenarios next to the scenarios presented in 3.3 are possible. This shows the potential of Active Haptic Environments and their advantages over feedback using other human senses.

List of Figures

References

[1] C. Basdogan, S. De, J. Kim, M. Muniyandi, H. Kim, and M. A. Srinivasan. Haptics in minimally invasive surgical simulation and training. *IEEE Comput. Graph. Appl.*, 24:56–64, March 2004.
[2] C. Basdogan, M. Sedef, M. Harders, and S. Wesarg. Vr-based simulators for training in minimally invasive surgery. *IEEE Computer Graphics and Applications*, 27:54–66, 2007.
[3] BBCMobile. Ultrasound to give feel to games. http://news.bbc.co.uk/2/hi/technology/7593444.stm, 2008.

[4] M. Benali-khoudja, M. Hafez, J. marc Alex, and A. Kheddar. Tactile interfaces: a state-of-the-art survey. In *Int. Symposium on Robotics*, pages 721–726, 2004.

[5] J. Blaurock. *Haptische Wahrnehmung rechnerinterner Freiformflächen*. PhD thesis, Technische Universitt Kaiserslautern, 2004.

[6] J. Dionisio. Virtual hell: A trip through the flames. *IEEE Comput. Graph. Appl.*, 17:11–14, May 1997.

[7] J. B. F. V. Erp. Vibro-tactile information presentation in automobiles. In *University of Birmingham*, pages 99–104, 2001.

[8] A. Gallace, H. Z. Tan, and C. Spence. The body surface as a communication system: The state of the art after 50 years. *Presence: Teleoper. Virtual Environ.*, 16:655–676, December 2007.

[9] K. S. Hale and K. M. Stanney. Deriving haptic design guidelines from human physiological, psychophysical, and neurological foundations. *IEEE Comput. Graph. Appl.*, 24:33–39, March 2004.

[10] A. Israr and I. Poupyrev. Exploring surround haptics displays. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, CHI EA '10, pages 4171–4176, New York, NY, USA, 2010. ACM.

[11] T. Iwamoto, M. Tatezono, T. Hoshi, and H. Shinoda. Airborne ultrasound tactile display. In *ACM SIGGRAPH 2008 new tech demos*, SIGGRAPH '08, pages 1:1–1:1, New York, NY, USA, 2008. ACM.

[12] M. Karam, C. Branje, G. Nespoli, N. Thompson, F. A. Russo, and D. I. Fels. The emoti-chair: an interactive tactile music exhibit. In *Computer Human Interaction*, pages 3069–3074, 2010.

[13] M. Karam, F. A. Russo, and D. I. Fels. Designing the model human cochlea: An ambient crossmodal audio-tactile display. *IEEE Transactions on Haptics*, 2:160–169, 2009.

[14] A. Law, J. Ip, B. Peck, Y. Visell, P. Kry, and J. Cooperstock. Ecotile niw ice cracking and snow. http://www.youtube.com/watch?v=AsKBigMD7fg, 2009.

[15] A. Maciel. Simulating minimally invasive surgery: create and touch your own virtual liver. http://vimeo.com/7189522, 2009.

[16] F. Mattern. Vom verschwinden des computers – die vision des ubiquitous computing. In F. Mattern, editor, *Total vernetzt*, pages 1–41. Springer-Verlag, Apr. 2003.

[17] B. McGrath, A. McKinley, M. Duistermaat, O. Carlander, C. Brill, G. Zets, and J. van Erp. Tactile actuator technology. In *Tactile Displays for Orientation, Navigation and Communication in Air, Sea and Land Environments*, pages 4/1–4/12, 2008.

[18] B. J. McGrath, A. Estrada, M. G. Braithwaite, A. K. Raj, and A. H. Rupert. Tactile situation awareness system flight demonstration final report. 2004.

[19] Y. Y. Mitsuhiro, M. Kakita, R. W. Lindeman, Y. Kume, and N. Tetsutani. Vibrotactile letter reading using a low-resolution tactor array. In *Proc. of the 12th Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 400–406, 2004.

[20] PSAPeugeotCitroenSustainableDevelopment. Avoiding accidents. http://www.sustainability.psa-peugeot-citroen.com/corporate-citizenship/priorities/overview/achievement.htm?id=2708, 2008.

[21] ShinodaLab. Touchable holography. http://www.youtube.com/watch?v=Y-P1zZAcPuw, 2009.

[22] H. Z. Tan, I. Lu, and A. Pentland. The chair as a novel haptic user interface. *Proceedings of the Workshop on Perceptual User Interfaces*, 1997.

[23] H. Z. Tan and A. Pentland. Tactual displays for sensory substitution and wearable computers. In *ACM SIGGRAPH 2005 Courses*, SIGGRAPH '05, New York, NY, USA, 2005. ACM.

[24] TheMusicGallery. http://www.musicgallery.org/node?page=19.

[25] Y. Visell. Tactile sensory substitution: Models for enaction in hci. *Interact. Comput.*, 21:38–53, January 2009.

[26] Y. Visell, A. Law, and J. R. Cooperstock. Touch is everywhere: Floor surfaces as ambient haptic interfaces. *IEEE Transactions on Haptics*, 2:148–159, 2009.

[27] M. Weiser. Human-computer interaction. chapter The computer for the 21st century, pages 933–940. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995.

[28] A. Wright. The touchy subject of haptics. *Commun. ACM*, 54(1):20–22, 2011.

Types of Public Interactive Display Technologies and How to Motivate Users to Interact

Neal Buerger

Abstract— Currently several different forms of public interactive displays are being researched. The two main categories are ”public displays in combination with mobile devices” and ”standalone displays”. Looking at all these types of displays, the most important design aspects for public interaction displays can be extracted: Non-distracting display of information, comprehension, notification, short-duration fluid interaction, immediate usability, shared use, combining public and personal information and privacy.
All public displays have a problem of acceptance: users have an aversion to use them as they are fearful of unfamiliar technology, but the bigger barrier is fear of social embarrassment. When using a public interaction display, users are aware that they are being watched. To motivate users to interact with the devices the devices have to be ”easy-to-use” and support ”natural” gestures. The user must have a positive tendency towards the device - there must be a benefit in some way by interacting with the device. This can be accomplished, for example, by rewarding the user for interacting with the device.
It has been shown that when a helper or another user introduces a new user to the public screen, the new user is more comfortable using the screen. When the display is in use, it automatically attracts more people. Most users learn how to operate the display by watching other users. Usually an individual from a group interacts with the public display while receiving comments from the group.

Index Terms—social factors, large displays, mobile device,interaction displays, input device, interaction techniques, multi-touch, augmented reality, multi-display environments, situated public displays, urban environments, multi-user interfaces, ubiquitous computing

1 Introduction

Currently more and more types of industry are starting to use various forms of digital signage. Most commonly large LCD screens are used for digital signage. Public displays allow the presentation of information in a more compelling fashion than ever before. Instead of having a single static content, the display can easily be adapted to dynamically display information. Current displays connect to a content management server to determine what information to display [1].

The next evolutionary step of this technology is to enable users to interact with these displays. Several prototypes of public interaction displays have been proposed. When the prototypes were evaluated, researchers discovered that a fear of social embarassement has to be overcome by users before they start using the display. The question remains how to design an ideal public interactive display and how to motivate users to activly interact with the device by helping them overcome their fears.

This paper covers three aspects of interative display technologies. It starts with the different interaction phases that were observed with many public interaction displays. Then it describes several different forms of public interaction displays. The paper continues with addressing general design aspects for creating a successful public interaction display, and with ways on how to encourage users to use these devices.

2 Interaction Phases with Public Interactive Displays

Several researchers have discovered that users react differently towards the display depending on the physical distance of the user to the display. They discovered four distinct phases of interaction [26] *(see figure 1)*.

Ambient Display Phase: The user passes by the display from a distance and does not interact with the device. All other types of interaction should not obscure this main interaction phase. Users should be able to process the displayed information at a quick glance.

- *Neal Buerger is studying Media Informatics at the University of Munich, Germany, E-mail: info@nealbuerger.com*
- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

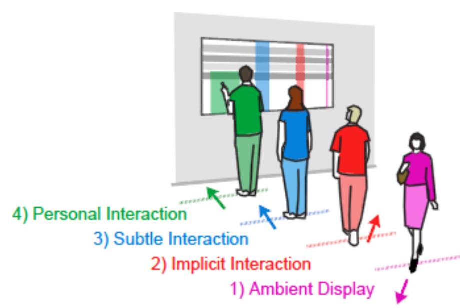


Fig. 1. The different types of interaction phases [27].

Implicit Interaction Phase: The display should be able to detect the position and orientation of the users body. If the user is moving towards the display or stops to watch the display, he is signalling that he is open to receive more information.

In this phase the user can be subtly notified and encouraged to come closer to the display. General notifications should be displayed. If there is a possibility to identify each user individually, personal notifications could be displayed. The user should have the possibility to signal to the display that he is not interested in interacting with it.

Subtle Interaction Phase: The user is moving towards the display. The display could expand the notifications and display more details. The user is approximately an arm's length away from the display. This is the ideal position to interact with the device directly by simple hand gestures, for example, to select items of interest. A single user at this position does not obscure the device, allowing multiple users to share the screen and simultaneously operate the screen. In this phase highly personal data should not be displayed.

Personal Interaction: The user can move in even closer to the display and, in addition to the hand gestures, is able to touch the display for interaction. The user can accurately interact with the device and interact precisely with the provided information. The user is also now standing very close to the display, so the user's body helps to occlude the view of other users. Still, very sensitive personal information is not protected and should not be displayed.

The users should be able to transition between these phases seamlessly [27]. ”Users initially signal a phase change using implicit interaction, such as body movement, body location, and head orientation, then gradually become more explicit with gestures and touch" [12].

3 Types of Interactive Public Displays

Currently several types of public displays are being researched. They can be categorized into two groups: "public displays in combination with mobile devices" and "standalone public displays".

The two groups of displays and existing applications are introduced in the subsequent chapters.

3.1 Public Displays in Combination with Mobile Devices

Large displays and mobile devices have become ubiquitous. Combining both technologies allow users to interact with the display with their mobile phones. Many screens would only need a minor upgrade to support interaction. But the major limitation of all these types of displays is that no system can efficiently support every type of mobile device and users must have a supported device to actively interact with the display. These displays are designed to work in the "Ambient Display Phase".

Interaction techniques can be categorized into three groups [21]:

1. *Extended input device*: The mobile device acts as a simple upload client that can provide text, image, and video input for the pubic display.

2. *Pointing device*: A cursor on the screen is controlled by the mobile device.

3. *Integral part of the interaction*: The mobile device serves as an extension of the display and allows with an additional interface to control various features of the display.

3.1.1 USIAlumni Faces

The USIAlumni Faces is a virtual yearbook application; the application was projected onto a large public screen. For all interaction with the screen, a customized Nintendo Wii remote control combined with an infrared pen built into a toy torch casing was used ("Interactive Artifact") *(see figure 2)*.

Fig. 2. A user interacting with the Alumni Application [19]

The screen was designed to enhance events like homecoming and alumni events. The artifact controlled the virtual yearbook application (photos of the alumni organized by year and faculty). Specific hand gestures controlled, for example, the flipping of pages.

The application was tested during a real life alumni event. Instead of explaining to the users how to operate the screen, the researcher simply informed the participants of the main function of the screen. This allowed to research social interaction around public displays as well as researching how gesture interfaces are expected to work. The interactions with the screen were videotaped and analyzed.

Users mainly learned how to use the application by observing others in a process of imitative learning. Some users played around with the artifact, while others watched them and tried to reproduce the same gestures to produce similar results.

Usually groups of people (from 2 to 8 people) used the display together; individual interaction was rather rare. While groups used the screen, one individual operated the screen, while the group suggested which information should be displayed.

The display stimulated social interaction. People, who met while interacting with the display, continued to talk with each other even when they stopped interacting with the screen. The display was used as an ice-breaker and allowed people that had never met before to start a conversation[19].

3.1.2 MobiLenin System

With the MobiLenin System users can use their mobile phones to interact with interactive video on large public displays. The system was specifically designed to allow multiple users to vote, which video would be next, and to allow collaborative and competitive interaction. The mobile phone has a specialized Symbian client installed to display the current video choices and to input a vote. The main server processes the votes, shows the result of the poll, and switches the video accordingly [22].

The MobiLenin system was evaluated in a real world setting in a restaurant in Oulu, Finland. The general feedback was that the system was "easy to use" and enhanced the social experience [22].

Users can only interact indirectly with the display and actually do not need to see the display. It seems that the large display is unnecessary and the results could as well be just displayed on the mobile devices.

3.1.3 Touch Projector

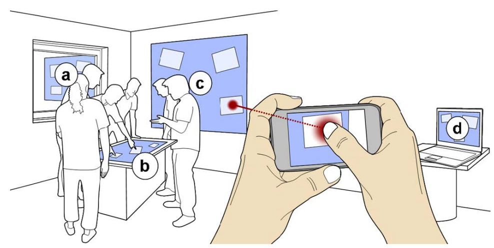


Fig. 3. "Touch Projector allows users to manipulate content on distant displays that are unreachable, such as (a) displays outside a window, or (b) a tabletop system. It allows users to manipulate devices that are incapable of touch interaction, such as (c) a wall projection or (d) a laptop. Users point the device at the respective display and manipulate its content by touching and dragging objects in live video. The device projects the touch input onto the target display, which acts as if it had occurred on itself" [3]

.

The touch projector system is based on the vision that users could remotely interact with devices via live video [25]. The basic functionality of the touch projector enables screens without touch capability to receive touch capability through the smartphone *(see figure 3)*. The system uses a smartphone (Apple iPhone) and a dedicated computer connected to the screens. The camera of the smartphone identifies screens by analyzing the objects currently displayed on the screen. In addition to directly interacting with a single screen, objects could be moved from one screen to another screen [3].

Interaction through live video is not limited to traditional display technology. The Touch Projector system was enhanced to control a media facade at the ARS Electronica Festival in Linz. The new system also allowed multiple users to interact simultaneously with the same facade *(see figure 4)* [4].

3.2 Standalone Public Displays

With standalone displays, no additional device is required to interact with the screen. This allows users to interact with the display in the implicit, subtle, and personal interaction phases. Most systems presented work best in the personal interaction phase.

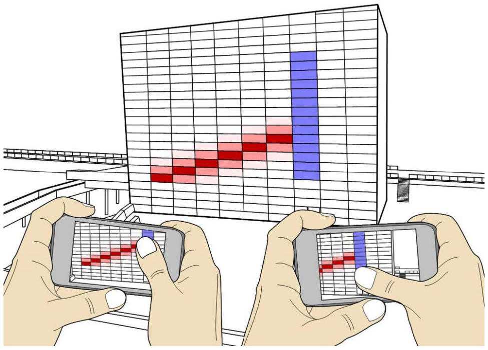


Fig. 4. "Interacting through live video allows multiple users to manipulate a media facade. Changes (also those of other users) are shown immediately on the facade and the mobile device. Colors denote actions from other users" [4]

Fig. 5. A user interacting with the Interactive Clothing Bins [23]

Fig. 6. The mixed-media products table [23]

3.2.1 Everywhere Display Projector

To enable dynamically projecting images and text everywhere around the projector, the Everywhere Display Projector was built. The LCD-projector is combined with a computer-controlled pan/tilt mirror. This particular projection system compensates for distortions that occur when changing the position of the mirror [16]. With the addition of an independent camera system, the system tries to track and identify hand motions. This is accomplished by analyzing frames to discover fingertips and their changing position between video frames. A touch sensing widget is also implemented in the device, enabling users to touch projected images and interact with sliders and buttons [23].

The system was created to provide additional information in an retail environment. Two types of interaction were implemented with this technology:

- *Interactive Clothing Bins:* The clothing bins are assembled in a checkerboard pattern of alternating merchandise and displayable space. Reaching into a clothing bin triggers to highlight the current bin and the displayable space to display information about the content of the bin. In addition, a touch interface enables to see further information for sizes, available stock, and customer service *(see figure 5)*.

- *Mixed-Media Products Table:* A round table with items placed on it, but leaving a 6 inch "'margin"' of table space as display space. The table displays keywords and highlights which items have the same keyword in common. When the user walks around the table, depending on his position, the display shows other keywords *(see figure 6)*.

The design was tested in a replicated 16x11 foot (5x3.3m) retail space in an laboratory. Users could use the "interactive clothing bins" filled with various types of pants and display detailed information about the pants. "The mixed-media products table" was stacked with Halloween-themed CDs, DVDs, and books. Keywords like "Ghosts", "Witches" etc. were used to describe the items.

During the evaluation of both systems several observations were made. Users did not know that their position was influencing the displayed information, leading to only unaware interaction with the system. Users simply assumed they were in an environment where information could change at any time.

Also, users initially were hesitant to touch projected images. After a short time of getting used to the system, users were comfortable using the touch interface [23].

3.2.2 City Wall

The City Wall interaction screen uses a rear view projection system. The screen can simultaneously track multiple hands and gestures. The interface was designed specifically to handle different media content, especially images. This system is ideal for placing a screen into an urban environment because of its multiple user support *(see figure 7)*. It is 2.5 meters wide and, during testing, up to 7 users used the display in parallel.

4 Design Considerations for Public Displays

Most humans are intimidated by new technology and are hesitant to adapt to new technologies. Public displays have additional social barriers, because of the fact that they are public, and that users are afraid to be publicly humiliated when using such an device. So besides the technical design challenges to be addressed, one must also address social acceptance issues.

4.1 Interesting Public Displays

When users pass a public display, the display must attract the attention of the user. People want to determin very fast, if something is interesting or not. The correct size and placement of the public display is key to attract users.

"Assume that viewers are not willing to spend more than a few seconds to determine whether a display is of interest. If the intent of the content is to be informative, present it in such a way that the most important information be determined in 2-3 seconds and avoid using more than minimal text [11]."

The display should be on eye level with the users to encourage them to glance at the display. The presented content should be updated dy-

Fig. 7. The City Wall used by two users [2].

namically to raise interest. The content should not abruptly shift to new content to encourage longer viewing of the display.

The placement of the display should be supported by its surroundings. The display should be the most interesting object to see and not be distracted by other information, for example, by posters or flyers. If it is possible, the display should be in the direction of the people's movement. Another possibility is that other objects draw attention to the display [11].

The size of the display is important for how the user perceives the experience. External factors play an important role, like how the user is exposed in public, or how much privacy is required. It is very important that the user is comfortable when using a public display. It may be beneficial to use multiple different screen sizes to comfort different types of users.

Observations have shown that when the same content is displayed on small and large screens, users pay more attention to the small screen, mainly because they feel as if it would be like a more private presentation and they are more comfortable looking at the content [11].

4.2 Design Principles

In 2004, Vogel and Balakrishnan compiled a list of fundamental design principles for public displays.

Non-distracting Display of Information lets users accept the display as a natural object that fits into its surroundings. Public displays should blend completely into their environment. It should fit with the surrounding architectural design and not disturb users in their common activities. Especially, content should assume a similar color palette as the surroundings of the display [7].

When the content of the display changes too fast, it disturbs and distracts users. The precise timing is key and has to be determined individually. A system with a too slow reaction time appears to be unresponsive and discourages users [8].

Comprehension of what the device can do and how to operate the device is important. "An interactive display should reveal meaning and functionality naturally" [27]. The user should not be intimidated by the screen and should be invited to use the display by its "easy-to-use".

Notifications have to be made in a socially acceptable manner. Displays should not forcefully interrupt passing people to attract users. The display has to determine the interruption tolerance of a potential user by analyzing the number, speed, and orientation of potential users [27]. In many cases, no explicit notification is needed, because active users automatically attract more users [15][6].

Short-Duration Fluid Interaction means the display integrates into common activites that take place in the surroundings of the display. The user should not be disturbed in his usual activities. Information should appear as naturally as possible. For example, in a retail environment the primary activity is to look and select merchandise. The display of additional promotional offers and other information should not intrude or distract from this activity [27].

Immediate Usability allows users to operate the devices with commonly known gestures. Visual aids are required to make it clear, if the screen is touchable, or if it is operable with hand gestures. Any form of explicit explanation, before using the device, discourages users to use the device.

Interaction techniques based on the position of the user as well as touchable projected images should be avoided, because users need prior training to operate the display.

If more complex interaction techniques are required, they should be discoverable by experimentation, or suggested while using the display. When the screen is used by multiple users, it is also possible that such knowledge gets passed on by observing other users operate the screen [19].

Shared Use is essential to create a positive experience for users. Many advantages from large public interaction displays come from the possibility that multiple users can share the display "either individually or collaboratively whether interacting implicitly, explicitly, or simply viewing the ambient display" [27] .

Even when a terminal is designed to only handle a single user at a time, the display should still be able to connect to an online social space to provide a possibility for discussion and collaboration [8].

Combining Public and Personal Information is a very delicate topic. Only "harmless" personal information should be displayed on an public screen. "By harmless, we mean information that one is not too concerned about others viewing" [27]. However, every person has to determine individually what he considers "harmless" information. "Information considered totally innocuous to some, is considered personally private to others" [14]. Before using the display with private information, the user should be able to set up the system in a way that he is comfortable with using the display.

Privacy cannot be maintained with large interaction displays, especially because people tend to be more voyeuristic with large displays [24] and, even standing in close proximity, does not occlude enough of the display to discourage eavesdroppers. When handling private information, the user should always have a simple gesture to hide their implicit interaction with the display.

When the display is used in an urban enviroment, the display needs additional qualities to be successful:

Placement of displays influences what kind of user experience gets created. "Public displays are generally deployed in socially active public environments and as such should be linked to the qualities that make such spaces special" [7]. For example, in a coffee house the public display could be part of the coffee table, letting it perfectly blend into the surrounding architecture as well as engaging users in interaction.

Communal and Shared Information is usually preferred over individual information, for example, customer pictures. All content should relate to the community and be created by the community.

Local and Location Dependent Applications on the public display are to encourage users to socialize.

Needing Nurturing : "Interactivity should be founded on repetitive social actions (e.g. participating in discussion, creating friendship links, collaborating in group activities)" [7].

Reward Systems can create loyal customers [5] and encourage usage in such environments. For example, "rewards in a coffee shop can be either a free coffee or an aesthetically pleasant visualization" [7].

Information presented in such community displays focuses on five types of information:

1. *Relations*: Users are mostly interested in continuing interacting with other users in the community. Information about quantity and quality of the current relations between users helps users to connect.

2. *Co-operations*: Overview of cooperate initiatives available and how users are participating in these initiatives. "e.g. voting for the same song in a vote-based music jukebox" [7].

3. *Nurturing actions*: Actions to maintain their participation level in the community.

4. *Visits*: "Information related to duration and rate of visits to the place, where the community display resides" [7].

5. *General*: Local weather information, local news etc.

4.3 Overcoming Social Barriers

Several studies have shown that users are reluctant to use public displays. It has been observed that a major factor is the fear of social embarrassment. An analogy can be drawn with a street performer, "who invites a participant from the audience to help out with their show. Such a person can often be wary of volunteering, not knowing what exactly will be required from them, especially if it entails making them look foolish in the eyes of the on-looking audience" [6].

Users do believe that they do not have the basic skills to operate such a device and they do not have time to learn how to use the device. Other initial barriers are that there is no established social etiquette how users should behave towards the device.

To study these kind of social interactions, the "Opinonizer" system was used [6]. It is a questioning program, running on a laptop, that could be projected onto a large wall. Users had the possibility to input comments on various changing topics. The system was evaluated at two real separate social events: a book presentation and a student welcoming party.

To overcome these initial barriers, a dedicated helper has to motivate and assist users to operate the device. When more users start to interact and actively use the display, bystanders learn how to operate the device through observing others interacting with the display.

In addition, the helper should actively interact with the device and provide initial input to the system. The resistance to inputing information is then lowered, because users see that other people are already participating[6].

Several observations were made:

Honey-Pot Effect describes the observation, when one user finds and uses the device, it attracts more users. Initially, the display is something new. Over time, more and more users overcome their fear of something new and start to participate. This in turn attracts more new users and the number of people in the vicinity increases. This leads to a physical "bottleneck": users have to determine the social practice of, for example, queuing up to use the input device.

Not knowing what social practice is being applied, leads to negative feelings and creates new barriers, thus discouraging users. Positive conceptions, like believing the display is enjoyable or worthy of attention, entice people to participate.

Besides the increased usage of the public display, the people around the display are open for discussion to talk about the display or the current topics, creating a social buzz around the display.

Low Self-consciousness of users prevents them from using the display. If the public display updates in real time, then all spelling mistakes and shaking with the mouse cursor are viewable by all. Users need a high self-confidence to cope with the feeling that other people are watching as well as the pressure of creating a socially accepted comment.

Remote input would remove social awkwardness, but also remove the "honey-pot" effect. "Remote input would reduce pressure on people, it would defeat the purpose of having a public display as a place for encouraging socializing" [].

A simple way to reduce this type of stress would be to remove real time updating of the screen, allowing users to edit and revise their comment before posting it on the display.

Nicknames: or first names were preferred by people when inserting a comment on the display, even though surrounding people could exactly identify who is inserting the comment. It allows people in the immediate vicinity to create social contact, but is at the same time, "vague enough to prevent social embarrassment and identification from a wider, unknown audience." [19]

Other possibilities to overcome the initial barriers are to increase the attractiveness of the display.

Reward Systems based on performance, for example, learning new aspects of the interface. This can be used with every type of display system. The reward itself should be something physical (e.g., discounts, product coupons).

A random lottery based system gives an incentive to use the system. In addition, it gives certain users a very positive feeling when they actually win the lottery. This type of reward system would be best when used in a bar-type environment [22].

Rewards encourage users to overcome their fear of the new as well as give them an attractive reason to continue using the display.

Usefulness of the display is very important for users so they immediately understand that using the display benefits them.

4.4 Multi-user Interaction at Urban Public Displays

The earlier mentioned CityWall project recorded the interaction of its users via webcam. Several observation were made of people interacting at the display with others [15]:

Parallel Use: Several people could interact with the device simultatiously and work parallel next to each other. Users acted as individuals, disregarding all actions of the other users. A different form of parallel interaction was observed when multiple users synchronized to simultaneously use the same interaction to manipulate displayed photos.

Teamwork and Playful Activities: People who came with friends to the display, clearly acted as a team to work towards the same objective. Depending on the social organization, different group interactions took place. In most cases, one individual approaches and uses the device, while the surrounding group comments and gives advice. Teamwork was also observed to overcome physical limitations. When a user held an item in one hand, another user provided a helping hand to perform gestures. Even though the display was only designed to manipulate photos, groups of users created their own games. "For example, people were playing Pong, throwing photos at each other, and soccer, building a goal out of two photos and trying to throw a third one in." [15]

Conflict Management: Conflicts using the display usually occur when users intrude on the territorial boundaries of other users. Such conflicts usually happen by accident, e.g., by extreme large scaling of an image that covers most of the display and disturbs other users in their work space. One way of resolving this conflict is that the group, that has been intruded on, withdraws from the display. Another way is to first look for support in the group, and then confront the intruding group. In some cases, the conflict was resolved by humoring the situation and making a joke like "It is mine, dont touch" [15].

Floor and Turn-Taking: By observing the actions of others, people can anticipate when it is appropriate to go and take the floor [20]. Such terminal activity has many different forms. One distinctive example is when users, just before their exit, moved towards the side of the screen without making any meaningful actions. No new items are introduced to the screen and the interaction with the existing objects is minimal.

Expressive and Pondering Gestures: Using expressive and grand gestures allow users to clearly signal other users that they are using the display and need a lot of space. On the other hand, people used a pondering grip while thinking about what they would want to do next with the photo. Both gestures signal to other users that they are busy. When other users want to interact with that user, they have to wait for another more suitable moment.

Concluding Actions: When individuals find interesting photos, they use physical and verbal signals to attract attention. Some users leave a mark before leaving the display,for example, leaving an embarrassing photo on the screen.

4.5 Limitations of Mobile Devices

It is obvious that mobile devices could control all applications from a remote distance. For many applications this can be desirable, for example, adjusting the lighting in a room or regulate the room temperature. But in many cases, where the user has to be present to operate the device like microwaves, DVD players, or ATMs, "it is not advisable to use handhelds as interaction devices in order to replace existing physical user interfaces. In most everyday situations, direct manipulation of the appliance is easier, faster, and more convenient than handheld-mediated interaction." [18]

Handheld devices can benefit user interaction when a special situation occurs. Special situations occur rarely, but usually present several problems at the same time: The user lacks the knowledge which individual steps are required to resolve the issues, and the user interface provides no simple interface for these uncommon tasks. Several examples for such tasks would be [18]:

- Programming an oven to start cooking at a user given time. (Usually the functionality is not used because it is to difficult to operate the user interface)
- Figuring out how to use the special programs of washing machines.
- Cryptic error codes of printers (e.g., F602) is incomprehensible and of little value for a user.

Mobile phones have two ways of interacting with the appliances [18].

1. *Information Provision*: The appliance could notify the mobile device about its current status and provide additional information that cannot be displayed on its own display. The user would be provided with a full instruction set what to do next to resolve the problem.
2. *Provide User Interface*: The mobile device could extend the haptic user interface and allow easy access to rarely used features.

When using a mobile device in combination with an public display the same interaction paradigm applies.

1. *Information Provision*: The display could notify users about new personalized information accessible via the screen. Other features could be that the display communicates with the handheld device to identify users and adjust the screen to the users preferences.
2. *Provide User Interface*: When the user can simply walk up to the display and manipulate it via touch, there is no clear benefit to have an additional user interface on an mobile phone.

   In some cases, when the public display is unreachable or cannot be manipulated by touch, creating interactivity via mobile device is an option.

When utilizing mobile phones as input devices one has to face the fragmentation of the cell phone market[9]. Currently there is no way to create a unified application for all devices. For the smartphone market, Adobe Flex 4.5 enables a unified experience for iOS, Blackberry, and Android devices. Even with such initiatives, some operating systems like Windows Phone 7 are currently unsupported [10].

5 Conclusion

The technical design principals found by Vogel et al. are the basis for public displays. Touchscreen technology, found in consumer products, like the Apple iPad, Apple iPhones, and Android devices, is the most promising technology for public displays. "With these, users can now interact directly with the displayed objects by simply touching the display, creating a sense of immediacy and natural interaction." [17]. Users are familiar with the technology and are comfortable using it.

To further enhance public displays, a Microsoft Kinect motion sensor could be used as proximity sensor and hand gesture tracking device. The Microsoft Kinect is a specialized gaming controller to accurately track multiple players. The system has proven to be very reliable in tracking human motions. Medical facilities are experimenting with Kinect controlled medical displays to assist surgents [13].

Combining a touchscreen and a Microsoft Kinect system is relatively cheap and would be an ideal public interaction screen for widespread usage.

At the current state of development of interaction displays, users perceive this type of technology as something new and unusual. The initial reaction of users will be to try to avoid the displays. In the long term, when interactive displays become more common, public interaction displays are going to find social acceptance. To accelerate this process, the public display has to have a solid intuitive design combined with a reward systems.

My research only focuses on Western culture. It would be very interesting to see if similar usage barriers of public displays exist in other cultures, e.g. the Japanese culture or the Arabic culture. With touch devices and proximity sensors becoming more common, new developments for additional ways of naturally interacting with public displays, evolvement of "easy-to-use" interfaces, and also new ways of mobile phone interaction are to be expected. Most interesting will be how these types of devices will integrate into our daily lives.

References

[1] B. C. Analysis. Digital Signage Revenue to Approach $4.5 Billion in 2016, 2011.

[2] G. Barr. Helsinki's "City Wall" -A Collaborative Social Space.

[3] S. Boring, D. Baur, A. Butz, S. Gustafson, and P. Baudisch. Touch Projector : Mobile Interaction through Video. *CHI*, 2010.

[4] S. Boring, S. Gehring, A. Wiethoff, M. Blöckner, J. Schöning, and A. Butz. Multi-User Interaction on Media Facades through Live Video on Mobile Devices. *German Research*, CHI(1):3–6, 2011.

[5] L. Brien and C. OJones. Do Rewards Really Create Loyalty? *Harvard Business Review*, 1995.

[6] H. Brignull, Y. Rogers, and B. Bnqh. Enticing People to Interact with Large Public Displays in Public Spaces. *Displays*, 2002.

[7] R. Calderon, S. Fels, R. Lea, and O. Neumann. Harvesting communal interaction with public displays through place-dependent " community displays ". *Displays*, pages 3–6, 2011.

[8] A. Churchill, Elizabeth, Girgensohn, Andreas, Nelson, Les, Lee. Blending Digital and Physical Spaces for Ubiquitous Community Participation. *Communications of the ACM*, 47(2):39–44, 2004.

[9] K. Gadhavi, Bimal, Shah. *Analysis of the emerging android market*. PhD thesis, San José State University, 2010.

[10] P. Goel. Build Mobile Apps for Android Devices, BlackBerry PlayBook, iPhone and iPad Today!

[11] E. M. Huang, A. Koster, and J. Borchers. Overcoming Assumptions and Uncovering Practices : When Does the Public Really Look at Public Displays ? *Pervasive*, 2008.

[12] G. Kortuem. Mixed initiative interaction: A model for bringing public and personal artefacts togehter. *UbiComp*, 2003.

[13] P. Loriggio. Toronto doctors try Microsofts Kinect in OR.

[14] L. Palen. Social, individual and technological issues for groupware calendar systems. *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*, CHI(May):17–24, 1999.

[15] P. Peltonen, E. Kurvinen, A. Salovaara, G. Jacucci, T. Ilmonen, J. Evans, A. Oulasvirta, and P. Saarikko. " It ' s Mine , Don ' t Touch !": Interactions at a Large Multi-Touch Display in a City Centre. *CHI*, 2008.

[16] C. Pinhanez. The Everywhere Displays Projector : Device to Create Ubiquitous Graphical Interfaces. *IBM Research*, 2002.

[17] A. Quigley, S. Subramanian, and S. Izadi. Special issue on interaction with coupled and public displays. *Personal and Ubiquitous Computing*, 13(8):549–550, July 2009.

[18] C. Roduner. The Mobile Phone as a Universal Interaction Device Are There Limits ? *Science*, 2006.

[19] E. Rubegni, N. Memarovic, and M. Langheinrich. Talking to Strangers : Using Large Public Displays to Facilitate Social Interaction. *HCII*, 2011.

[20] G. Sacks, H., Schegloff, E.A. and Jefferson. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language 50*, 50(4):696–735, 1974.

[21] C. Sas and A. Dix. Designing and evaluating mobile phone-based interaction with public displays. *Proceeding of the twenty-sixth annual CHI conference extended abstracts on Human factors in computing systems - CHI '08*, page 3941, 2008.

[22] J. Scheible and T. Ojala. Mobile group interaction with interactive video on large public display. *ACM SIGGRAPH 2005 Posters on - SIGGRAPH '05*, page 18, 2005.

[23] N. Sukaviriya, M. Podlaseck, R. Kjeldsen, A. Levas, G. Pingali, C. Pinhanez, I. B. M. Tj, and S. Drive. Embedding Interactions in a Retail Store Environment : The Design and Lessons Learned. *Technology*, 2002.

[24] D. S. Tan and M. Czerwinski. Information Voyeurism : Social Impact of Physically Large Displays on Information Privacy. *New Horizons*, pages 748–749, 2003.

[25] S. Tani, Masayuki and Yamaashi, Kimiya and Tanikoshi, Koichiro and Futakawa, Masayasu and Tanifuji. Object-oriented video: interaction with real-world objects through live video. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 593—-598, Monterey, California, United States, 1992. ACM.

[26] D. Vogel and R. Balakrishnan. Interactive public ambient displays. *UIST*, page 137, 2004.

[27] D. Vogel and R. Balakrishnan. Interactive Public Ambient Displays : Transitioning from Implicit to Explicit , Public to Personal , Interaction with Multiple Users. *UIST*, 2004.

Collaborative Work on Interactive Surfaces in Household and Office Environments

Corinna Lins

Abstract— In the ubiquitous computing era, surfaces become interactive. This paper gives an overview of different collaborative working possibilities on interactive surfaces in household and in the office environments as well. Not every surface is adequate for collaborative interaction. There are a lot of approaches in form of applications and independent projects, which are presenting different ways of collaboration. In household environments we can control our devices in the house and also collaborate with our friends and family by doing leisure activities on a interactive surface. In office environments the collaborative work is not the same in the different areas of work. In consultancy the customer and advisor are visualizing the information together, in medical education the users are taking different roles, and for promoting the creativity users have to make complementing inputs.

Index Terms—collaborative work, interactive displays, household environment, office environment, creativity, consultancy, medical education, decision making

1 Introduction

The goal of this paper is to give an overview of the possibilities of collaborative work on interactive surfaces in household as well as in office environments. In the history of human computing are distinguished three eras of development trend. The first one, named Mainframe era, begun in 1950. The idea was, that lots of people used one Mainframe. In 1975 begun the Personal Computer era, where everyone had his own PC. Since 2000 we are in the Ubiquitous computing era, where one user uses more devices. Weiser [31] defines ubiquitous computing as the opposite of the virtual reality. "Ubiquitous computing forces the computer to live out here in the world with people" in contrast to virtual reality which "puts people inside a computer-generated world".
Nowadays we are surrounded by lots of interactive devices and interactive surfaces became more important in our daily life. At home, in educational institutions, in companies, in public domains, on mobile devices like cell phone or music players and other locations, interactive surfaces are more and more common. This paper approaches only interactive surfaces in household and office environments and has the main focus on collaborative work possibilities on displays in such environments.

The first part introduces the term of interactive surface. Due to the fact that not every interactive surface is adequate for collaborative work, that ones which are convenient are watched closer. In the main part of the paper, are presented collaborative applications and projects in the domain of the two analyzed environments. In the office Environments the projects and applications are divided in two parts. The first one relates about software solutions which are designed to support the collaboration on multi-touch displays. The second part is presenting projects which are providing the collaboration on interactive surfaces. This surfaces are not obligatory multi-touch displays and are supporting different types of collaboration.

Collaborative work is "the act of people working together toward common goals"[13]. The collaboration between people is very important, because in this way can arise new ideas, ideas for improvement, creative approaches and a lot of other advantages which are supporting better collaboration.

- *Corinna Lins is studying Media Informatics at the University of Munich, Germany, E-mail: Corinna.Lins@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

2 Interactive Surfaces

Collaborative work on interactive surfaces is significant because people can work together from different locations or they can do different tasks on the same surface.
Now raises the question: What is a interactive surface? The term "interactive surface" is not really defined and the word "interactive" can be interpreted in different ways. A surface is interactive when the user can interact with it. This can happen with the aid of a mouse, joystick or other input device. More recent ideas, define "interactive surfaces" as a surface which is responding to direct input interactions. This surface could be big or small, hard or soft, vertical or horizontal, planar or non-planar. In this paper, the term "interactive surface" defines every surface that reacts on direct interactions.
Interactive surfaces, also named tabletops can support single-touch or multi-touch interaction. The single-touch interaction does not sustain collaborative work because it allows only one user to interact with the display. The multi-touch allows the users to interact simultaneously with the display and the actions are passing in real time. For this reason the most applications that support collaborative work are designed for multi-touch displays.

2.1 Adequate Interactive surfaces for collaborative work

As mentioned above, there are planar and non-planar interactive surfaces. The planar surfaces are flat and are the most common. The non-planar surfaces have different shapes which are not flat and are not so popular because they aren't yet accepted by the majority and the technical implementation is more complex.

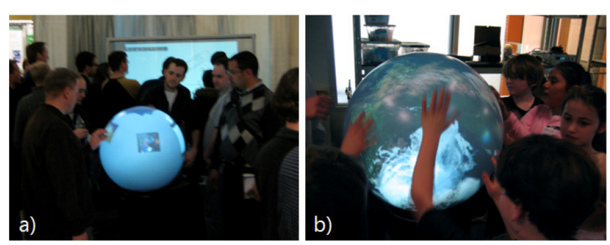


Fig. 1: Collaboration on Sphere in high-traffic locations: (a) 5 adults browsing videos; (b) 7 children interacting with a globe. [11]

The Sphere [11] is a project that implements a prototype of a spherical display which was designed to facilitate the collaborative interaction on the display (see figure 1). The advantage of this curved surface is, that there is no master user and so it offers an egalitarian user experience for each person. Other advantages are that each user has the view to another portion of the display and the surface has no borders.

Another surface with non-planar shape is the i-m-Tube [19], a tubular display designed for artistic and educational purposes. As we can see, the non-planar surfaces have interesting and maybe useful approaches but until the acceptance does not raise, collaborative work is done on the planar surfaces.
There are a few well-known interactive surfaces, the most one from the industry, which are adequate for collaborative work applications.
The most popular multi-touch display, which is also commercially available, is the Microsoft Surface [4]. The display has a rectangular shape and has a horizontal orientation, the ideal assumptions to enable a truly collaborative computing experience [5]. There are already two generations of Surface on the marketplace: The Surface 1 (see figure 2) was launched in 2008 and the Surface 2.0 (see figure 3) was launched in January 2011. This multi-touch display can respond to many touches at a time, this means it is also multi-user able. Microsoft has already 120 partners, which are developing applications for the interface of this tabletop [6]. The Surface is also shipped to the partners with a portfolio of basic applications which can be customized. There are a lot of domains like restaurants, casinos, educational units, retail, and banks where the Surface is deployed.

Fig. 2: Microsoft Surface 1 [4]

Fig. 3: Microsoft Surface 2 [4]

Other well-known systems are the DiamondTouch and the SmartSkin. Both are using capacitive input to enable multiple users to work together on a shared surface.
The Diamond Touch [14] appeared in 2001 and was first sold in 2006. Diamond Touch was the first system which was able to recognize and distinguish between different users. It was developed to facilitate approaches like face-to-face collaboration or brainstorming. [3]
The SmartSkin [24] was made public in 2002, but it is not available on the market place. The SessionDesk [7] is also a multi-touch display, often used to develop collaborative applications on it.
The iPad [1] could also be a good device for collaboration, due to its portability and size. But this paper only expands on large size displays, which are not personal.
A lot of interactive surfaces are giving to the developers the opportunity to choose that one, which is convenient for their concept or to design a new interactive display which has to be customized to the requirements.

3 Applications and Projects Designed for the Collaborative Work on Interactive Surfaces

To analyze collaborative work on interactive displays, there were designed applications and projects in many different fields. This section elucidates various applications facilitating the collaboration in office and household environments on interactive surfaces. Furthermore there are illustrated individual projects which are using their own interactive display and are analyzing proper applications on it. In the next step following issues are addressed: firstly, collaborative work in private environments and after that, in office environments.

3.1 Household Environment

In household environments, interactive surfaces are not so common like in other areas of use. But displays in our household can make our everyday activities easier and funnier. Every device in the house could be controlled on the display, for example the washing machine, the air conditioning system, the TV, the audio system or even the lighting in the house. It would be also interesting to share our photos or videos with friends and family. Advantages of using such a device are:

- we can control devices in the house for the timeframe we are not at home
- automated completion of definite tasks (e.g. switching on the coffee machine and the audio system at a certain hour) is simple done
- social interaction with other persons

Imagine that you are sitting on the sofa with your friends. You want to share some memories from your last holiday. After that, you want to watch a movie. It is cold and you turn up the heater and it is too bright, so you want to dime the light down. Finally, you and your friends want to choose a suitable movie. All this tasks could be done by one device. *Remotable* [10], *Cristal* [27] and *TViews* [20] are living room coffee tables which are able to control household devices and are also supporting collaboration.

The **Remotable** project was presented in 2007. This coffee table has a built-in media center, which enables users to handle different media contents. The device has a touch-sensitive surface and when it is not active, it looks like a common coffee table.
In contrast to the Remotable, the **Cristal** uses a DiamondTouch multi-touch display for the table surface. The goal of this project was to control all electronic devices in one room, as well as to encourage social interaction with friends and family. Concerning the control of devices, there are a few objects which are possible to control: light sources, TV, music player or the movement and position of the robotic vacuum cleaner. The GUI of the coffee table display is a live video image of the room (see figure 4). The video tracking is done by using a ceiling mounted camera. The video image shows the entire living room with all its devices that can be controlled. By controlling them, the user gets direct feedback. When the light is turned off, the video ambient light will be darker. Slider widgets are also necessary to control other devices which don't give visual feedback. A robotic vacuum cleaner can be controlled by sketching paths to be followed.
With regard to the usability, the user study conducted to evaluate the Cristal revealed that the table was too large for being used as a interactive coffee table.

Fig. 4: Cristal: A Collaborative Home Media and Device Controller Based on a Multi-touch Display [27]

TViews [20] is a also a coffee table, but in contrast to other projects the interaction with the display is done with *tangible objects*. This approach has the benefit that the objects can be customized to the used application. TViews allows multi-user interaction through an extensible set of tagged tangible objects. Over ten types of applications were developed to analyze the TViews and four of these were closer

scrutinized: a picture sorter, a map browser to organize images on a geographical map and two games. Figure 5 shows two participants playing a game on the TViews by using the tangibles as pucks.
The table could be used as a everyday coffee table by placing objects on it or eating while playing a game on the tabletop. Some everyday tasks which are done on the PC, could be done on the TViews. Because in the household, designed modernist furniture is wanted, the tangibles of the coffee table could be personalized to individual demands.

Fig. 5: TViews: Users playing a game [20]

There are existing a lot of control systems on the marketplace. This systems were designed to control many different environment sectors. Unfortunately, the most of them does not support collaborative work.
One example of control system in the household environment is the **HouseGenie** [29] project, an Smartphone application presented in 2010. HouseGenie is a universal monitor and controller of networked devices through touch-screen phone in smart home environment. Similar to Remotable, the HouseGenie lets the user view the entire controlled room. Instead of a video image, the user can see a 2D panoramic view. The application also integrates speech and handwriting recognition. All this features could be a good starting position for a collaborative application.

A new possibility of interaction with a surface in the household environment is the proxemic interaction. The following approach is based on the importance of spatial relationship of people or devices to other objects in our environment.
Published in 2010, the project **Proxemic Interaction** [9] uses a home media player application to illustrate the possibilities of proxemic interactions in different environments. The system recognizes the user and customizes its activities to those of the user (e.g. The user watches a movie, the phone is calling, the user answers and the tv is set on pause during the phone conversation). Figure 6 shows a person in four different positions to the display.

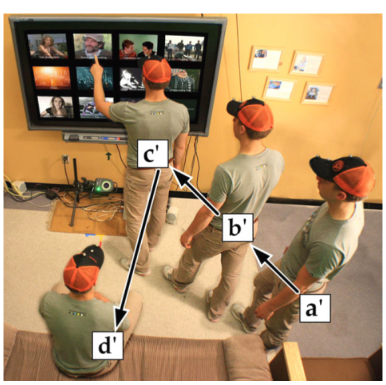


Fig. 6: Proxemic Interaction: User-position to the display [9]

In figure 7 is displayed what the user would see at the four distances illustrated in figure 6 as follows [9]:

a. activating the system when a person enters the room,

b. continuously revealing of more content with decreasing distance of the person to the display,

c. allowing explicit interaction through direct touch when person is in close distance, and

d. implicitly switching to full screen view when person is taking a seat.

Like every new approach, the proxemic interaction also raises difficulties. How to react when two people converse while watching a movie? To pause it or to turn down the volume. How to identify different users to personalize data? How to save user activities in a history? How reacts the system if there are multiple persons doing different activities in a room? All this questions have solutions. The main problem is if the solutions are corresponding to the expectations, or rather how to configure "the rules of behavior".

Fig. 7: Proxemic Interaction: proxemic media player visualisation [9]

It is assumed that even if almost everyone of us possesses a device with a touch display, the acceptance of new technology in household is not so advanced.

3.2 Office Environment

In our days, more and more companies are focusing on collaborative work and for this purpose proven collaboration techniques are needed. Working in groups is very efficient and the results are always proficient because the participants can complete their ideas each other. By designing collaborative applications in different office and working environments, must be taken into consideration the fact that not every shape, orientation and size of the interactive surface is suitable for active collaboration. Depending on the amount of users, type of use and the interaction possibilities a suitable display must be chosen. The most common shapes of such a surface is the flat and horizontal one. It seems for us the most natural, because the collaboration takes place around the table.
In the most diverse working sectors were developed many applications and projects. This section presents firstly some applications designed for the collaboration on *multi-touch* displays and second are introduced some examples from the multitude of different projects drawn up for collaboration on surfaces which are able to be interactive.

3.2.1 Software solutions on multi-touch surfaces

In the early years of the UbiComp era, the project **UbiTable** [28] was developed to support the face-to-face collaboration. This project should enable *spontaneous and unplanned collaboration* where the participants share contents from their mobile devices like laptops or PDAs. The content sharing is divided in three privacy and visibility access possibilities: public, private, and personal. Private data is not visible or accessible to others, personal data is visible but not electronically accessible by others and public data allows shared

visibility and access. Figure 8 shows the separation between the personal and public areas on the GUI of the UbiTable. The project was developed on a DiamondTouch display, which can identify every single user. The goal of this project was to ”provide support for impromptu chance encounters where people need to collaborate on-the-go without prior preparation” [28].
In office environments the employees often encounter unexpected occurrences and spontaneous collaboration is the best solution to handle such occurrences. It is also important to have a delimitation between private and public data, because in a company are a lot of confidential information which are not available for every employee. Such information are an example of private data in a company.

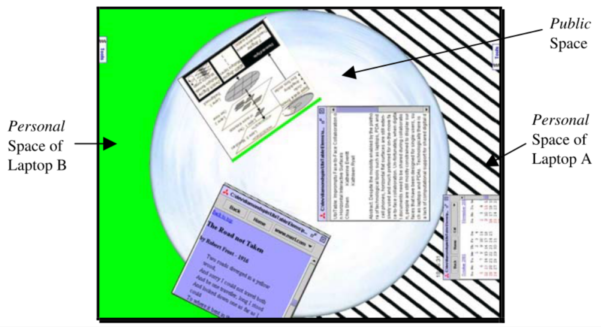


Fig. 8: Screenshot of the UbiTable [28]

WeSearch [21], published in 2010, is a collaborative web search project intended for education and workspace environments. The application is suitable for groups of up to four users which are co-located. As well as UbiTable, the WeSearch GUI has a private area for the web search. Figure 9 shows the user interface. Each group member has a color-coded toolbar in which they can enter queries or URLs, and a marquee containing awareness information. The WeSearch allows following functions [21]:

- Web pages can be divided in multiple smaller chunks by holding the button ”clips”
- Clips can be organized within ”containers”
- The records of a session can be exported at the end

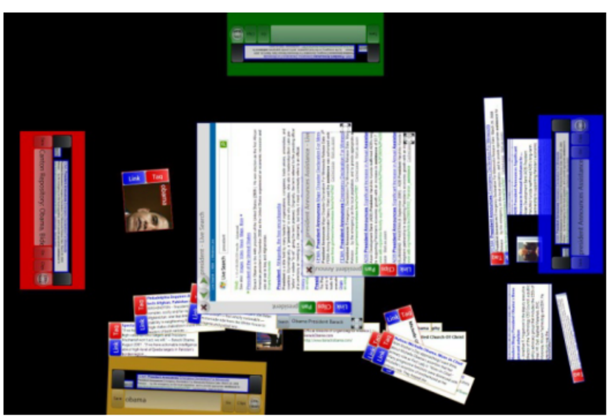


Fig. 9: A WeSearch session [21]

In the *consultancy sector*, a good collaboration with the customer can be the key to success. Using a interactive surface for consultancy, is advantageous because both positions can interact simultaneous with that surface. The customer can see every step in the advisory activity and can be better convinced to choose the offered service.

The **SmartTravel** [22] is an application developed for tourist agencies. The customer itself, can explore active for details of the desired holiday. The visualization is made on a flat vertical display, because in this way all the participants have direct access to the information. The surface orientation is dependent on the number of users. Due to the fact that in consultancy every customer is counseled individually and the number of participants is small, vertical displays are adequate. Figure 10 shows a collaborative interaction between a travel agent and a customer.

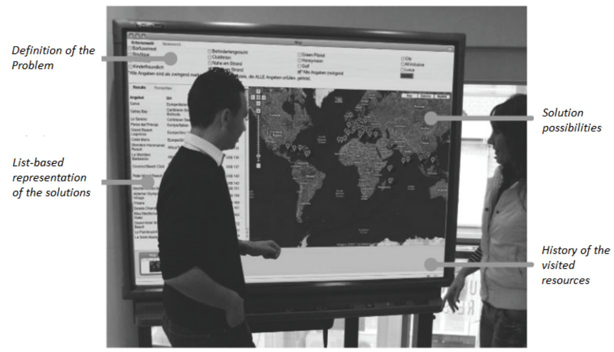


Fig. 10: SmartTravel. Display for cooperative travel planning [22]

Another advisory application is the **Financial advisory support of the Sparkasse Bank** [8]. The application runs on a Microsoft Surface display and has the goal to make adequate proposals to the financial position of the customer. The application functions as follows: The user is identified by its own EC card and its personal data is shown on the display. The advisory cannot start before the consultant validates the advisory process with his card. After authentication, the system proves which product the user already has and makes proposals of new products adequate for his needs. By selecting a product, the user can interact with the application and determine by himself the convenient option. Confirmations by signature are made on the display and compared with the stored original signature.
This were only two consultancy application examples of the multitude of such applications.

In *medical education* collaborative working on a display could be also very helpful mostly for students, but also for health professionals. During the university studies the medicine students must learn theoretical bases as well as practical bases. How else could students learn better to practice the medicine as on a virtual patient.

The **SimMed** [18] project is a game-based learning environment, running on the SessionDesk multi-touch display. The goal of this game-based application is to train medical skills in a team. The users can diagnose and cure the virtual patient. They also should learn how to interact with a patient and not with the simulation on the virtual patient.

Another example in the medical education, available on the marketplace, is the **Sectra**[2] visualization table. Sectra enables collaboration and interaction with 3D images of the patient, which are rendered from a CT or MR scan. Virtual autopsies can be made by the physicians and has the advantage that unlike a physical autopsy it does not alter evidence. Figure 11 shows a collaborative interaction on the Sectra.
In the healthcare sector collaborative working researches are expanding continuous. This is a working area where the interaction with real objects can be simulated and be replaced. For example interaction with patients, biological investigations, surgery simulation and preparation. Because touch interactions does not deliver haptic feedback, the introduction of haptic objects could let the interactions seem more realistic.

Fig. 11: Sectra visualisation table [18]

Collaborative working on interactive tabletops can be prolific by using applications for *promoting the creativity*. The most projects of this type are designed for interaction on a horizontal surface, because users interact frequent with the display.
A project of this type is the **Pictionaire** [16], with the goal to enhance creative collaboration across physical and digital artifacts. Pictionaire supports co-located group design sessions though interaction techniques for searching and tagging, physical-to-digital transitions, digital-to-physical transitions, remote highlighting, and image organization. The application allows users to trace images from the tabletop into sketchbooks like in figure 12 or digitalize paper drawings. Additionally to the multi-touch tabletop, a wireless keyboard is used for user input and the image transfer can be made with a mouse or direct touch input.

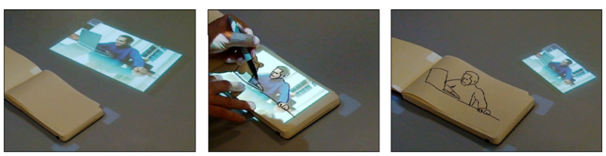


Fig. 12: Using Pictionaire to trace images in sketchbooks [16]

The collaborative tabletop interface **WordPlay** [17], generates, organizes and explores ideas. The surface of the FTIR display has a cubed shape and was built especially for this application. The ideas can be added by speaking, manipulate the properties of words with hand gestures, and explore related concepts by tapping them. The application supports two types of meetings: the idea generation or the decision between a set of alternatives. The idea input is made by speaking into a microphone or writing the input on the multi-touch keyboard. WordPlay is one of many "design possibilities at the intersection of social conversation spaces and multi-touch computing platforms" [17].

In working environments, mostly in companies, the support of *decision making* plays an important role.
A project that focuses intensely on this fact, is a **naval planning support application** published in 2010 [25]. The application is running on a pen-based horizontal interactive surface and incorporates the tracking of maritime vessels. The initial point of the project was to explore different possibilities of application fields in naval planning. The goal of the concept is to have a basic map display system, the possibility to show and edit ship tracks, and top support data input from an arbitrary data source. The actual prototype enables only the collaborative exploration of a dynamic maritime tactical picture and of related information sources. Some functionalities of this application are:

- The geospatial data content is provided in individual windows.
- Users with lower authority have less access to confidential information.
- The system should provide access to dynamically updated, map-based data sources.
- Information and data windows are flexible and adjustable to enable the use from any side of the table(see figure 13).

Fig. 13: Naval Planning Support Application [25]

3.2.2 Projects providing collaboration on interactive surfaces

Different to other interactions on an interactive surface, is the **collaborative interaction with volumetric displays** [15]. Its unique properties are making this displays suitable for collaborative 3D applications. This type of multi-user display has the advantage that it has a 360 field of view and allows user-interaction from anywhere around the display. An example of use would be to examine a virtual model of an anatomy specimen. There are also a few problems by implementing an application for such a surface. The interface elements must be simultaneous available for all users, the feedback has to be available also for all the users, and simultaneous navigation may not be possible. To enable users to walk around the display, the viewing position and the input devices had to be tracked (see figure 14). The users has to wear hats with reflective markers and the 3D positions of this markers were tracked by six cameras. For the interaction with the display, the participants used a 3D presentation mouse. The application implements following interaction techniques:

- interface control through a 3D radial menu
- the navigation operation is done by one user to avoid possible problems
- markup and manipulation: a highlighting tool to define areas of interest

This project was designed to be used in numerous working areas, especially for studying and analyzing 3D images.

Fig. 14: Collaborative Interaction with Volumetric Displays [15]

ConnectTables [30], published in 2001, is a mobile, networked device which supports collaborative work. The height of the display can be adjusted according to the needs (figure 15). Two pen-based displays can be coupled by moving them close to each other (figure 15b) and allow users to interact with a shared workspace. Information exchange can be done by shuffling it over to the other display (figure 15c). Furthermore each user can have its own, but shared view of the same information (figure 15d). This system allows users a great flexibility in work trough the different possibilities of use, especially on meetings and workshops.

(a) (b) (c) (d)

Fig. 15: ConnecTables: Dynamic Coupling of Displays for the Flexible Creation of Shared Workspaces [30]

The **Everywhere Display** [23] is a portable system which enables the delivery of interactive multimedia content on ordinary objects, like walls, tables or floor. The use of the system is demonstrated on a Microsoft PowerPoint slide presentation. The navigation on the slides is made by touching buttons on the projected image like in figure 16. The touch-interaction can be made by hand, but also by feet on the floor. This system could support various applications in environments like conference rooms, smart homes and offices, manufacturing, or health care. Unfortunately, the system does not support multi-touch interactions, but it could open up entirely perspectives concerning collaborative working due to its portability.

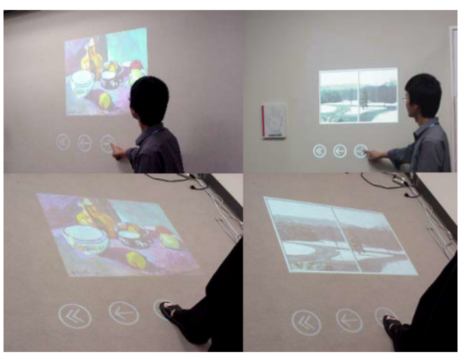


Fig. 16: Example demonstration of moving interactive content across different surfaces [23]

The last project presented in this paper is the **Impromptu** [12], published in 2008. The project presents a *framework for collaboration in multiple display environments*(MDEs). This framework was evaluated in the domain of face-to-face group software development. The collaboration happens by placing information, especially applications on a shared display by the users, to discuss and reflect the focused topics. The advantage of this system is, that it supports any number of shared displays. The user interface of Impromptu is simple and comprehensible and is comprised of three main parts:

- The collaboration control - lets the user configure if the information is shared, showed or hidden (figure 17a)
- A collaborator bar - lists all users with theirs photo and displays for every user which information he has shared or showed (figure 17b)
- The shared screen dock - thumbnails of all opened applications (figure 17c); by moving the cursor over the thumbnails, the applications are opened (figure 17d)

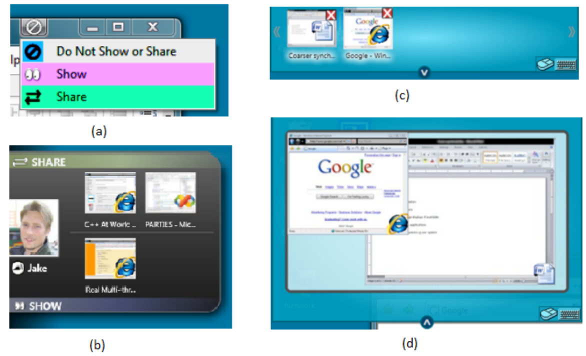


Fig. 17: The IMPROMPTU user interface [12]

A user study which was conducted to evaluate the Impromptu project, has shown that users shared mostly code editors, notes and documents, web browsers, instant messaging or diagrams. The key benefit of this project is that it supports focused problem solving by sharing information with other users.

4 Conclusion

All the presented projects supported co-located collaborative work. Regular desktop computers can support collaborative work but not the co-located one because they support only one input device.
By designing projects and applications for collaborative work, must be considered some skills. Scot et. al. [26] presented design guidelines according to the technology on co-located collaborative work. Some of that guidelines are: the support of interpersonal interaction, switching between activities should be simple, the possibility to switch between individual and group or external work, and support of multi-user interaction. This are only a few guidelines which should be followed to create a successful co-located collaborative work environment.

This paper presented different applications which are supporting collaborative working.
In private environments the most projects are designing a coffee table because it is assumed that the majority collaboration interactions in a household are taking place in the living room, sitting on the sofa. Maybe in a few years, interactive coffee tables will be part of our lives, but until then interactive surfaces in household environments are a delicate topic.
In office environments is the situation quite different. Companies which are focusing on collaborative working, are trying to introduce as fast as possible new technologies based on interactive surfaces. The variety of shape, orientation, size, and interaction possibilities of interactive surfaces is large and can be customized individually for every need. A case study is the Microsoft Surface [6], debuted in 2007. Within four years, the surface became popular all over the world. The number of partners raised very fast and numerous companies, educational institutions, or finance and health area are using it.

It is assumed that the different projects and collaborative applications in the presented environments are expanding our knowledge, our collaboration abilities, and our creativity. In a few years, interactive surfaces will surround our environments and we will be able to benefit of the advantages of collaboration.

REFERENCES

[1] iPad. http://www.apple.com/de/ipad/. visited 04.07.2011.

[2] Sectra Visualisation Table. http://www.sectra.com/medical/visualization/pdf/product-leaflet_Sectra_visualization_table_2.0.pdf. visited 10.06.2011.

[3] DiamondTouch. How it works. http://www.circletwelve.com/products/howitworks.html, 2001. visited 30.05.2011.

[4] Microsoft Surface. http://www.microsoft.com/surface/default.aspx, 2007. visited 30.05.2011.

[5] Microsoft Surface. Fact Sheet. http://www.microsoft.com/presspass/presskits/surfacecomputing/docs/MSSurfaceFS.doc, 2007. visited 09.06.2011.

[6] Microsoft Expands Surface to Europe, Middle East. http://www.pcworld.com/article/160473/microsoft_expands_surface_to_europe_middle_east.html, 2009. visited 09.06.2011.

[7] Session Desk. http://www.archimedes-products.de/session_table/session_desk/, 2009. visited 30.05.2011.

[8] Beratungsunterstuetzung der Zukunft. http://bit.ly/qSzff8, 2010. visited 30.05.2011.

[9] T. Ballendat, N. Marquardt, and S. Greenberg. Proxemic interaction: Designing for a proximity and orientation-aware environment. In *ACM International Conference on Interactive Tabletops and Surfaces*, pages 121–130. ACM, 2010.

[10] J. Beijar, D. Leinerud, R. Nilsson, P. Thorin, and G. Weimar. Remotable: Managing a Built-in Media Center with the Table Top Surface. In *Conference Supplement of Tabletop07*, 2007.

[11] H. Benko, A. Wilson, and R. Balakrishnan. Sphere: multi-touch interactions on a spherical display. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 77–86. ACM, 2008.

[12] J. Biehl, W. Baker, B. Bailey, D. Tan, K. Inkpen, and M. Czerwinski. Impromptu: a new interaction framework for supporting collaboration in multiple display environments and its field evaluation for co-located software development. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 939–948. ACM, 2008.

[13] T. N. B. Centre. What is collaborative working? http://www.nb2bc.co.uk/images/article/pdf47.pdf, 2007. visited 30.05.2011.

[14] P. Dietz and D. Leigh. DiamondTouch: a multi-user touch technology. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 219–226. ACM, 2001.

[15] T. Grossman and R. Balakrishnan. Collaborative interaction with volumetric displays. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 383–392. ACM, 2008.

[16] B. Hartmann, M. Morris, H. Benko, and A. Wilson. Pictionaire: supporting collaborative design work by integrating physical and digital artifacts. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 421–424. ACM, 2010.

[17] S. Hunter and P. Maes. WordPlay: A table-top interface for collaborative brainstorming and decision making. *Proceedings of IEEE Tabletops and Interactive Surfaces*, 2008.

[18] M. Kaschny, S. Buron, U. von Zadow, and K. Sostmann. Medical education on an interactive surface. In *ACM International Conference on Interactive Tabletops and Surfaces*, pages 267–268. ACM, 2010.

[19] J. Lin, Y. Chen, J. Ko, H. Kao, W. Chen, T. Tsai, S. Hsu, and Y. Hung. im-Tube: an interactive multi-resolution tubular display. In *Proceedings of the seventeen ACM international conference on Multimedia*, pages 253–260. ACM, 2009.

[20] A. Mazalek, M. Reynolds, and G. Davenport. TViews: An extensible architecture for multiuser digital media tables. *Computer Graphics and Applications, IEEE*, 26(5):47–55, 2006.

[21] M. Morris, J. Lombardo, and D. Wigdor. Wesearch: supporting collaborative search and sensemaking on a tabletop display. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 401–410. ACM, 2010.

[22] J. Novak, S. Schmidt, M. Aggeler, and G. Schwabe. Kooperative Medienumgebungen fr das Reisebro der Zukunft. *Tagungsband Mensch und Computer*, 2008.

[23] C. Pinhanez, R. Kjeldsen, L. Tang, A. Levas, M. Podlaseck, N. Sukaviriya, and G. Pingali. Creating touch-screens anywhere with interactive projected displays. In *Proceedings of the eleventh ACM international conference on Multimedia*, pages 460–461. ACM, 2003.

[24] J. Rekimoto. SmartSkin: an infrastructure for freehand manipulation on interactive surfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, pages 113–120. ACM, 2002.

[25] S. Scott, A. Allavena, K. Cerar, G. Franck, M. Hazen, T. Shuter, and C. Colliver. Investigating Tabletop Interfaces to Support Collaborative Decision-Making in Maritime Operations, 2010.

[26] S. Scott, K. Grant, and R. Mandryk. System guidelines for co-located, collaborative work on a tabletop display. In *Proceedings of the eighth conference on European Conference on Computer Supported Cooperative Work*, pages 159–178. Kluwer Academic Publishers, 2003.

[27] T. Seifried, M. Haller, S. Scott, F. Perteneder, C. Rendl, D. Sakamoto, and M. Inami. Cristal: A collaborative home media and device controller based on a multi-touch display. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 33–40. ACM, 2009.

[28] C. Shen, K. Everitt, and K. Ryall. UbiTable: Impromptu face-to-face collaboration on horizontal interactive surfaces. In *UbiComp 2003: Ubiquitous Computing*, pages 281–288. Springer, 2003.

[29] Y. Suo, C. Wu, Y. Qin, C. Yu, Y. Zhong, and Y. Shi. HouseGenie: Universal Monitor and Controller of Networked Devices on Touchscreen Phone in Smart Home. In *Ubiquitous Intelligence & Computing and 7th International Conference on Autonomic & Trusted Computing (UIC/ATC), 2010 7th International Conference on*, pages 487–489. IEEE, 2010.

[30] P. Tandler, T. Prante, C. M
"uller-Tomfelde, N. Streitz, and R. Steinmetz. Connectables: dynamic coupling of displays for the flexible creation of shared workspaces. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 11–20. ACM, 2001.

[31] M. Weiser. The Computer for the 21st Century. http://www.ubiq.com/hypertext/weiser/SciAmDraft3.html, 1991. visited 30.05.2011.

Implicit Authentication On Mobile Devices

Tobias Stockinger

Abstract— One major disadvantage of mobile devices is their liability to theft. Since people make more intense use of their smart phones to browse the web and store potentially sensitive data, it is proposed to use additional measures to secure one's device. Implicit authentication for mobile devices is a promising solution to the problem. Analyzing user behavior and biometric characteristics can be used as additional measure to verify the user of a handset. Thus, sensors that are already inside contemporary smart phones are efficiently utilized for an area, that they were not originally designed for. Not only security is increased through implicit authentication techniques, but also usability. Inconvenient password requests by a phone can be reduced to a minimum. However, there are certain challenges in finding solutions to implement continuous, implicit authentication. Among these are the limited battery capacity and computational resources. Smart phones do not offer as much CPU power as cloud computing, so it is suggested to make use of these external resources, that are even dynamically adjustable.

Index Terms—Security, Smartphone, Implicit Authentication, Usability, Mobile Devices, Privacy

1 Introduction

The main purpose of authentication is to ensure that only the rightful owner of a certain device, account or document is granted access to it. In other words, this thought includes sealing the device off from possible adversaries such as thieves, impostors or also curious friends. However, current authentication fails in certain cases where an adversary somehow found out the credentials, e.g. Personal Identification Numbers (PIN) or any password.

One can imagine an everyday scenario: Alice's smart phone was stolen while she was at university. In the auditorium, Bob could have seen her typing in the PIN while he was sitting behind her ("Shoulder Surfing" [30]). As soon as Alice remarks the missing phone, she is likely to take measures to have the SIM card locked and request a remote deactivation of most phone functionality to protect her data (which is possible on newer phones [3]). Until that lock is established, Bob - the thief - has access to sensitive information and can even cause financial as well as reputational damage to Alice.

Although it is harder for an attacker to access and use the phone without knowing the PIN, many people do not secure their devices effectively with this type of authentication: PINs are usually only required at startup. After successfully verifying the identity of the user, the phone is henceforth unlocked. A lot of phones offer to re-authenticate after recovering from stand-by, but users seem to refrain from PINs/Passwords/Passcodes for usability reasons (or even renounce the PIN entirely) [7], [17]. Mobile devices complicate password entry because they have smaller keyboards. As a consequence, users are motivated to opt for simpler, therefore weaker, passwords which do not withstand common attacks [21].

There are a number of reasons, why securing personal data is especially important on mobile devices. For example, if the smart phone's email application is accessible for anyone, one part of the most sensitive data is vulnerable without much effort. Some emails include passwords for specific accounts, where an attacker can then log into. Furthermore, the majority of on-line services only requires a valid email address to reset the password. The impostor can reset many passwords for different accounts entailing a complete exclusion of the legitimate user. This weighs heavy, notably for financial accounts, such as PayPal [22].

In this paper, the principle of *implicit authentication* is presented, which tries to thwart the above-noted scenario and protect the user's data from misuse. Its main idea is to utilize user biometrics or behavioral characteristics to verify the rightful owner of a device in addition to traditional authentication like passwords.

The rest of the paper is structured as follows: *Section 2* states some necessary definitions before a short overview of authentication methods is given in *section 3*. Hereafter, the functional principles of implicit authentication in general (*section 4*) and for smart phones in particular (*section 5*) are explained. A summary of adversary characteristics is presented in *section 6*. The paper concludes by showing the limitations of implicit authentication (*section 7*) and a look into possible future development (*section 8*).

2 Definitions

When talking about authentication, it is helpful to give some basic definitions regarding this matter. The process of *authentication*, or also *verification*, validates a claimed identity by matching it to a known set of identities [5]. In other words authentication answers the question *"Am I who I claim to be?"* [13]. The result of the one-to-one test has a binary output: The answer can be `true` or `false`. However, there are certain degrees, i.e. thresholds, of deciding whether the identity can be confirmed. Mostly, this happens through the calculation of a one-to-one matching *score* (see *section 4.3.3*) [5].

In contrast, *identification* has a different purpose. When *identifying* a person, that person does not his or her identity [5]. Rather, the system has to find out itself who is interacting, through matching certain characteristics of a client to models in the database. This is accomplished in a one-to-many matching process. As a consequence, *identification* searches for an answer to the question *"Who am I?"*[13]. It is then assumed, that the dataset with the highest similarity represents the individual. Thus, identification returns a vector or tuple closest to the person's characteristics instead of binary answers.

3 Authentication Methods

After defining the most important terms, some of the most common features that can be utilized to perform authentication are explained. This section covers the general authentication area and is not necessarily limited on mobile devices, which are presented in *section 5*, but it still provides the basis for understanding the specific conditions for portable devices. Note that *identification* does rely on some of the following cues as well. However, the focus of this paper remains on authentication, so identification is not treated in detail. Also, there is no distinction between implicit and explicit authentication just yet.

3.1 Passcodes

The most common way to authenticate a user is to explicitly ask her for a certain passcode [13]. The term "passcode" includes Personal Indentification Numbers (PINs), alphanumeric passwords and other graphical passwords [30]. This authentication approach relies on the

• *Tobias Stockinger is studying Media Informatics at the University of Munich, Germany, E-mail: stockinger@cip.ifi.lmu.de*

• *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

human brain's capability of remembering a preferably unique combination of numbers, letters or symbols. However it suffers from certain disadvantages:

(1) Even though people can remember a certain amount of different passcodes, the maximum capacity still seems limited. In 2006 it was found that a "heavy" user has an average of 21 passwords to remember [13]. As Gafurov et al. do not further specify the term "heavy" user, it is implied that a normal user is likely to have less than 21 passwords to bear in mind. This is either because the number of accounts is smaller compared to "heavy" users, or because a normal user is inclined to re-use a certain password for multiple accounts. This is aggravated by the fact that people usually are not motivated enough to linger over security issues [29]. Instead, about 81% of the users choose common words as their passwords, which are more susceptible to dictionary attacks [15]. (2) Another problematic issue are malware and other attacking methods. Even the strongest password can be compromised if a computer or mobile device is infected with keyloggers. Observing people while they type in their password is often referred to as shoulder surfing [30].

Despite their shortcomings, passcodes possess a huge advantage in comparison to biometrics: They are changeable. Once a password is compromised, it can easily be reset whereas for example face-, iris- or fingerprint-authentication lack this feature.

3.2 Tokens

Especially in business environments, the use of a special hardware device for a second factor authentication has established [26]. In case of SecurID the hardware token is a rather small device which displays a randomized number that is used as a second-level password [24]. Furthermore, this password is changed every 60 seconds which is supposed to make it even more secure. If used in combination with another portable device, this authentication method shows its strengths and weaknesses: On the one hand it is almost perfectly secure because the session timeout on the device that one tries to secure can be very low or even adapted to the presence of the device. If the token is not in the vicinity of the device, the session can be terminated, making it impossible for an attacker to access sensitive data. On the other hand, the token is as easily stolen or lost as the actual device. Moreover, this method requires a costly, highly evolved wireless network infrastructure to repeatedly send passwords to the token.

3.3 Biometric Cues

As we have seen in the previous paragraphs, passwords and tokens have some disadvantages, so that one should examine other authentication features. This section summarizes the different (physiological) biometric cues that can be used for authentication.

3.3.1 Face and Iris

Person identification through face recognition can be seen as the most intuitive way, because humans themselves are highly dependent on visual cues when it comes to recognizing someone. This approach already delivered working authentication systems in the early 1970s [6]. The general principle hereby is to capture a person's face through a camera. After digitizing the image, the pixels of certain regions of the face are compared to images that have already been acquired. Iris recognition operates a similar way, but only captures a high-resolution image from a person's eyes.

However, just like any other method this authentication scheme shows certain disadvantages: If the person's image isn't taken while in a similar position (e.g. because the camera angle is different), the matching faces difficulties. Moreover, people's faces change over time as people age, so it is suggested to implement an adaptive system.

Algorithms in this field are for example the Eigenfaces approach or the Fisherfaces algorithm [11]. The Fisherface-algorithm which is based on linear discriminant analysis has been found to produce low error rates, compared to the Eigenface approach [4]. Especially in different lighting situations "Fisherfaces" hold the upper hand.

3.3.2 Fingerprint

Fingerprint matching takes place by comparing minutiae (ridges and furrows on the finger), which are acquired using a fingerprint reading sensor [28]. This type of biometric authentication has become one of the most popular and is applicable to smart phones due to the rather small size of the sensor: A few solutions for business customers as well as consumer products already exist, such as the Motorla ATRIX 4G. Like almost any authentication method, it is possible to spoof a fingerprint. However, it takes a lot effort to successfully obtain and forge a person's fingerprint. A picture of a contemporary fingerprint reader on a smart phone is shown in *figure 1*.

Fig. 1. Fingerprint reader on the Motorola ATRIX 4G [20]

3.3.3 Voice / Speech

While speech recognition is to a high extent aimed at recognizing words and phrases independent of the speaker, voice- or speaker-recognition focuses on finding out *who* is speaking [6]. Voice recognition analyzes the acoustic characteristics of a speaker, such as pitch and phonetics. Especially in security applications, it is desirable to challenge a user with different phrases for each verification, so that the system is more robust against prerecorded samples. Voice recognition is more difficult or even impossible, if the user is sick or sore. Furthermore, the voice of a certain person might sound different in the morning than it does in the evening due to stressing during the day, e.g. singing or loud talking [11]. Additional background noise in loud environments might hamper any voice recognition.

3.3.4 Other Biometric Cues

Beside the above presented biometrics, a few other methods can be found in the literature, which shall not be explained in detail here. Handwritten signature comparison is usually done by visual inspection. That means a person usually checks another person's signature. However this authentication method also shows some potential to be exploited for automatic authentication [16]. Other possible authentication methods are based on electroencephalograms (EEG) and electrocardiograms (ECG) [11]. Therein special sensors measure electrical activity on the head (EEG) or around the Heart (ECG). Their major advantage is the fact, that they do not exclude any human being, since everyone has a heartbeat and a brain. However, one sensor does not suffice, which causes usability issues.

3.4 Behavioral Cues

Some papers in the area of biometrics consider behavior *biometric* as well. However, throughout this paper biometrics are seen as pure physiologic attributes. For example, someone can't immediately change his or her fingerprint whereas typing speed or gait can be manipulated on command.

3.4.1 Gait

First approaches of analyzing the individual characteristics of human gait were found in the 1970s [10]. Since then, researchers have tried to automatically identify people from their walk. After finding ways of visual gait recognition, technology has evolved and gait data can now be collected using acceleration sensors [1]. An image of gait patterns, is shown in *figure 2*, where one can see two examples of gait cycles. On the left, the graph rises because the persons carrying the accelerometer were standing still at the beginning. After a short timespan both graphs show characteristic slopes which represent steps or gait cycles. Examining the two different plots, one can notice that the slopes differ in duration, for example. That examination can take place automatically, which is the main idea of gait recognition.

Due to the fact, that modern smart phones have already built-in gyroscopes, this method is applicable for mobile devices without hardware changes. However, one must consider, that there are a lot of things that can alter gait. Just to name a few: Footwear, ground surface, carrying load and injuries are likely to create notable impact on a person's walking style [13]. Furthermore, gait recognition fails if the user does not walk at all, because she has been seated for some time.

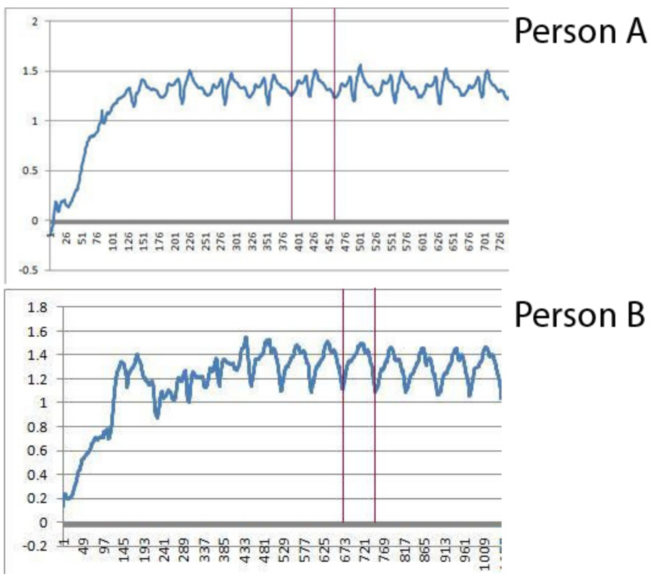


Fig. 2. Gait patterns (combined for x,y,z axes) gained from two different persons [27]

3.4.2 Keystroke

First keystroke authentication algorithms were found in the late 1980s by Card et al. [8]. They measured and analyzed the time span between two keystrokes, which they termed *diagraph*. Ever since then, this approach has been enriched with other factors such as key hold time and error rates [21]. In order to provide more reliable scores, keystroke based authentication usually requires a rather long training phase [31]. It's challenging to apply this method on mobile devices, especially if there is no hardware keyboard but a soft input method, like touchscreens. The hold time is a less meaningful factor and thus less applicable on mobile devices than on normal sized keyboards [18].

3.4.3 Location

"People are creatures of habit" [26], so they usually visit the same places every day like an office at work. Location based authentication exploits this very routine by tracking an individual's whereabouts. This approach is predestined for implicit authentication: It would be outrageous if the mobile device asked its user to move to a certain location in order to verify her, which would indeed become a hide-and-seek like game. However, if the user moves out of her regular action space, the phone might take measures to ensure that it's in the hands of the rightful owner. Evidently, it's necessary to have additional authentication mechanisms, since the user might actually one day go to unfamiliar places. So there are further steps to be taken.

3.4.4 Network Usage and File Access

A fairly new mechanism was proposed in 2009 by Yazij et al. [31]. They investigated if it were possible to identify a user from examining her network activities and file access behavior. So if a certain individual visits a particular website about 6 times per day, and another website about 10 times, one can build a user model for normal behavior. This approach is especially strong if an attacker does not know that there are certain patterns to be followed, in order to stay logged in. However, if the impostor knows what he has to do to mimic the victim, the system fails very soon. Yazij et al. therefore propose to use this kind of authentication in combination with other features.

3.4.5 Other Behavioral Cues

Some people do have specific **call patterns** - they call the same person every other day and are called by only a limited number of contacts on their phone. These call patterns are user specific and can also be exploited to calculate a precise authentication score [17],[26].

To conclude, an overview of the different authentication features is shown in *table 1*.

4 Functionality of Implicit Authentication

In this section, the workings and approaches of implicit authentication is explained. Although there is a separate section dealing with mobile device specific advantages and problems, the paper tries to maintain the focus on mobile devices.

4.1 Implicit vs. Explicit Authentication

Traditionally, a user authenticates herself after the device asks her for a certain proof of identity (see *section 3*). This process is called *explicit authentication*. This type of verification becomes a time-consuming procedure if it is necessary to explicitly authenticate for many different services or accounts. As we have seen in *section 3.1*, heavy users have an average of 21 passwords to remember. Typing in 21 passwords and the according user names appears to be a lot of effort and suffers from major usability issues. It seems comprehensible that users work around this problem by re-using passwords or choosing words they can easily remember. The problem is aggravated if the user is not challenged to enter a password regularly. In that case the session is kept alive for a rather long period, sometimes until the device is turned off or even longer. The Single-Sign-On (**SSO**) paradigm is supported by password managers that reduce the problem of frequent authentication [26]. Password managers therefore increase usability but drastically reduce privacy.

On the one hand, implicit authentication is designed to minimize memory challenges and time consuming verification procedures. On the other hand, continuously verifying the holder of a device brings about an increase in security, such that personal data is not compromised through session hijacking. Thus, implicit authentication might one day be able to replace explicit authentication, but the current technologies do not meet the requirements to do so. As a consequence, current research focuses on enhancing explicit authentication and thus both usability and security by adding implicitly gathered authentication cues.

Finally there are certain scenarios in which implicit authentication acts as a fraud indicator not requiring any user interaction on the actual device [26]. This can be the case when a credit card has been stolen and the thief wants to make an on-line purchase with it. An implicit authentication system could detect that the legitimate owner of the credit card is currently busy. For instance because her or she is making a phone a call in a certain location. Knowing this, the system can then take measures to prevent the purchase with the stolen credit card, because it is not plausible.

This paper mainly focuses on implicit authentication as a usability advantage and as second factor security enhancement. Fraud indication as stated above is not primarily in the scope of this work.

Feature	Capturing Method	Implicit / Explicit	Spoofing Threats	Applicability on Mobile Devices	Problems
Passcode	Hard / Soft keyboard Input	Explicit	Keyloggers, Shoulder Surfing	No constraints	Guessable passwords are still the most used
Token	Hardware Device shows a password, that expires after a short time	Mainly explicit, but implicit authentication through Bluetooth possible	None	Mobile devices are rather small, carrying around two devices seems too much to ask from a user.	Easily stolen or lost
Face & Iris	Camera	Both	Photographs of the legitimate user	Front Camera necessary (for face recognition)	Lighting situation and make-up
Fingerprint	Fingerprint reader	Currently only explicit, implicit authentication difficult	Play-Doh casts or Scotch Tape	No constraints	Injuries on the finger alter the minutiae pattern
Speech	Microphone	Both	Recordings of the user's voice	No constraints	Sickness, natural voice changes, background noise
Gait	Camera or Accelerometer	Both	Gait imitation (difficult)	No constraints	Injuries, carrying load, footwear, ground surface, being seated
Keystroke	Hard/soft keyboard	Primarily implicit, explicit possible	Typing imitation (difficult)	Possible, but difficult	Long training phase, reliability
Location	GPS or infrastructure calculated position	Primarily implicit	Informed strangers	No constraints	Traveling outside the regular scope, precision
Network/File Access	Software protocol (exemplary tool: WireShark)	Implicit	Informed strangers	File access is restricted, since the file paradigm is not too widespread on smart phones. No constraints for network usage	Precision

Table 1. Comparison of different authentication methods

4.2 The Imprinting Paradigm

In order to better understand the idea of implicit authentication, some research papers suggest to regard the relationship of human and computer as a *parent-child-relationship* found in the animal world. For example: " Geese [...] imprint on the first suitable moving object they see shortly after hatching, and will ever after treat it as their parent"[14].

Applying this imprinting paradigm to computers would allow for a more secure handling of the device. The first time the computer or mobile device is used, it imprints on the user relying on as many cues as possible to recognize her. For security reasons, not even the user should be granted access to the storage of the data model that represents her [14]. The hereby established trust relationship is stronger than using a challenge/response authentication scheme, as the user can be assured that the computer will not give away personal data freely.

Imprinting is a process that presumably cannot take place in a single moment. Rather, it is necessary to run through a special training phase, where the imprinting happens. The duration of the training phase is a critical point. On the one hand the system should work as securely as possible, which a longer training phase could ensure. On the other hand the timespan in which the device is still unprotected, respectively *less* protected, should be as small as possible. These two aspects have to be considered when taking a decision concerning the training duration. In the literature one finds suggestions that training should at least take two months to yield a False Rejection Rate (**FRR**) of 2% or less [31], which is an acceptable value in terms of security. Implicit authentication mechanisms are not active during that enrollment.

Furthermore, there exist two types of training phases, analogue to authentication: explicit or implicit training. A user-initiated training phase is regarded problematic, since users are not motivated enough to secure their devices [29]. User studies have shown that explicit enrollment of authentication data, e.g. keystroke dynamics, is a bothering task [7]. One can conclude that people would then refuse to train their devices properly, which would in turn render implicit authentication useless. So it is suggested to perform an implicit training phase, as well. This means that the device does not ask the user to take certain steps, but rather informs her, that data will be collected for a certain time. After that duration, the system becomes active [7]. If calculation and storage resources are sufficient, there is no reason to stop collecting enrollment data after this point, so that the system is strengthened.

If a person other than the legitimate owner uses the device during the training phase, the enrollment data is biased. At this point, it appears no reasonable solution to this problem has been found because research papers in this area do not treat this case at all.

4.3 Algorithms

This section shortly presents some algorithms that are used to perform anomaly detection afterwards. When going through literature, only few papers explicitly describe the functionality of the used algorithm in detail, as some of these are expected to be well known. Rather, concepts are shown, leaving room for further ideas.

4.3.1 Prerequisite: Evaluation and Metrics

In many research papers that performed user studies (with prototypes), the tested algorithms were measured and compared through False Acceptance Rates (**FAR**), False Rejection Rates (**FRR**) and in some cases through Equal Error Rate (**EER**). FAR refers to the number of times an impostor is mistaken for the legitimate user, while FRR tells how often a legitimate user has not been recognized. Both measures are expressed through a percentage relative to a total number of impostor or genuine authentication attempts. Thus, FAR and FRR are calculated as follows [13]:

$$FAR = \frac{N_{accepted\_impostors}}{N_{total\_impostors}}; FRR = \frac{N_{rejected\_genuines}}{N_{total\_genuines}}$$

In general one can state that "FAR relates to the security of the system, while FRR to the usability" [13]. Balancing those two rates while still minimizing both is the primary goal of every authentication technique. That balance is expressed through the EER, that represents the state where FAR equals FRR [13].

4.3.2 Anomaly Detection

Given a user model and certain features (see *section 4.3.3*), the core functionality of implicit authentication are classification- and scoring algorithms, which are a sub-area of data mining. They either compare a sample to the user model or calculate a probability score.

Neural networks. When the result should be a classification (legitimate user / impostor) the use of neural network algorithms is suggested, because the yielded FARs and FRRs are usually quite low, thus more accurate [7]. They take an input vector containing the measured feature data and try to find known patterns in the user model, resulting in classified data. On the downside, the complexity of neural network algorithms is rather high especially when applied for large data sets, so they require higher CPU performance. As a consequence, this sort of classification technique is less applicable on mobile devices themselves, but potentially suitable in a powerful cloud.

Statistical and Heuristic Methods. More often, statistical algorithms are used for anomaly detection. Their major advantage are low processing requirements while maintaining acceptable results [31],[7]. They perform very fast and energy efficient even for large data sets. Exemplary algorithms in this area are K-Means Cluster, FFT, correlation and Bayesian networks - each one has its advantages in different areas. For example, gait recognition is especially suitable for correlation (see *figure 3*).

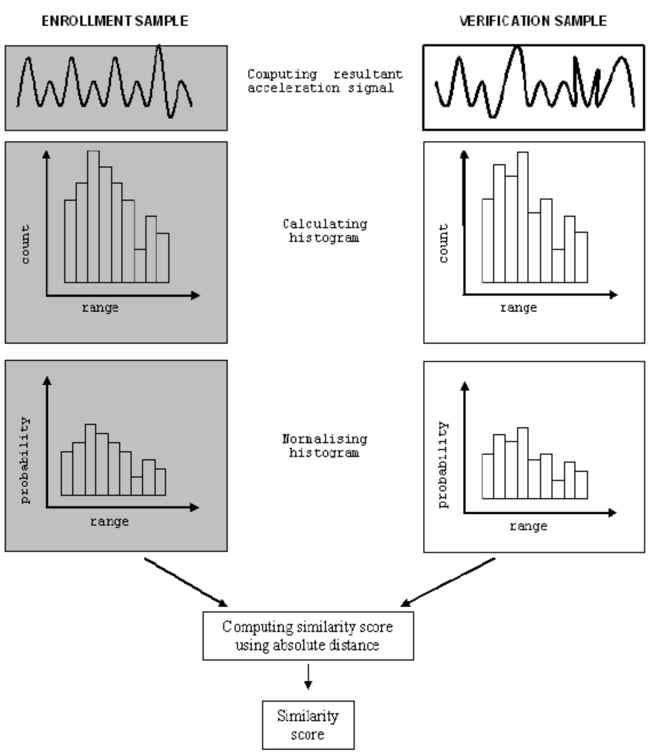


Fig. 3. The process of applying the histogram similarity method [13]

4.3.3 User Models and Scores

In order to make an authentication decision, most approaches utilize a **scoring function**. This function usually requires a previously generated **user model**, which represents the rightful user of the device. When combining multiple features (which is recommended, see *section 5.1.2*), each one of these features has to be rated individually by the scoring algorithm. Since some biometrics are more reliable than others, some algorithms weight the resulting scores, based on the feasibility and precision of the measured cues. After that, a final score is generated that is then used to either reject or accept the authentication.

4.3.4 Exemplary Scoring Functions

Jakobsson et al. [17] as well as Shi et al. [26] propose the introduction of "good" and "bad" events that influence the current authentication score. In their approaches, the user's recent behavior is described by a tuple $(t, v_1, v_2, ..., v_k)$ where t is the current time and $v_1, ..., v_k$ denote the values of independent feature-variables at time t. After computing a separate score for each feature, a function f is then used to combine these scores into a final score. Additionally, a probability density distribution is included into the scoring, meaning the score has to decrease over time. "Good calls" are made to or received from known numbers, e.g. from the contact list stored on the device. Contrarily, "bad calls" are all the others, i.e. made to unknown numbers.

Let $V_1 =$ *time elapsed since last good call*, then one has to adapt the probability according to the daytime, for example. As a consequence, for someone who usually makes phone calls in the afternoon, but rarely in the morning, the score is expected to **decrease faster** in the afternoon, because phone calls during this time are more probable. This allows for a faster anomaly detection.

4.3.5 Exemplary Algorithm

If one wants to summarize how user models are built and at what point, the score calculations take place, it is helpful to illustrate the procedure with a pseudo-code algorithm. An example of such an algorithm based on work by Yazij et al. is presented in *algorithm 1* [31]:

Algorithm 1 General User Model Building and Implicit Authentication Algorithm according to [31]

```
For each user U_n do the following:
while TRUE do
    Receive data from capturing system
    if NewUser = TRUE then
        if time < TrainingDuration then
            Add log data to the data source
        else
            Build the user's profile
            NewUser = FALSE
            Send user's profile to the cloud
        end if
    else
        {user model already exists}
        Perform anomaly detection
        if abnormal behaviour then
            Take action, like re-authentication through PIN, lock de-
            vice, or warn the user
        else
            continue
        end if
    end if
end while
```

5 Mobile-Device-Specific Implicit Authentication

We have seen the advantages and workings of implicit authentication in the previous chapter. These findings are now substantiated in the following sections.

5.1 Smartphones

This section is about further unique conditions of implicit authentication for mobile devices and for smart phones in particular.

5.1.1 Suitable features

As one can see in *table 1* certain features/cues are more suitable for implicit authentication on mobile devices than others. This is because mobile devices usually have less capabilities of capturing behavior through sensors and software. However, the literature proposes a few cues that have already been evaluated or will be tried out in the future. The three major approaches so far were **gait recognition**([13],[27],[19]), **keystroke analysis**([7],[21]) and **behavior patterns**([26],[17],[9]).

First approaches relying on (mobile) gait recognition made use of an external accelerometer attached to a specific body part, e.g. to the lower leg [27]. They are now replaced by an internal gyroscope which many current smart phones possess. Keystroke recognition has mainly been evaluated on phones with hardware keyboards - touch-screen based phones are not entirely excluded, because they offer other usable characteristics: the size of the area produced by finger pressure on the screen gives hints on who is using the phone [25]. The mostly suggested behavioral patterns were user actions (e.g. phone calls or web-sites) in combination with location information. Location based continuous authentication is predestined for mobile devices.

5.1.2 Combining Multiple Features

Since single cues have rather large error rates (cf. *section 4.3.1*) it is reasonable to fuse many features to calculate an authentication score. For example, if a device collected information from 20 different cues, with independent error rates of 20%, a 2/3 vote has only a theoretical chance of 1:500,000 of taking a false decision [14]. Similar results have been presented by Yazij et al.: Their network, respectiveley filesystem, based approach had maximal FARs of 65%, respectively 94% - which are unacceptable values [31].

It is also important to have fall-back options when one feature is currently not available at all.

5.1.3 Further Considerations

At this point, there is no "real" solution or software for implicit authentication on mobile devices, that one can buy and install on one's device, which stands in contrast to explicit authentication that has a lot of different solutions to offer. However, a stable implementation might not be too far ahead, since the algorithms and techniques already exist. Several prototypes have been evaluated in exemplary studies [9].

One challenge that has to be taken are hardware constraints: More devices would profit from implicit authentication if the functionalities are realized without having to add hardware to the handset. The suggested Trusted Platform Modules are not available in every device - so this approach seems less promising, although the security would increase.

5.2 Local vs. Remote Calculation

Basically, the scoring calculation can take place on the device itself or on a more powerful remote system. Due to the limited calculating capacity on mobile devices (see *section 7*), it is recommended by the majority of recent research papers to make use of *Remote Attestation* [21]. According to this approach, log data and variable states are repeatedly packed and then transmitted to a remote server. Naturally a nearly permanent mobile Internet connection is required, because data is transmitted in short time intervals. Since this server possess higher capabilities, the calculation of the authentication score is performed quickly. The result containing the decision if the session ought to be quit or not is immediately sent to the phone, which can take the according measures.

It has to be noted, that all data transmitted through the network is highly sensitive. Thus, it is mandatory to establish trusted connections through encrypted channels [21]. Furthermore, suggested solutions also demand the above mentioned *Trusted Platform Module* (*TPM*) in order to ensure that the endpoints of the communications have not been compromised and the data is valid.

5.3 Cloud Based Implicit Authentication for Mobile Devices

A recent development enhancing the principle of remote attestation are cloud based authentication frameworks. One of the reasons for the increased demand of cloud computing could be the ongoing sales of smart phones. These devices offer access to information that is spread across the Internet. It is necessary to store data externally because typical handset storage capacity is rather small compared to desktop PCs. In order to enable complex calculations and applications, mobile devices outsource difficult tasks. Consider Anti-virus software for smart phones: the analysis of a file can be very intensive and is likely to use up a lot of battery, as well as CPU capacity. However, it would be preferred if this was accomplished through a background process, which ought not to slow down the phone. In this scenario an external service, which often is realized using cloud technology, receives the signature of a file, that has to be checked. As soon as all the necessary data has been transmitted, the cloud service performs the analysis-job and sends back the result to the device, e.g. "the file is malicious" or "no suspicious signatures found". Although this procedure could be implemented deploying a single server, cloud computing offers functionalities and opportunities, that traditional client-server-architectures lack: High dynamic scalability, data transparency and high throughput computing armbrust2009above. This comes in handy, when there are many thousands of users accessing the authentication framework services from the cloud.

Such a framework was proposed by Chow et al. [9]. Their so-called *TrustCube* uses **policies** to support authentication decisions, which are taken based on calling patterns, SMS activity, website access and location. Services are modularized into a star-shaped topology generating privacy benefits, as only the central node needs to collect user-specific data. Among these services one finds data aggregators (i.e. the implicit authentication server), an authentication engine and several authentication consumers. The client software was built for the Android operating system. The elementary authentication procedure is as follows: After the handset has collected **user data** for a specific time window, the data is packed and reported to the data aggregator (and afterwards deleted from the device, to free memory). After that, the client agent collects **phone-specific information**, such as firmware version and running applications. Finally the authentication service calculates a score and makes a decision based on the give **policy**. Such a policy may look like this: "(1) the device should run Android 2.1 update 1 or above AND the WiFi SSID should be "hospital" AND only default and hospital applications can be installed; (2) the minimum score to view medical data is 800; (3) if the authentication score is below the minimum, the user must use a PIN pad to further authenticate."

5.4 Implicit Authentication on Other Portable Devices

Some solutions presented in the literature do not primarily target smart phones. Yazij et al. developed a prototypical implicit authentication framework that runs on laptop computers [31]. These devices offer a mature file-system implementation, which the researches used for authentication beside network usage and location. In their conclusion they stated that in the future they as well will focus on smart phones.

Lastly, Shi et al. [26] and Jakobsson et al. [17] suggest the use of implicit authentication for portable medical devices. They think of a digital clipboard that doctors carry around. These devices contain highly sensitive information, which on the one hand has to be accessible very fast in case of an emergency - typing in a password might waste critical seconds and continuous authentication could speed up getting to the information. On the other hand has to be protected from unauthorized access, for example the data for a certain patient should only be available for doctors of a specific position at a hospital.

6 Adversarial Models

Any form of authentication tries to defend a certain device, account, area etc. against foreign attacks to protect sensitive data. The attackers - or adversaries - possess different incentives and capabilities, which have to be specified in order to be able to shield personal data. For implicit authentication, some particular characteristics of possible adversaries have to be taken into account. Shi et al. provide a detailed adversary model in which they characterize adversaries by roles, incentives and capabilites [26]. Their model is summarized below.

6.1 Roles

If a smart phone, for example, gets into the hands of someone other than the user, this person is not necessarily an *attacker*. In certain cases, *Friends or co-workers* got hold of the handset. When at home, *family members* can easily access phones not originally belonging to themselves. *Strangers* might have stolen the device or found it in a public place where the user accidentally left it. Finally, *enemies or competitors* could try to reap information for political or industrial espionage. This leads right to the different types of incentives, that an adversary might possess.

6.2 Incentives

When it comes to the question "Why would someone want to get hold of my phone?", one can describe certain incentives that motivate adversaries to try and capture a device. If an adversary wants to gain *financial advantages* this may happen for the following reasons: (1) He can make free phone calls or use the Internet at the owner's expense. Or he uses the device for entertainment, maybe because he does not own one himself. (2) He might sell the device or parts of it. (3) If personal data, such as emails, reveals access information for bank accounts, the attacker is inclined to take advantage of this. Thus, he may perform on-line purchases or transfer money to his own account at the cost of the owner of the stolen device. The latter is a major issue and the most difficult to resolve.

Another incentive is sheer *curiosity*. Family members, co-workers or the significant other might want to know read a user's emails or SMS messages. Possibly the browsing or phone call history are also interesting for others. The issue becomes more serious if an adversary is driven by espionage. He might want to obtain sensitive data concerning business matters.

Additionally, some illegitimate users are not willing to take any advantage, but to *sabotage* a victim, either financially or in reputation matters. This means that they buy things on-line just to cause trouble for their victim. Aside from that, the phone can be used to publish embarrassing remarks or pictures on social networks - recently informally termed "frape", a combination of the words 'Facebook' and 'rape' [23].

Finally, a stolen phone can cater for the thief's *anonymity*. He could access illegal web-sites or share illegal material and not get caught, because the traces lead to someone else.

6.3 Capabilities

Capabilities are established especially for implicit authentication. While every attacker being asked for a password immediately knows, that there probably exists one, this is not the case if the user is verified implicitly. However one can attribute three different capabilities to an adversary.

The *uninformed stranger* is not aware, that implicit authentication takes place. Thus, after capturing the device, the person is likely to use the handset as his own, or use it to achieve different incentives as described above.

In certain cases, implicit authentication becomes more challenging when dealing with an *informed adversary*, who is aware of the existence of implicit authentication. It might be easier to imitate certain features, if the attacker knows the legitimate owner in person. The more features are used for verification, the more difficult it becomes even for the informed adversary to game the system.

Lastly, if an attacker infects the handset with *malware*, behavioral patterns can be logged and mimicked afterwards, which would invalidate the whole implicit authentication system. There are certain measures to be taken to protect the implicit authentication software from malware, like installing Anti-Virus solutions.

7 Limitations of Implicit Authentication on Mobile Devices

This section covers the deployability of implicit authentication on mobile devices. It remains open what factors do have to be considered when designing and programming continuous authentication software that is targeted at mobile devices.

7.1 Battery Usage

Implicit authentication happens in the background, therefore implementations use background processes to provide continuous authentication. However, these processes use up a part of the battery capacity, so the phone has to be recharged more often. This in turn reduces the battery lifetime, since charging cycles are not infinitely repeatable.

A detailed calculation of energy consumption through implicit authentication was carried out by Yazij et al. [31]. They examined how implicit authentication relying on network- and file access plus location influenced the battery lifetime of a laptop. It was shown, that remote attestation consumes less energy: Only **6.6%** were used up by implicit authentication mechanisms if deploying an external service whereas 42.6% of battery capacity were wasted if the anomaly detection took place on the device itself.

7.2 Calculation Speed / Detection Latency

Another problem with implicit authentication is detection latency. The paradigm states that authentication happens unobtrusively, so that a secured device should not interrupt the user and divert her from tasks. Since other factors have impact on how often the authentication score can be calculated, there still is a certain time window, where the handset can be used by an attacker without the system noticing. If battery lifetime and processor performance increase one can minimize the detection latency by locally calculating authentication scores.

7.3 Data and Traffic Amount

Collecting data eventually leads to storing it somewhere on the device - either only temporarily or permanently. The log data becomes larger if multiple cues are used for authentication, which we have seen, is recommended. In a user study conducted by Yazij et al. participants generated log data between 80MB and 25GB within two weeks on laptop computers. However, probably none of current handsets could cope with log data above 10GB. Besides, users actually would like to use the limited space for features for which perceived usefulness is a lot higher - log data has no obvious benefits.

One already addressed solution are data aggregators or external storage in general. This in turn suffers from the fact that the data has to be transfered from the mobile device to the cloud, for example. Mobile phone contracts usually cover a certain traffic amount that is free of extra charge. It is probable that transferring a lot of log data through UMTS or LTE networks might eventually lead to higher monthly costs, which some users will not be willing to take. Thus, a good balance between locally stored data and traffic has to be found. In order to minimize traffic, one could also find better ways to compress the data. If the network isn't available at all, for example because the user travels to a foreign country and refuses to pay extra fees for roaming, the calculation of the authentication scores will have to take place on the device. Otherwise implicit authentication is deactivated.

7.4 Account Sharing

In certain cases, a user might allow a friend to use her phone, for example to look up something on the Internet or for entertainment. If the phone uses implicit authentication mechanisms, the phone is able to detect the illegitimate user and possibly shuts down. This process might even reoccur when the correct PIN is entered afterwards by a different person. Therefore solutions to this problem have to be found, without burdening the user of switching on and off the authentication services.

7.5 Reliability

Even though combining multiple features to authenticate a user implicitly yields relatively good results, there still remains a certain percentage of impostors who are mistaken for the legitimate user. In other words, an FAR of 0% is desirable but in most cases not possible. The same holds true for usability matters, where an FRR of 0% would mean no bothering re-authentication, but one cannot avoid it - for the moment.

8 Conclusion

In this paper, a detailed view of authentication and implicit authentication has been described. The most valuable features that implicit authentication relies on are gait, keystroke and behavioral patterns such as phone call- or web-surfing-activity. Although the computational power of smart phones rapidly increases, outsourcing implicit authentication tasks, like decision taking processes, seems the most reasonable at the moment. Specially designed frameworks for cloud computing might become a de facto standard in the near future, unless users refuse the new technology. Fortunately, current studies indicate that users would welcome the establishment of implicit authentication [7], [12]. Also, creating software that runs on as many operating systems as possible is a major task that the developers are yet to face.

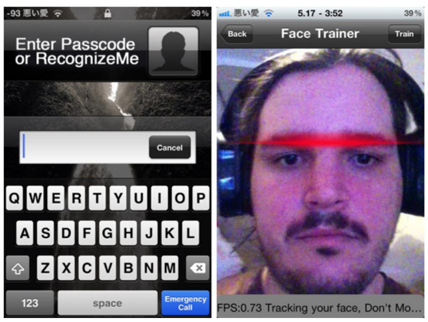


Fig. 4. The RecognizeMe iPhone app uses the front camera to explicitly authenticate a user [2]

Although implicit authentication might not be able to replace explicit authentication entirely, and although parts of it show some disadvantages, the principle is really promising in terms of both security and usability.

The author would like to suggest face recognition techniques for implicit authentication on mobile devices, which have not been under research recently. As new generation smart phones have additional cameras oriented to the user's face, one could think of taking pictures every few minutes to see whether the face matches the legitimate user. There is already an iPhone app called RecognizeMe (see *figure 4*) that can authenticate the user through the iPhone 4's front camera, given it is jailbroken [2]. It seems rather easy to enhance this application to become suitable for implicit authentication.

References

[1] H. Ailisto, M. Lindholm, J. Mantyjarvi, E. Vildjiounaite, and S. Makela. Identifying people from gait pattern with accelerometers. In *Proceedings of SPIE*, volume 5779, page 7, 2005.

[2] Apocalipse. RecognizeMe iPhone App on Cydia. `http://modmyi.com/cydia/com.apocolipse.recognizeme`. visited 18.06.2011.

[3] Apple. Set a Passcode Lock with Find My iPhone. `http://www.apple.com/mobileme/news/2009/09/set-a-passcode-lock-with-find-my-iphone.html`, 2009. visited 02.06.2011.

[4] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE transactions on pattern analysis and machine intelligence*, 19(7):711–720, 1997.

[5] J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Combining biometric evidence for person authentication. *Advanced Studies in Biometrics*, pages 1–18, 2005.

[6] R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE transactions on pattern analysis and machine intelligence*, 17(10):955–966, 1995.

[7] A. Buchoux and N. Clarke. Deployment of keystroke analysis on a smartphone. In *Proceedings of the 6th Australian information security management conference. Perth, Western Australia: SECAU-Security Research Centre*, pages 40–47, 2008.

[8] S. Card, T. Moran, and A. Newell. Computer text-editing: An information-processing analysis of a routine cognitive skill. *Cognitive Psychology*, 12(1):32–74, 1980.

[9] R. Chow, M. Jakobsson, R. Masuoka, J. Molina, Y. Niu, E. Shi, and Z. Song. Authentication in the clouds: a framework and its application to mobile users. In *Proceedings of the 2010 ACM workshop on cloud computing security workshop*, pages 1–6. ACM, 2010.

[10] J. Cutting and L. Kozlowski. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the psychonomic society*, 9(5):353–356, 1977.

[11] I. Damousis, D. Tzovaras, and E. Bekiaris. Unobtrusive multimodal biometric authentication: The humabio project concept. *EURASIP journal on advances in signal processing*, 2008:1–11, 2008.

[12] S. Furnell, N. Clarke, and S. Karatzouni. Beyond the pin: Enhancing user authentication for mobile devices. *Computer fraud & security*, 2008(8):12–17, 2008.

[13] D. Gafurov, K. Helkala, and T. Søndrol. Biometric gait authentication using accelerometer sensor. *Journal of computers*, 1(7):51–59, 2006.

[14] R. Greenstadt and J. Beal. Cognitive security for personal devices. In *Proceedings of the 1st ACM workshop on AISec*, pages 27–30. ACM, 2008.

[15] G. Hayday. Security nightmare: How do you maintain 21 different passwords. `http://tinyurl.com/silicon-security-nightmare`. visited 03.07.2011.

[16] A. Jain, F. Griess, and S. Connell. On-line signature verification. *Pattern recognition*, 35(12):2963–2972, 2002.

[17] M. Jakobsson, E. Shi, P. Golle, and R. Chow. Implicit authentication for mobile devices. In *Proceedings of the 4th USENIX conference on hot topics in security*, pages 9–9. USENIX Association, 2009.

[18] S. Karatzouni and N. Clarke. Keystroke analysis for thumb-based keyboards on mobile devices. *New approaches for security, privacy and trust in complex environments*, pages 253–263, 2007.

[19] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. Makela, and H. Ailisto. Identifying users of portable devices from gait pattern with accelerometers. In *IEEE international conference on Acoustics, speech, and signal processing, 2005. Proceedings (ICASSP'05).*, volume 2, pages ii–973. IEEE, 2005.

[20] Motorola. Fingerprint reader of the motorola atrix 4g. `http://mediacenter.motorola.com/ImageLibrary/Detail.aspx?MediaDetailsID=1472`. visited 09.06.2011.

[21] M. Nauman and T. Ali. Token: Trustable keystroke-based authentication for web-based applications on smartphones. *Information security and assurance*, pages 286–297, 2010.

[22] PayPal. `http://www.paypal.com`. visited 03.07.2011.

[23] Phailure. Definition of frape on urban dictionary. `http://www.urbandictionary.com/define.php?term=Frape&defid=2463827`. visited 17.06.2011.

[24] RSA. SecurID. `http://www.rsa.com/node.aspx?id=1156`. visited 03.07.2011.

[25] H. Saevanee and P. Bhattarakosol. Authenticating user using keystroke dynamics and finger pressure. In *Consumer communications and networking conference. CCNC 2009. 6th IEEE*, pages 1–2. IEEE, 2009.

[26] E. Shi, Y. Niu, M. Jakobsson, and R. Chow. Implicit authentication through learning user behavior. *Information Security*, pages 99–113, 2011.

[27] M. Tamviruzzaman, S. Ahamed, C. Hasan, and C. O'brien. epet: when cellular phone learns to recognize its owner. In *Proceedings of the 2nd ACM workshop on assurable and usable security configuration*, pages 13–18. ACM, 2009.

[28] S. Tulyakov, F. Farooq, P. Mansukhani, and V. Govindaraju. Symmetric hash functions for secure fingerprint biometric systems. *Pattern Recognition Letters*, 28(16):2427–2436, 2007.

[29] A. Whitten and J. Tygar. Why johnny can't encrypt: A usability evaluation of pgp 5.0. In *Proceedings of the 8th USENIX Security Symposium*, pages 169–184. Citeseer, 1999.

[30] S. Wiedenbeck, J. Waters, L. Sobrado, and J. Birget. Design and evaluation of a shoulder-surfing resistant graphical password scheme. In *Proceedings of the working conference on advanced visual interfaces*, pages 177–184. ACM, 2006.

[31] S. Yazji, X. Chen, R. Dick, and P. Scheuermann. Implicit user re-authentication for mobile devices? In *Ubiquitous intelligence and computing: 6th international conference, Brisbane, Australia, July 7-9, 2009, Proceedings*, page 325. Springer-Verlag New York Inc, 2009.

Implicit Authentication On Mobile Devices

Hans - Peter Dietz

Abstract— With the proliferation of ubiquitous and pervasive computing more and more services and applications become accessible through mobile devices accompanying users in their daily lives. While hard- and software adapt to this trend and improve usability, security, especially authentication techniques, remain behind in terms of old fashioned password-entry approaches *(see e. g. [38], [58], [15])* . It is shown that well-established secret-knowledge techniques do not fit into this development. They offer weaknesses both in usability and security.
Reasons for the implementation of implicit authentication on mobile devices are introduced and discussed. A thorough overview of possible cues, encompassing physical and behavioral biometrics is given. Furthermore the concepts of multimodal systems are investigated and evaluated. Finally the presented concepts and techniques are discussed and evaluated.
The paper is concluded with an overview of the subject matter.

Index Terms—Authentication, Implicit, Mobile Devices, Verification, Biometrics, Ubiquitous Computing

1 Introduction

The market for portable devices grows rapidly and steadily. According to Gartner's mobile market report [47] the sales in mobile devices exceed 428 million in the first quarter of 2011. This represents a 19% increase year-over-year. Furthermore, with the advent of tablet devices like Apple's *iPad* [2], the mobile devices circulating in households gain another growing branch [29]. This trend towards more and more mobile, wirelessly communicating devices populating modern societies demands for attention not only of market analysts and sales people but IT-specialists who have to adapt existing access-, interaction- and security-mechanisms to this development.
Moreover the emergence of Cloud Computing through services like *Google Music* [23], *Windows Live* [36], *Dropbox* [18] and others enables users to access data and services from nearly everywhere. The combination of these increasingly popular *Cloud-Services* [49] and mobile devices able to communicate with them, forms the context for this paper. It specifically focusses on two aspects: *security* and *calm computing*.
In [56] Weiser predicts *"the third wave of computing [...]"* as *"[...] that of ubiquitous computing, whose cross-over point with personal computing will be around 2005-2020"* [26] and that *"calmness is a fundamental challenge for all technological design of the next fifty years"*.
Calmness in the sense Weiser meant it describes the shift of focus and attention from the center to the periphery - in other words: computers should stay out of our way.
In conclusion this implies that security topics should move to the background of our attention. Therefore this paper provides an overview of possibilities to move authentication from happening explicit to implicitly taking place. This implies serving the demand of the ubiquitous era for calmness while retaining, possibly even increasing, the necessary security level. The importance of a high level of security, especially on mobile devices, is proven by the facts that they are frequently stolen [3] and the growing amount of personal data stored and/or accessible through them [1].
To begin the examination of implicit authentication on mobile devices section 2 provides insights into the flaws of traditional secret-knowledge techniques and illustrates how implicit authentication can help to improve them. Section 3 introduces some definitions of commonly used terms in authentication while section 4 engages in the technical possibilities implementing implicit authentication. These are grouped into physical biometric, behavioral biometric and physical approaches. The section ends with a discussion of the introduced techniques. Finally the last section concludes this paper.

- *Hans - Peter Dietz is studying Media Informatics at the University of Munich, Germany, E-mail: dietzh@cip.ifi.lmu.de*

- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

2 Why implicit authentication is desireble

This section elucidates the reasons that make implicit authentication desirable. First of all the term *implicit authentication* needs to be clarified. Jakobsson et al. [48] define implicit authentication as:
"the ability to authenticate mobile users based on actions they would carry out anyway".
This definition serves as the basic notion of the term in this paper.
As can be seen from the 3G security specification [1] there are a number of requirements and specifications for security on mobile devices. At the moment these issues are mostly implemented through passwords, such as the personal identification number *(PIN)*. In their paper Nauman et al. [38] point out the problem of entering passwords, especially long ones, possibly composed of mixed uppercase, lowercase letters, digits and special characters on mobile phones. They state that miniature keyboards and on-screen touch keyboards lead users to choosing simpler, and thus weaker, passwords [38]. This implication of input devices antagonizing security on mobile agents illustrates the need to adapt security mechanisms and serves as the first argument for a shift from explicit to implicit authentication.
Another concern with passwords, being the main way of user authentication, is their weakness and insufficiency [58]. In general a well chosen password cannot be classified as a weak authentication mechanism, but the circumstance of having to enter it through uncomfortable ways of input, oftentimes leads users to either choose easily enterable and therefore weak ones or even abandon them completely [15]. In their survey, Clarke et al. [15] also point out that a password, being a secret-knowledge technique, has long-established drawbacks. These are often introduced by the authorized users themselves. Examples for such weaknesses include the aforementioned *"selection of weak(guessable) strings, as well as sharing passwords with other people, writing them down and never changing them"* [15]. The general tendency of choosing weak passwords can, for instance, be inferred from [33], which states that most passwords can be cracked in less than a minute. Though the example applies to passwords used in health care systems it can be related to the situation on mobile devices and the outcome can be assumed to be even worse, due to the oftentimes limited amount of symbols available to compose secrets such as a PIN.

Figure 1 shows another issue with PINs: people tend to not change them, compromising proper secret usage. This behavior leads to an additional concern: static authentication.
Point-of-entry and single-sign-on techniques do not provide continuous authentication. Continuous authentication means that the

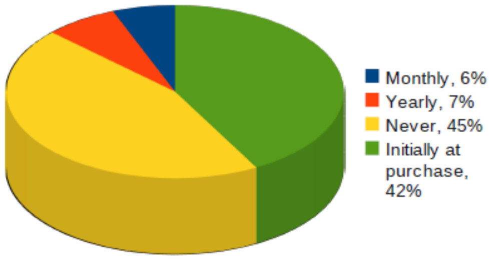


Fig. 1. Changing PIN-code behaviour, derived from [15]

user's identity is verified continuously even after the initial login. The reasons this is desirable can be devided into two categories. The first one is given by the tendency of users to rarely log out, i. e. restart their device and re-authenticate [15]. This leaves the adversary, e. g. a thief, with no barrier at all to acquire data and services after appropriating the device [48]. The second issue lies within the flat or static nature of point-of-entry authentication: it is assumed that all accessible information, applications and services are of equal value, therefore not requiring any further access control restrictions [21]. The emerging single level of security is depicted in figure 2.

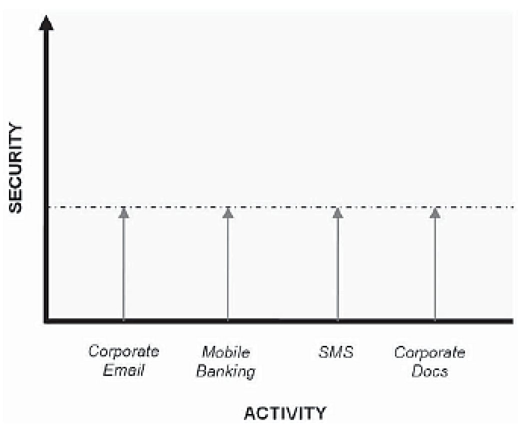


Fig. 2. Current authentication situation: after loging in there is only a single level of security [21]

Having investigated the flaws of traditional password based authentication there are some other advantages of implicit authentication techniques left to mention. A device secured by continuous implicit techniques clearly implies higher *"hacking costs"* [24]. The cost of hacking security mechanisms based on what the user is *(e. g. her behavior)* is significantly higher *(i. e. requires more work)* than hacking a password, while, at best, reducing user efforts such as entering her PIN-code.
Furthermore, implicit authentication can help to improve user experience. In their study Falaki et al. [19] state that *"[...] mechanisms to improve user experience or energy consumption will be more effective if they learn and adapt to user behavior"*. The techniques of biometric authentication, which are examined later in this paper, work on this foundation: the user's behavior.
Many more references can be found that indicate that adaptive devices, capable of learning user behavior improve user experience (e. g. [58], [19], [21]).
Summing up, this section pointed out that implicit authentication is desirable for a number of reasons:

- provide stronger cues
- provide non-intrusive continuous authentication
- improve user experience
- enable different levels of security

Some of these reasons, e. g. the possibility to provide different levels of security, will be explained further in later sections.

3 Definitions

When referring to authentication and its applications a few terms need to be introduced and defined. The definitions given are derived from Bigun et al. [6].
First off the users of such applications are often referred to as *clients*, whereas the *impostor* is the adversary. Adversary in this context means not only malicious people, but possibly the whole world population. When the client provides her identity, i. e. enters her password, the application performs a matching, yielding a *score*. If this score is higher than the *verification threshold* the claim is accepted and the client's access granted. Different criterions, e. g. the password string, the fingerprint or the gait pattern, are referred to as *cues*. In multimodal authentication frameworks various subsystems, called *experts* evaluate the different cues and the scores produced by the experts serve as the input for a *fusion strategy*, which yields a combined score, deciding if the claim is accepted or rejected.

4 Techniques for Implicit Authentication

Having introduced the general goals of implicit authentication, this section provides insights in techniques available to implement implicit authentication. These techniques are grouped into three different concepts: physical biometric, behavioral biometric and purely physical approaches.

4.1 Physical Biometric Approaches

Security systems incorporating physical biometric cues concentrate on the physical conditions of the client. The three most important of these are the face, the fingerprint and the voice.

Face recognition In human interaction face recognition plays an important role for person authentication. A great part[1] of humanity has a unique and individual face, differentiating one from the other. Furthermore face analysis engages special regions of the brain, other than those used for analysing objects [20]. Since face recognition for person authentication is a natural process, it leads to the idea of imitating nature and therefore performing face recognition for authentication purposes on mobile devices.

The general procedure for face recognition is depicted in figure 3. At first the face has to be detected within the image, then facial feature points, such as corner points of eyes or the mouth have to be found. The image data at the feature points is then extracted and the resulting feature vector is compared to the stored feature vectors of authorized users for verification.
Usually face recognition can be used for two purposes: identification and verification. The former is used to identify people, the latter for *"matching the input identity with [the] registered identity in the database"* [27]. Therefore verification forms the main application of face recognition in implicit authentication. State of the art face recognition systems, such as the one proposed by Ijiri et al. [27], are efficient enough in terms of processing speed and memory usage to be a viable choice for mobile devices. Furthermore, a great part of the mobile devices in use today, such as smart- and cellphones, are already shipped with one or even two integrated cameras. These cameras

[1] excluding monozygotic twins or the like

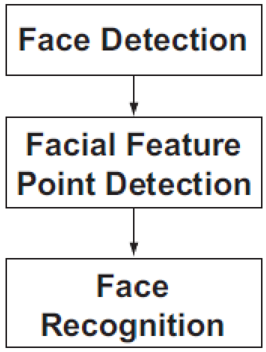


Fig. 3. General face recognition process [27]

oftentimes provide resolutions far exceeding the requirements of face recognition systems [8].
In conclusion, face recognition can be implemented on a large portion of mobile devices without the requirement for new hardware. Given front-mounted cameras the client does not even have to perform a specific action for authentication, but the face recognition system could work in the background, performing the verification anytime the client holds the device in front of herself.

Fingerprint matching Every human being has a unique fingerprint. This natural circumstance can be exploited for human identification, authentication and verification.

Fig. 4. Fingerprint feature extraction: (a) original image, (b) edge detection, (c) feature points (minutiae) [14]

Fingerprint recognition in general is a computationally complex task. It involves *"a lot of trigonometric computation"* [11] and can therefore be classified as a rather time-consuming process. Fortunately there are fingerprint matching systems, like the ones proposed by Chan et al. [11] or Tang et al. [53] which focus on fast execution speed and are specifically designed for embedded and mobile systems.
Figure 5 depicts the general workflow of a minutiae based fingerprint verification process. The general process consists of two steps. At first the fingerprint image is captured by a sensor and the minutiae *(see figure 4)* are extracted. The minutiae are then processed to be stored as a 'master-template', i. e. the feature vector of the authorized client. In the verification phase this process is repeated and a feature vector, i. e. the 'live-template' is generated. A matching between these two templates finally determines a similarity score of the two fingerprints [11].

Due to recent progress in manufacturing techniques, fingerprint sensors small enough to be integrated in mobile devices can be produced [53]. Such small devices combined with efficient algorithms like the ones mentioned above, make fingerprint matching a viable candidate for implicit authentication on mobile devices. The fingerprint sensors could be mounted on top of a button, e. g. the iPhone's home button, which is pressed frequently, enabling non-intrusive, continuous authentication.

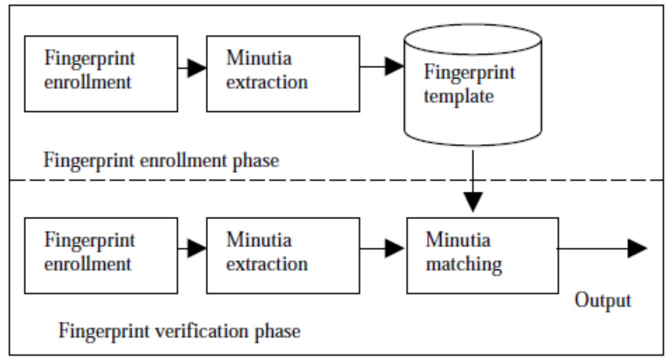


Fig. 5. Fingerprint matching workflow [11]

In summary this implies that it is possible to implement efficient fingerprint matching through additional hardware, working on unique cues.

Speaker Recognition Every living person has a characteristic voice[2], therefore we are able to identify a person based on her voice, e. g. during a phone call. These characteristic properties of a speaker's voice make automatic speaker recognition systems possible. As with face recognition and fingerprint matching it can be used for two tasks: speaker identification, i. e. *"[...]determining who is talking from a set of known voices or speakers"* [43], and speaker verification, i. e. *"[...] determining whether a person is who she claims to be [...]"* [43]. Further, speaker recognition systems can either be text-dependent or text-independent. While the constraint for prior knowledge of the text claimed by the former can increase the performace, the latter is more flexible and enables completely non-intrusive authentication.

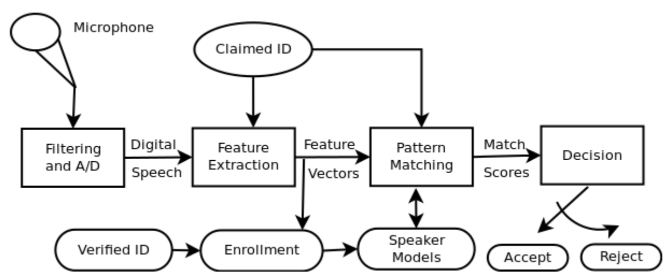


Fig. 6. Generic speaker verification system, derived from [9]

The general model of speaker recognition systems encompasses capturing, filtering noise, feature extraction and matching (see figure 6). Capturing is done by a microphone, which nowadays is present in almost any mobile device. Next the environmental noise is filtered through the application of a bank of bandfilters. Although the speech signal does not contain any exclusive feature conveying the speaker's identity, feature extraction is based upon speech spectrum shape, resonances and pitch harmonics [43]. Oftentimes some form of channel compensation, i. e. normalization of the signal, is performed to compensate for channel effects (e. g. a person calling from a cellphone sounds different than a person calling from her office telephone). The resulting feature vectors are then fed into a matching algorithm yielding a score.
To enroll a speaker recognition system a speaker model[3] has to be established. This means that the characteristics of the authorized client's

[2]excluding mute or otherwise handicapped people
[3]also referred to as *Voice Print*

voice have to be learned. To achieve this different techniques offering different trade-offs regarding performance, ease of training and updating, storage and computational complexity can be applied. Some frequently used examples are template matching, nearest neighbor or hidden markov models. Within the first a sequence of template feature vectors from a fixed phrase are stored, making it a text-dependent approach. In the second *"[...] no explicit model is used; instead all feature vectors from the enrollment speech are retained to represent the speaker"* [43]. The last technique stores the temporal evolvement of the feature vectors and models their statistical variations yielding *"a statistical representation of how a speaker produces sounds"* [43]. Analogous to the speaker model an impostor model can be established, acting as a normalization.
In summary speaker recognition systems can be implemented on most mobile devices without the need for additional hardware. Especially on mobile devices used for telephony, where speech is one of the main modalities, it can provide a non-intrusive, implicit way to authenticate the client using a unique cue.

4.2 Behavioral Biometric Approaches

Behavioral biometric approaches concentrate on how the client usually behaves. They analyze the clients bearing and infer feature vectors representing a client model which can be used for client authentication and impostor detection. This subsection describes five possible cues: keystroke-dynamics, gait, filesystem and network access, location and general usage behavior. All behavioral biometric techniques follow a general workflow - at first a client model has to be established. This is done in the learning phase, where some kind of classification or machine learning algorithm is trained. When the client model is set up the system can enter the detection mode, where client behavior is observed and in case any anomalies are detected some form of escalation to a verification module is performed (see figure 7).

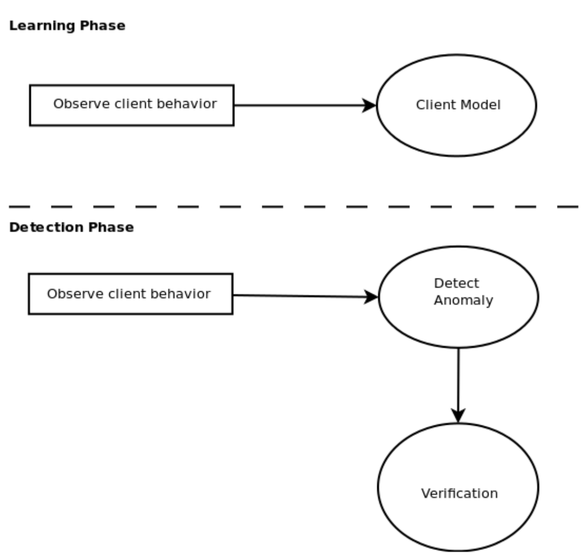


Fig. 7. General Workflow Of Behavioral Biometric Systems

Keystroke-Dynamics With the advent of mobile devices able to access the internet, new use cases apart from telephony became relevant, such as email writing and browsing the internet. Therefore most state of the art mobile devices are shipped with soft- and/or miniature keyboards improving character input. This development leads to the approach of porting keystroke analysis techniques available from traditional desktop and laptop computers to mobile devices and using the resulting data for authentication purposes.
The most elementary metric for keystroke-dynamics is based on the time span between successive keystrokes [10]. This duration is referred to as *digraph*. Since soft- and/or miniature-keyboards vary significantly from traditional desktop and laptop keyboards, however, more parameters have to be included. One of the most successful approaches was reported by Zahid et al. [59], which measures the key hold time, i. e. the time span between key press and key release events of one key, the digraph and the error rate, i. e. the number of key press events for the backspace key. The resulting feature vectors are then classified using fuzzy logic [31] to establish the client model. In addition to the creation of an initial client model Zahid et al. continue to optimize and adapt the model through dynamic optimizers like particle swarm optimizers [30] and genetic algorithms [22].
Other possible pattern recognition techniques include linear and non-linear distance techniques [55], *z*-tests [37], Bayesan classifiers [39] and neural network approaches [13], [40], [39].
Overall implicit authentication based on keystroke-dynamics does not require new hardware, is non-intrusive and can be performed continuously. Furthermore, since keying patterns are unique to individuals [32], this technique works on a unique cue.
Smartphones and tablets seem to be especially well suited platforms for the deployment of this cue for authentication purposes, since they are frequently accessed through character input.

Gait Pattern The gait pattern of a person, i. e. her walking style, is a robust measure, i. e. it is distinctive [5]. Combined with the facts that a great part of mobile devices comes equipped with accelerometers and that they are carried around a lot, this leads to the idea of using gait pattern recognition for authentication.
In [34] Mäntyjärvi et al. describe their approach using accelerometer data and correlation to identify people. In the learning phase the collected data (see figure 8) is normalized and divided into parts representing steps. These steps are then grouped into *a*- and *b*-steps representing but not identifying left and right steps. A- and b-steps are then averaged to form a template, i. e. the client model. During the detection phase the signal is processed analogous, resulting in *c*- and *d*-step *(c- and d- are simple variable names, in analogy to a- and b- steps)* models, which are cross-correlated with the a- and b-steps yielding the verification score.

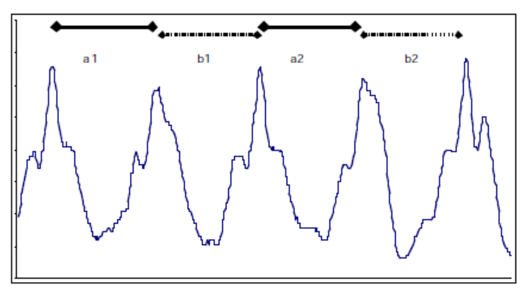


Fig. 8. Typical acceleration signal [34]

Tanviruzzaman et al. [52] also include the gait pattern in their multimodal *(see next section)* approach. Instead of correlation they apply a Dynamic Time Wraping algorithm [4] to match between different modes of walking. Their system, called *ePet*, was implemented in Java and they delegate an implementation on the iPhone to their future works.
Generally speaking gait patterns offer a unique cue which can be obtained in a non-intrusive way facilitating implicit authentication. Furthermore a great part of the mobile devices circulating, like smartphones, come with integrated acceleration sensors, which provide the necessary data. Since deliberate imitation of another person's walking style is considered difficult [34], this technique can be credited as another viable choice for the implementation of implicit authentication on mobile devices.

Filesystem and Network Access Using mobile devices means accessing applications, files and network resources. This observation lead Yazji et al. [58] to the idea of monitoring the client's network and filesystem access to identify anomalous behavior and therefore detect impostors respectively verify clients. Consequently their approach is based around the idea that *"different individuals have differing computer use patterns"* and *"operating systems have access to a great deal of information"* [58]. Figure 9 provides an overview of the proposed system architecture.

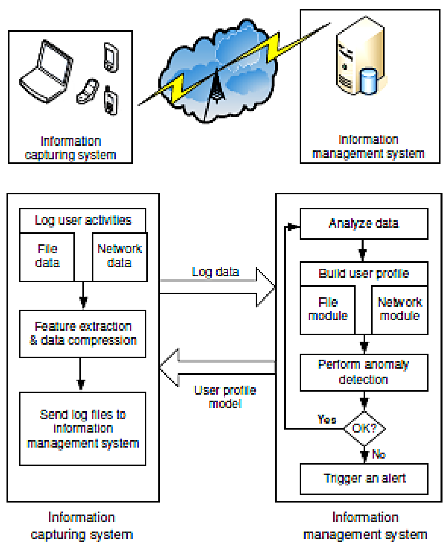


Fig. 9. Filesystem and network access monitoring system overview [58]

The data collection module captures file access patterns with the help of a modified version of *FileMon* [45]. It generates an access record on each system call while filtering records generated by system services to reduce transmission energy and reduce noise. Network activities are captured using *WireShark* [57] with a focus on user-dependent network events. The features extracted to form the feature vectors are *process*, *time* and *location* for file access and *destination IP*, *time* and *protocol* for network access respectively.
The final client model is obtained by applying K-Means clustering [25]. To detect anomalies without previous knowledge of anomalous behavior a *distribution vector* (DV) of a data set is constructed, where the *i*th element *"is equal to the percentage of records in the ith cluster"* [58].
The distribution vector of the current data, i. e. DV(*evaluation*) is then compared to DV(*training*) using *Euclidean distance* yielding the verification score.
Though Yazji et al.'s implementation incorporates third party software, it still serves as an example for the general concept of monitoring filesystem and network access to detect impostors. Every mobile device capable of running applications and accessing network resources has an operating system and oftentimes an application programming interface (API) providing the necessary data. Therefore an implementation of the proposed architecture should be possible at least for the most popular devices, such as the iPhone, the iPad, Android Phones, etc.. As a result this system constitutes a non-intrusive way of authentication working on arguably distinctive cues which can be implemented without the need for new hardware.

Location Tracking As the name *"mobile device"* implies these devices are designed to be carried around for use anywhere. Therefore location as a cue naturally comes to mind.
In their papers Shi et al. [48] and Jakobsson et al. [28] propose a framework for implicit authentication on mobile devices incorporating exactly this cue. The client, respectively the client model, is therefore composed of characteristic locations in relation to the time of day.
As they point out in their introductions most people have daily routines - they go to work, have lunch, return home and maybe have specific evening activities, e. g. sports, that they carry out on a regular basis. These traces can be used to formulate feature vectors composed of coordinates, e. g. GPS, per daytime.
Having established such a client model the system is able to spot anomalies and therefore recognize possible impostors.
The flaw of this cue lies within the fact that location on it's own is not a reliable factor - people can move to new and unknown locations having their mobile devices with them, without the emergence of an access violation. Therefore it should only be used as a supporting cue within *multimodal* (see next section) systems, like both authors did. Nevertheless, most mobile devices are shipped with location sensors, e. g. GPS sensors, enabling the implementation without the need for additional hardware. Though location traces per se are not unique upon the whole world population, they are distinctive enough to be included in a multimodal architecture. Furthermore it is easy to implement and does not require a lot of computing power. Concluding, it constitutes a viable, non-intrusive *support-cue* in multimodal authentication systems implementing implicit authentication.

General Usage Behavior The term *Behavioral Biometrics* includes all kinds of aspects of human behavior. This implies that the preceding paragraphs by no means encompass all possible cues. This paragraph therefore tries to present some other interesting ideas for client authentication.
In their work Chang et al. [12] present a method of distinguishing TV viewers through analyzing people's hand motions and button press sequences on remote controls. The necessary data is captured by acceleration sensors mounted to the remote controls. In their experiment they trained Support Vector Machines [51] and Max-Margin Markov Networks [54] to create what resembles a client model. Their feature vectors were composed of *motion-features* and *button-features*, i. e. accelerometer data and button press/release events and timestamps respectively. The achieved accuracy varied from 70-92%, which denies it's application in authentication systems. However, the idea of monitoring how people hold, move and interact with their devices seems to be a promising idea to the author and it might be possible to optimize the system's performance to make it a viable, non-intrusive option for implicit authentication on mobile devices, that does not necessarily require additional hardware.
Another interesting approach is the previously mentioned work of Shi et al. [48], which is a multimodal approach comprising various usage-behavioral features. The proposed system uses a client's short-messaging, phone call and browsing behavior supported by location traces. These features are combined into a user-model through probability density functions conditioned on the time of the day and the day of the week. The applied fusion strategy then calculates the probability of the observed behavior, described as a tuple of variables $v_1,...,v_k$ and a timestamp t being valid, i. e.:

$$score := p(v_1|t) * p(v_2|t * ... * p(v_k|t) \text{ [48]}$$

This general concept of observing how, where and when the client usually interacts with her device is a very robust approach, because imitating all of these aspects would be very difficult for an impostor. Furthermore the collected data is readily available and therefore does not need any new hardware. It can be collected in a non-intrusive way and, as figure 10 depicts, performs well:

On the x-axis the number of times the legitimate user used the device before she failed to authenticate is depicted, therefore higher values imply better performance. The y-axis depicts the number of times an impostor was able to use the device before she was detected as such, therefore lower values imply better performance. The data for location tracking only reveals that, as stated before, it cannot provide a good enough performance to serve as the only cue, since good performance on the x-axis induces bad performance *(i .e. high values)* on the y-axis. Generally speaking, cues performing well produce low

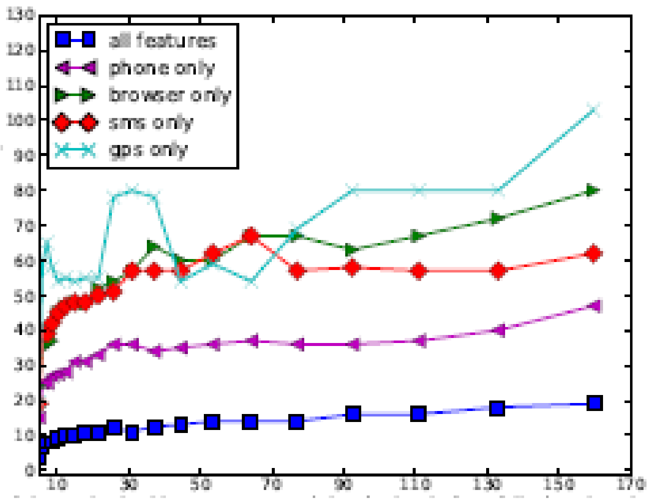


Fig. 10. Evaluation of Shi et al.'s fusion strategy (95 percentile) [48]. X-axis: Number of times the legitimate user used the device before failed authentication. Y-axis: Number of times an impostor used the device before detection.

(in terms of y-axis values) lines, which, applied to the image, reveals that the cue location is outperformed by browser data, which in turn is outperformed by short-messaging behavior, which in turn is a weaker cue than calling-behavior. The combination of all of these in turn outperforms every single cue.

4.3 Physical Approaches

Physical approaches rely on physical devices interacting to authenticate the client. These physical devices are referred to as *tokens*. Token based authentication techniques require a token to be near or plugged into the authenticating device, it's presence therefore verifies the user. An example of such a system is proposed by Corner et al. [16]. The problem with these systems, when applied to mobile devices, is, that the token has to be carried around, too. This implies that the token must reside in immediate proximity to the mobile device, leaving both susceptible to theft. Furthermore the desired implicit and non-intrusive aspects become attenuated, since the user might have to plug the token into the device, constituting an explicit action.
Other token based systems work with *radio-frequency identification* (RFID). In their papers Patterson et. al [41] and Philipose et al. [42] describe an RFID based application to infer activities from interactions with objects. RFID-tags are small enough to be included in items the client would be expected to always have with her, such as jewelry. However, as Furnell et al. [21] state, such approaches seem impractical, at least in some scenarios, and it is thus unlikely that both users and technicians focus on further development in this direction.
It therefore remains to record that physical approaches constitute a possible way of realizing implicit authentication but there are issues to be solved first.

4.4 Discussion

This subsection tries to relate and discuss the techniques described above. When reviewing different authentication techniques, their performance always is an important factor. Since a complete comparison of the performance of the presented techniques in terms of *false rejection (FRR)*, *false acceptance (FAR)* and *equal error rates (EER)* lies beyond the scope of this paper, it shall be stated that all of the described approaches perform well enough to be used as cues for implicit authentication if not stated otherwise. To provide a rough overview however, the findings of Snelick et al. [50] indicate that tolarable error rates range between 0.1% and 1%.

In terms of performance, the prior mentioned concept of *multimodality* plays an important role. Multimodal concepts combine different cues, therefore deploy several experts and a sophisticated fusion strategy then calculates the final score. On an abstract level this means that different information sources are queried, the responses evaluated and fused into a decision - a concept humans utilize every day [46].
The system described by Brunelli et al. [7] for example combines two physical biometric cues, namely face and voice to yield an improvement from 91% and 88% respectively in the *correct identification rate* to a combined rate of 98%. The fusion strategy employed uses cross-correlation.
Also Bigun et al. [6] confirm through their experiments that combining different biometric traits (voice and fingerprints in their case) results in better verification performance. Their report also includes the concept of adaptive fusion strategies, which i. e. apply weights to the expert scores based on their quality, to further improve verification performance.
Rokita et al. [44] describe a multimodal biometric system working with face and hand recognition. The fusion strategy uses Support Vector Machines to model each client. In their experiments they show that, depending on the number of facial and hand feature points, an average accuracy of 99.82% is achievable.
An interesting and promising approach can also be found in Damousis et al.'s HUMABIO Project [17], which is an unobtrusive, multimodal, biometric authentication system. The cues incorporated include face, gait, voice, tokens and even anthropometric measures. Figure 11 shows an overview of their proposed architecture.

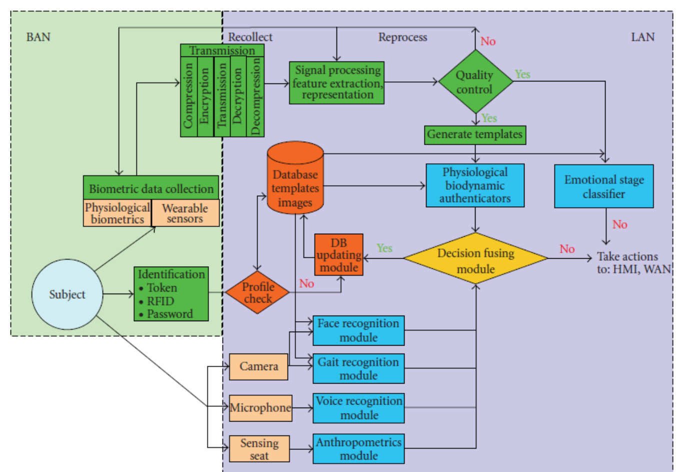


Fig. 11. HUMABIO architecture overview [17]

Concluding this demonstration of the power of multimodal concepts in terms of performance improvement the author would like to cite Bigun et al. [6]: *"[...] even one of the best known mono-modal recognition engines (human face recognition) is not able to reach a recognition rate beyond 80% when it is limited to a single view"*.
Multimodal concepts not only offer advantages in performance, they also enable continuous non-intrusive implicit authentication. The scenario of *Alice*, the client, who accessed her smartphone, equipped with a multimodal, implicit authentication system shall serve as an example. In this scenario it is winter and Alice walks outside, wearing gloves and therefore denying the fingerprint reader attached to the home-button of her smartphone to take her fingerprints. But thanks to the multimodal concept the front camera is still able to provide an image of her face, verifying her as the authorized client. If in this scenario however Alice would be wearing sunglasses and therefore excluding the face-recognition expert, too, still the acceleration sensor could verify her as the authenticated client, due to her gait pattern.
This example scenario illustrates another advantage of multimodal systems compared to mono-modal approaches: even if one or more

experts fail to compute a score, as long as there are others available, the system can still verify the client. Finally, in case all biometric experts fail, an escalation to a secret-knowledge technique, such as PIN, is still possible.
Another advantage of multimodal systems is the possibility to establish different levels of security mentioned in section 2. Point-of-entry and single-sign-on techniques assume that all services and applications are of equal value - an assumption that does not hold true for today's possibilities on mobile devices (see figure 2). For example text-messaging clearly requires a lower security level than accessing a bank account, while both services are available on the same device. Multimodal techniques however enable different levels of security through different kinds and numbers of experts. One could for example imagine that for basic text messaging only a subset of the available experts has to yield a sufficient score, while accessing a bank account requires all available experts to verify the client. The resulting security assessment is depicted in figure 12.

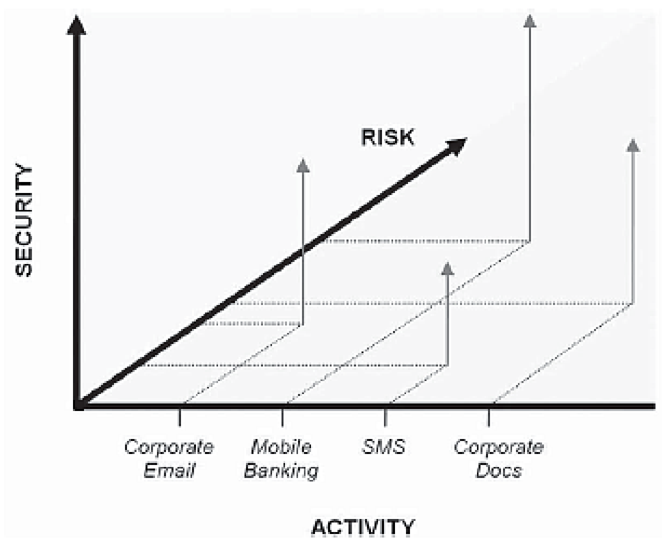


Fig. 12. Multiple levels of security [21]

Having shown the advantages of biometric systems, the problems introduced by them also need to be inspected. An obvious problem with any biometric cue is the discrimination of groups of people whose biometrics cannot be recorded well [17]. An example for this scenario are monozygotic twins, whose faces are so similar to each other that a face recognition system cannot distinguish between them. Another example is given by disabled people who are bound to a wheelchair - they cannot be verified by gait patterns. Though many more of these scenarios are thinkable, a possible solution is already given: multi-modality. If a system incorporates multiple cues the probability that a person cannot be verified by all of the deployed experts can be assumed to be rather low. Another issue is given through the scenario of identity theft. If a person can imitate another person well enough, she can circumvent the biometric experts and access the device at will. Basically this problem is analogous to the theft of a password in secret-knowledge techniques, though it can be assumed that given high performance experts and fusion strategies, cheating biometrics is harder, i. e. connotes higher hacking costs, than hacking or stealing a password.
After pointing out these problems an interesting strategy needs to be mentioned: the concept of *imprinting*. Greenstadt et al. [24] describe this concept in their paper and propose it for future security applications. The idea is analogous to the biological notion of imprinting: When the client has her first interaction with the device, it should imprint on her. This implies that the device stores as many characteristics of the client as possible in a way not even the client can access or change them. Such a system implies a very strong, hardly adaptive relationship between the device and the client and there is no reference in the literature of it's implementation yet. Implementing it would require a very compact learning phase, resembling the *first encounter* and only limited and slow adaption to maximize the value of the imprinted characteristics. Nevertheless this concept seems very promising and worth further investigation to the author.
Adaption per se is an important factor in biometric authentication. When the deployed system first has established a client model based on biometric cues, it must not stop adapting to the client. In fact it is essential that the system continues to observe the input features and adapt the model to that effect. If, for example, a system based on location traces would not adapt, a client moving from *place A* to *place B* could not be verified any more. Furthermore, several biometrics are subject to change: a face can change over time, the way one walks can change, even one's voice is not resistant to change. These changes do not have to happen at a single point in time, but slowly over time. An adaptive system can accommodate to these changes and retain a high level of usability and security.
Concluding this discussion, the author wants to mention a different approach to the problem with secret-knowledge and single-sign-on techniques based on passwords. Mayrhofer et al. [35] describe the idea of generating authenticated, secret keys for small, mobile devices that communicate with each other, through shaking them. This idea could be ported to constitute a, albeit explicit, replacement for password-entry. Humans handle their gadgets in different manners, e. g. they grab and hold their mobile phones differently. The way people grab and hold their devices could constitute a cue for multimodal, biometric systems. It is measurable through acceleration sensors and could either happen explicit, i. e. through shaking the device to unlock it, or implicit, if the data is sufficient, which has to be proven by respective studies first. Even if such studies prove that it is not possible to use it as a cue for implicit authentication, shaking a device would constitute a less intrusive and user friendlier way of performing explicit authentication than password entry.

5 Conclusion

This paper gave an overview of implicit authentication on mobile devices. The current state of authentication on mobile devices was investigated and analyzed with the result that secret-knowledge and single-sign-on techniques are inadequate for state of the art services and applications accessible through mobile devices. Reasons for a switch to implicit techniques were given, including usability and security concerns. Possible cues and techniques for an implementation were reviewed and discussed. The resulting conclusion is that multimodal systems incorporating various biometrics with a possible escalation to explicit authentication techniques are worthwhile and can be implemented without the urgent need for new hardware.
The author hopes that his work can spur research in this area and he intents to expend further efforts to this topic.

References

[1] 3GPP. 3g security; security threats and requirements. http://www.3gpp.org/ftp/Specs/html-info/21133.htm, 2001. Accessed on 27.05.2011.
[2] Apple Inc. iPad. http://www.apple.com/ipad/, 2011. Accessed on 13.06.2011.
[3] BBC News UK. Huge surge in mobile phone thefts. http://news.bbc.co.uk/2/hi/uk_news/1748258.stm. Accessed on 27.05.2011.
[4] D. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *AAAI94 workshop on knowledge discovery in databases*, pages 359–370, 1994.
[5] L. Bianchi, D. Angelini, and F. Lacquaniti. Individual characteristics of human walking mechanics. *Pflügers Archiv, European Journal of Physiology*, 436(3):343–356, 1998.
[6] J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Multimodal biometric authentication using quality signals in

mobile communications. In *Proceedings of the 12th International Conference on Image Analysis and Processing*, pages 2–11. IEEE, 2003.

[7] R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):955–966, 1995.

[8] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.

[9] J. Campbell Jr. Speaker recognition: A tutorial. *Proceedings of the IEEE*, 85(9):1437–1462, 1997.

[10] S. Card, T. Moran, and A. Newell. Computer text-editing: An information-processing analysis of a routine cognitive skill. *Cognitive Psychology*, 12(1):32–74, 1980.

[11] K. Chan, Y. Moon, and P. Cheng. Fast fingerprint verification using subregions of fingerprint images. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):95–101, 2004.

[12] K. Chang, J. Hightower, and B. Kveton. Inferring identity using accelerometers in television remote controls. In *Proceedings of the 7th International Conference on Pervasive Computing*, pages 151–167. Citeseer, 2009.

[13] S. Cho, C. Han, D. Han, and H. Kim. Web-based keystroke dynamics identity verification using neural network. *Journal of Organizational Computing and Electronic Commerce*, 10(4):295–307, 2000.

[14] T. Clancy, N. Kiyavash, and D. Lin. Secure smartcardbased fingerprint authentication. In *Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications*, page 52. ACM, 2003.

[15] N. Clarke and S. Furnell. Authentication of users on mobile telephones-a survey of attitudes and practices. *In: Computers & Security*, 24(7):519–527, 2005.

[16] M. Corner and B. Noble. Zero-interaction authentication. In *Proceedings of the 8th annual international conference on Mobile computing and networking*, pages 1–11. ACM, 2002.

[17] I. Damousis, D. Tzovaras, and E. Bekiaris. Unobtrusive multimodal biometric authentication: The humabio project concept. *EURASIP Journal on Advances in Signal Processing*, 2008:1–11, 2008.

[18] Dropbox Inc. Dropbox. `http://www.dropbox.com`, 2011. Accessed on 13.06.2011.

[19] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin. Diversity in smartphone usage. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 179–194. ACM, 2010.

[20] M. Farah. Is face recognition special? evidence from neuropsychology. *Behavioural Brain Research*, 76(1-2):181–189, 1996.

[21] S. Furnell, N. Clarke, and S. Karatzouni. Beyond the pin: Enhancing user authentication for mobile devices. *Computer Fraud & Security*, 2008(8):12–17, 2008.

[22] D. Goldberg. Genetic algorithms in search, optimization, and machine learning. 1989.

[23] Google Inc. Google music. `http://google.music.com`, 2011. Accessed on 13.06.2011.

[24] R. Greenstadt and J. Beal. Cognitive security for personal devices. In *Proceedings of the 1st ACM workshop on Workshop on AISec*, pages 27–30. ACM, 2008.

[25] J. Hartigan and M. Wong. A k-means clustering algorithm. *Journal of the Royal Statistical Society C*, 28(1):100–108, 1979.

[26] IDC. Transition to the information highway era. *Information Industry and Technology Update*, page 2, 1995. Cited from Weiser et al., 1996.

[27] Y. Ijiri, M. Sakuragi, and S. Lao. Security management for mobile devices by face recognition. In *Proceedings of the 7th International Conference on Mobile Data Management*, page 49. IEEE Computer Society, 2006.

[28] M. Jakobsson, E. Shi, P. Golle, and R. Chow. Implicit authentication for mobile devices. In *Proceedings of the 4th USENIX conference on Hot topics in security*, pages 9–9. USENIX Association, 2009.

[29] G. Keizer. ipad lead shrinks as tablet sales grow. `http://www.computerworld.com/s/article/9207319/iPad_lead_shrinks_as_tablet_sales_grow`, jan 2011. Accessed on 27.05.2011.

[30] J. Kennedy and R. Eberhart. Particle swarm optimization. In *IEEE International Conference on Neural Networks, 1995. Proceedings.*, volume 4, pages 1942–1948. IEEE, 1995.

[31] G. Klir and B. Yuan. *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A. Zadeh*. 1996.

[32] J. Leggett, G. Williams, M. Usnick, and M. Longnecker. Dynamic identity verification via keystroke characteristics. *International Journal of Man-Machine Studies*, 35(6):859–870, 1991.

[33] R. Lemos. Passwords: The weakest link? hackers can crack most in less than a minute. `http://news.cnet.com/2009-1001-916719.html`, may 2002. Accessed on 27.05.2011.

[34] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. Makela, and H. Ailisto. Identifying users of portable devices from gait pattern with accelerometers. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05).*, volume 2, pages ii–973. IEEE, 2005.

[35] R. Mayrhofer and H. Gellersen. Shake well before use: Authentication based on accelerometer data. *Pervasive Computing*, pages 144–161, 2007.

[36] Microsoft Corp. Windows live. `http://explore.live.com/windows-live-essentials`, 2011. Accessed on 13.06.2011.

[37] R. Napier, W. Laverty, D. Mahar, R. Henderson, M. Hiron, and M. Wagner. Keyboard user verification: toward an accurate, efficient, and ecologically valid algorithm. *International journal of human-computer studies*, 43(2):213–222, 1995.

[38] M. Nauman and T. Ali. Token: Trustable keystroke-based authentication for web-based applications on smartphones. *Information Security and Assurance*, pages 286–297, 2010.

[39] M. Obaidat and D. Macchairolo. A multilayer neural network system for computer access security. *Systems, Man and Cybernetics, IEEE Transactions on*, 24(5):806–813, 1994.

[40] M. Obaidat and B. Sadoun. Verification of computer users using keystroke dynamics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 27(2):261–269, 1997.

[41] D. Patterson, D. Fox, H. Kautz, and M. Philipose. Fine-grained activity recognition by aggregating abstract object usage. 2005.

[42] M. Philipose, K. Fishkin, M. Perkowitz, D. Patterson, D. Fox, H. Kautz, and D. Hahnel. Inferring activities from interactions with objects. *IEEE Pervasive Computing*, pages 50–57, 2004.

[43] D. Reynolds. An overview of automatic speaker recognition technology. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2002*, volume 4, pages IV–4072. IEEE, 2002.

[44] J. Rokita, A. Krzyżak, and C. Suen. Cell phones personal authentication systems using multimodal biometrics. *Image Analysis and Recognition*, pages 1013–1022, 2008.

[45] M. Russinovich and B. Cogswell. Filemon. `http://technet.microsoft.com/en-us/sysinternals/bb896642`, 2006. Accessed on 16.06.2011.

[46] C. Sanderson and K. Paliwal. Identity verification using speech and face information. *Digital Signal Processing*, 14(5):449–480, 2004.

[47] S. Schroeder. Smartphone sales up 85% year-over-year. `http://mashable.com/2011/05/19/smartphone-sales-q1-2011-gartner/`, `http://www.gartner.com/resId=1688625`, 2011. Accessed on 27.05.2011.

[48] E. Shi, Y. Niu, M. Jakobsson, and R. Chow. Implicit authentication through learning user behavior. *Information Security*, pages 99–113, 2011.

[49] A. Shukla. Rising popularity of cloud computing in asian markets. 2009. Accessed on 13.06.2011.

[50] R. Snelick, M. Indovina, J. Yen, and A. Mink. Multimodal biometrics: issues in design and testing. In *Proceedings of the 5th international conference on Multimodal interfaces*, pages 68–72. ACM, 2003.

[51] I. Steinwart and A. Christmann. *Support vector machines*. Springer Verlag, 2008.

[52] M. Tamviruzzaman, S. Ahamed, C. Hasan, and C. O'brien. ePet: when cellular phone learns to recognize its owner. In *Proceedings of the 2nd ACM workshop on Assurable and usable security configuration*, pages 13–18. ACM, 2009.

[53] T. Tang, Y. Moon, and K. Chan. Efficient implementation of fingerprint verification for mobile embedded systems using fixed-point arithmetic. In *Proceedings of the 2004 ACM symposium on Applied computing*, pages 821–825. ACM, 2004.

[54] B. Taskar, C. Guestrin, and D. Koller. Max-margin markov networks. nips, 2003.

[55] D. Umphress and G. Williams. Identity verification through keyboard characteristics. *International journal of man-machine studies*, 23(3):263–273, 1985.

[56] M. Weiser and J. Brown. The coming age of calm technology [1]. *Xerox PARC. Retrieved July*, 8:2007, 1996.

[57] Wireshark Foundation. Wireshark. http://www.wireshark.org, 2011. Accessed on 16.06.2011.

[58] S. Yazji, X. Chen, R. Dick, and P. Scheuermann. Implicit user re-authentication for mobile devices. *Ubiquitous Intelligence and Computing*, pages 325–339, 2009.

[59] S. Zahid, M. Shahzad, S. Khayam, and M. Farooq. Keystroke-based user identification on smart phones. In *Recent Advances in Intrusion Detection*, pages 224–243. Springer, 2009.

Peripheral Interaction

Philip Nitsch

Abstract— Since Weiser's and Brown's [20] vision of Calm Technology, many approaches for designing Calm Technology have been published. The main topic of this paper addresses systems the user can interact with. Since this area of research is quite new, there are only a few approaches today. An essential task for developing such solutions is to understand the theories of attention. This paper provides an overview of relevant attention theories, such as the theory of selective and divided attention and sums up the approaches for creating interactive and ambient systems like Edge's Token System or Ishii's ambientROOM, harnessing the given theories. This theoretical background shall be used to give an overview about systems, where interaction takes place in the periphery of attention. Finally, a new system shall be introduced, which calculates the personal and team workload based on upcoming appointments and unread emails. It uses an object which shines in different colours depending on the workload and also gives the user the opportunity to set the personal workload manually. The whole system is connected to the Microsoft Office Communicator and Outlook.

Index Terms—Peripheral Interaction, Ubiqitous computing, Tangible Interaction, Perception Psychology, Attention Theory, Peripheral Embodied Interaction, Workload Calculation

1 Introduction

In the end of 1997, Weiser and Brown[20] first mentioned the so-called "Calm Technology". They realized that most common information technology during this time was exactly "the enemy of calm"[20], as it frequently overloads us with information. In contrast to this, there are technologies, which really encalm someone, e.g. a writing pen or a pair of shoes[20]. The question why one of these technolgies encalms someone while the other enrages someone led Weiser an Brown to the idea that it depends on the way how technology engages our attention[20]. "Calm technology engages both the center and the periphery of attention, and in fact moves back and forth between the two"[20]. Bakker et al.[4] mentioned, that "Calm Technology enables users to monitor information without specifically paying attention to it, while at the same time facilitating them to focus on it if it desired"[4]. Many technolgies have been invented making use of this phenomen, such as the Dangling String[20] or the ambientROOM[10], mostly the human skill to monitor something while not really attending to it, is not used for interaction with technology. Bakker et al. see "major opportunities for this skill to be leveraged in order to avoid information overload[4].

For example Darren Edge and Alan Blackwell were one of the first to develop such interactive systems on the basis of an analytical design process, where interaction takes place in the periphery of attention[7]. So how can technology be designed and what approaches have already been done, so that the interaction ostensibly takes place in the periphery of attention and moves to the center of our attention when desired? To answer this, it might be valuable to have a closer look on attention theories or perceptual psychology from the 19th and 20th century, most notably to understand the design of existing interactive systems and to support further design approaches for "peripheral interaction systems", which shall be the focus of this paper.

2 Attention Theory and Perception Psychology

Looking at our daily environment, everyone is surrounded by lots of different stimuli. An example could be concentrating on wirting a research paper in different environments, such as a rather quite library or a public place that has many intensive stimuli. Looking at this example it should be clear, that it is not possible to attend to all the stimuli which are around. Anderson[2] defined attention as "the cognitive process of selectively concentrating on one aspect of the environment while ignoring other things. Attention has also been referred to as the allocation of processing resources"[2]. James[11] also stated that attention can be devoted to stimuli that we perceive through our senses (sensorial attention), but also to cognitive processes (intellectual attention)[11].

Over the last decades, several theories and functions of attention have been developed. For this paper it is useful to have a closer look especially on the functions of divided and selective attention. Both functions are based on the fact, that nobody can "fully appreciate all that takes place at any one time"[13].

Selective attention is the process of selectively focusing the attention on one stimulus while intentionally ignoring others[18].

Divided attention is the process in which we carefully divide our attentional resources over multiple attentional tasks at once[18]. The next two chapters will focus on these two functions of attention.

2.1 Selective Attention

Psychologists stated that there must be serial bottlenecks within human information processing. In this state it is no longer possible to process everything parallel[3]. Solso[17] compared selective attention to the situation when lighting up a dark room with a torch: Things of interest are lightened up while others stay in the dark. Psychologists are unsure about the time these bottlenecks of human information processing appear: Either, they appear before attending to a stimulus or after attending to a stimulus, yet before thinking about it[3]. In general, these stimuli must be processed before they can be perceived, which leads to the fact that it is possible to dinstinguish a friend's voice from the voice of a passerby[4].

Referring to the bottleneck, psychologists differentiate between the early and late selection theory, depending on the time when these bottlenecks first appear[3].

2.1.1 Early Selection Theory

Early selection models maintain that the main problem facing the organism is the richness and complexity of information presented to the sense at any one time, which may confuse and overload high-level processing mechanisms. Selective attention is linked to a filter which attenuates and excludes irrelevant information from further analysis of meaning or storage in long-term memory[8], as shown in figure 1. The arrows represent words or messages someone could be attratcted to. Before words or messages are filtered by the selective filter, they pass the senses. After passing the selective filter, "words in the rejected messages are not remembered"[4].

2.1.2 Late Selection Theory

Late selection models have contested the assertions of early filtering by designing experiments and tasks that demonstrate high-level analysis of all simultaneous information, both irrelevant and relevant. According to these approaches, the bottleneck and confusion are on the

---

- *Philip Nitsch is studying Media Informatics at the University of Munich, Germany, E-mail: nitschp@cip.ifi.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

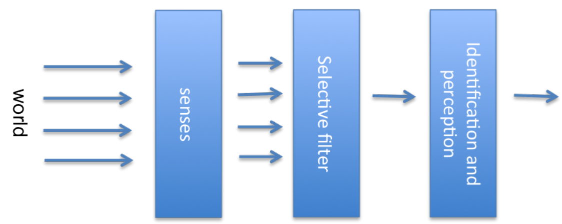


Fig. 1. Simplified illustration of Early Selection Theory, adapted from Sternberg[18]

response tendencies that are instigated at one time[8]. In addition to this, Moray[12] showed that words of interest, such as your own name for example, are noticed when present in a rejected stream. This shows that, in contradiction to the early selection theory, the meaning of some words can be extracted before the whole channel is selected as a "channel of interest", as shown in figure 2.

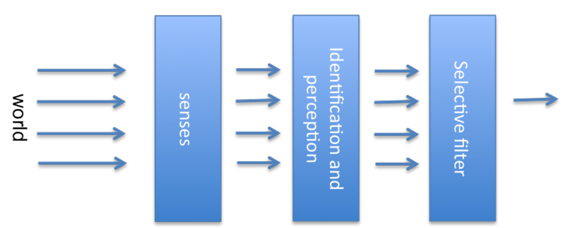


Fig. 2. Simplified illustration from Late Selection Theory, adapted from Sternberg[18]

2.1.3 Attenuation Theory

In contradiction to the early- and late selection theories, where different stimuli are totally eliminated through a filter, the filter in the attenuation theory attenuates the unattended material. "Attenuation is like turning down the volume so that if you have four sources of sound in one room (TV, radio, people talking, baby crying) you can turn down or attenuate three in order to attend to the fourth"[1]. That means that the unattended material appears to be lost. However, if an unattended channel includes a word of interest (such as your own name), you will probably hear it because the material is not totally lost. In general, Treisman[19] found "that when words in the rejected channel are relevant to the information in the attended channel, they conciously or unconciously influence the perception of the information in the attended channel[4]. She suggests that the selection process is additionally influenced by the relevance of information in the incoming channel[4]; this is called priming. As an example for this study, someone could imagine the situation when talking about last night's football game with a friend. In this case, the names of the top players or the name of the stadium could be primed. When a passerby says one of these primed words, one is more likely to recognize and suddenly attend to this previously-unattended channel. Someone could also be primed not only for words of interest, such your own name or topics being relevant to the current attentional focus, but also to topics which are in the back of our mind[4].
In contradiction to the early selection theory, where recognition of relevant words in rejected channels is unclear and the Late Selection theory the Attenuation Theory can explain the results of many shadowing experiments[4].

2.2 Divided Attention

The mentioned theories of selective attention mostly concern our sensorial attention or just the auditory modalities[4]. However there are also studies which take a broader approach as they try to explain how someone can perform multiple attentional tasks at one time. Sternberg[18] defined the so-called divided attention theories as the assignment to a limited amount of mental resources over different activities[18]. Different tasks require different mental effort. Looking at the fact that a human being has limited mental resources, it is not possible to participate any number of tasks. However, the amount of mental resources someone needs for a task decreases with practice and experience[4]. Therefore it is possible to do several tasks, which each require less mental resources or effort at one time. While reading a book, it is nearly not possible to do another task in parallel, because the task of reading a book requires the most mental resources. This process, where only one task can be performed at once, is called "controlled process", whereas highly trained processes, such as walking, can be done in parallel with other processes and are called "automatic processes"[4].

2.3 Center and Periphery of Attention

For this paper it is quite useful to define what is exactly meant by the center and the periphery of attention. The technological solutions and approaches, which will be mentioned in the following chapters, try to make use of the different psychological theories of attention as the information flow between technology and a human being should take place in the periphery of attention.
Weiser and Brown[20] explain the word periphery as "what we are attuned to without attending to explicitly"[20].
Bakker et al.[4] see this not very explicit definition of the periphery of attention as quite "important to inform the design process" of the mentioned technical solutions[4].
Furthermore, Bakker et al.[4] define the center of attention as "the one activity that most resources are allocated to"[4].
Grounded on these two definitions, the following figures 3 and 4 show different tasks and the mental resources needed to participate, the total amount of mental resources and the "place" of attention where the different tasks are located.

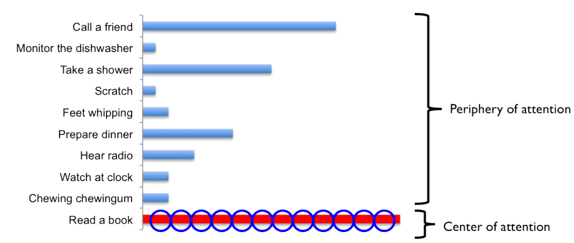


Fig. 3. Center and Periphery of Attention, Controlled Process, adapted from [4]

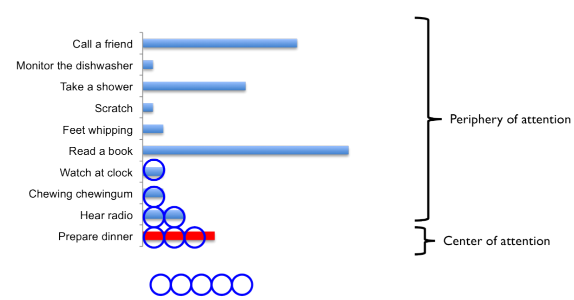


Fig. 4. Center and Periphery of Attention, Automated Process, adapted from [4]

The first example illustrates a task which requires a high amount of mental resources and therefore belongs to the "controlled processes". While reading a book, all the mental resources are allocated to this task and there are no resources left to use for other tasks. This task takes place in the center of attention.

Having a closer look on the second example, the task in the center of attention requires little mental resources and therefore belong to the "automated processes". Hence it is no problem to perform other tasks at the same time. All tasks except the task in the center of attention take place in the periphery of attention. Tasks located closer to the center of attention are more likely to shift there[4]. "At any moment one may (...) be attracted to pay more attention to the radio as one's name is suddenly heard in that stream. This would change the resource demand of some activities as well as the priming of certain activities"[4]. The attentional process therefore is highly dynamic. For example, while watching a football game and talking to a friend at the same time, both activities will move between the center and the periphery of attention.

3 Examples for Ambient Information Systems and Peripheral Interaction

In this section, systems which make use of the different attention theories shall be mentioned. The first attempts in this research area were made by Weiser and Brown[20]. They envisioned the idea where the interaction with computers vanishes in the background, the so-called "Calm Technology". Famous examples for Calm Technology are the "Dangling String" or a simple window, which connects an office to the hallway[20] and therefore gives someone informative clues like somebody is working late, when a light shines out into the hallway[4].

3.1 Ambient Information Systems

Based on Weiser's and Brown's vision, research was led to "Ambient Information Systems". Information conveyed via calm changes in the environment made users more able to focus on their primary work[14]. For a better understanding, the most used terms for describing these systems like "Peripheral Display, "Ambient Display" or "Notification System" should be untangled. Not all of these terms describe exactly the same system, but as the focus of this paper is not on Ambient Systems we claim the following dependencies:

- All ambient displays are peripheral displays,
- Some notification systems are peripheral displays (some notification systems are not peripheral but are instead the object of focused work and attention)[14]

3.1.1 Dangling String

The Dangling String was invented by Weiser and Brown[20]. The system is quite simple: It consists of an eight foot plastic string mounted on an electric motor which is connected to an ethernet cable. Whenever a bit of information is sent through the cable, the motor makes a little twitch which makes the string whirl around. Therefore the string heavily whirls around when the network traffic is high and, in contrast to this, only moves little when the traffic is low[20]. After getting used to this systems one can easily imagine that network activity can be monitored without directly paying attention to the string and therefore moves to the background of attention (not only because it makes sound while whirling around).

3.1.2 ambientROOM

Another approach for augmenting an environment with technology where information processing happens in the background of awareness is the so-called ambientROOM[10]. Ishii et al.[10] built a huge steelcase as a detached office room. Different areas inside this cube were augmented with different technologies, as figure 5 shows. The following list explains the most important interfaces placed in the ambientROOM.

- Water Ripples: This display allows the user "to have some awareness of the activity of a distant loved one"[10]. Ishii et al.[10] used a so-called "phicon" (physical icon) which vibrated when any motion was detected (they monitored a hamster inside a hamster wheel). This phicon was placed in a water tank with built-in lights shining out and then producing rippling shadows on the ceiling when the hamster moved.

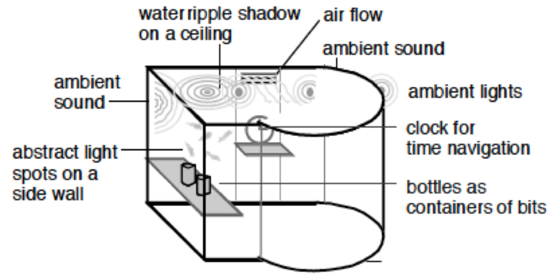


Fig. 5. Overview about ambientROOM and built-in interfaces, adapted from [10]

- Light Patches: This display also allows the user to monitor the presence of others in a working area. Human movement is monitored via electric field sensors and, according to the amount of movement, illuminated patches are projected onto a wall. Unless a sudden change of human activity happens, this "active wallpaper" is rarely noticeable[10].
- Natural Soundscapes: Through this subtle soundtrack of birds and rainfall connected with variations of the room lighting, approximate quantities such as unread emails or the value of a stock portfolio can be monitored[10].

3.2 Interactive Systems

As we have already seen, there are many approaches for communicating information in the periphery of attention like the mentioned ambient systems. However, most of these solutions do not give the user a chance to interact with the system. Darren Edge and Alan Blackwell[6] were among the first ones to build a system with which the user could interact with directly. Edge and Blackwell introduced a concept called "peripheral tangible interaction". They defined it as "episodic engagement with tangibles, in which users perform fast, frequent interactions with physical objects on the periphery of their workspace, to create, inspect and update digital information which otherwise resides on the periphery of their attention"[6, 7]. They also see the "combination of calm peripheral interactions and engaging tangible interactions"[7] as the essence of their understanding of peripheral tangible interaction.
Hausen and Butz[9] mentioned an "interaction style that is carried out alongside the user's current primary task without asking for their full attention"[9] and created two systems where interaction takes place in this way.

3.2.1 Token Systems by Edge and Blackwell

Edge and Blackwell[7] offered an "alternative perspective on the use of tangibility in interaction, in which meaning is created not through precise manipulations of a computationally-interpreted spatial syntax, but through imprecise interactions with independently meaningful, digitally-augmented physical tokens"[7] in their paper, see figure 6.

These physical tokens represent unfinished tasks, shared documents and contacts. In general, the domain of their TUI (Tangible User Interface) is personal and group task management[7]. While someone is a member of a team, the system can be used to "track and update task progress and dependencies between tasks"[7]. Their aim is to support "fast and fine-grained management of group activity"[7].
The system consists of various items: An interactive surface is placed on the left side of an office desk. The mentioned tokens can be placed on this surface and their position and identity can be recognised by a camera mounted above the surface. After a token is placed on the interactive surface, the tokens are visually augmented through "conventional display halos displaying their attributes"[7]. On the top edge of the interactive surface is a timeline-based calendar. Depending on the task-tokens placed on the surface the calendar shows planned comple-

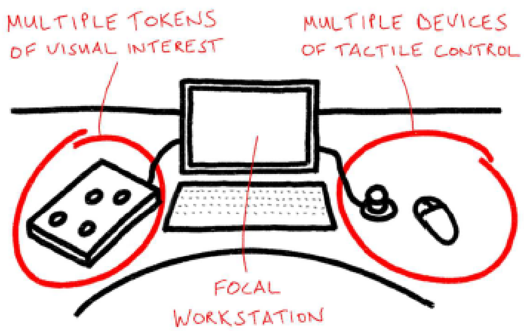


Fig. 6. Interactive Token System, adapted from [6].

tion dates and "visualises the effects of manipulating the factors that impact upon workload"[6]. These factors are:

- Estimated work time remaining for this task
- Sharing time between overlapping tasks
- Number of hours dedicated to task work in one or all working days.

Users therefore can adjust different attributes of a task(-token) like the completion date, a to-do list and many more. It is always possible to compare the user's estimates to reality. In general, the coloured halos belong to different tokens: The following figures 7 and 8 show the design of the surface and the corresponding tokens.

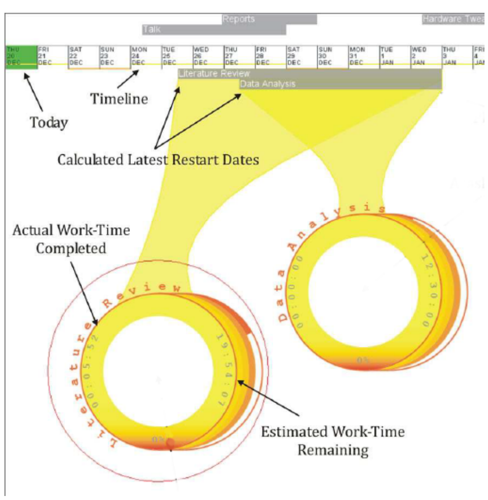


Fig. 7. Interactive Token System with yellow and red halos, adapted from[6]

- Yellow and red halos: These belong to task tokens, provide visualisation of multiple task attributes, and permit rapid, low attention interaction[6].
- Simpler yellow and blue halos: belong to document tokens, act as a physical shortcut or hyperlink to an underlying document, names of other team members who are working on the document can be monitored[6].
- Green and blue halos: Belong to contacts or team members. While placed on the surface, the user's workload is displayed over your own calendar[6].

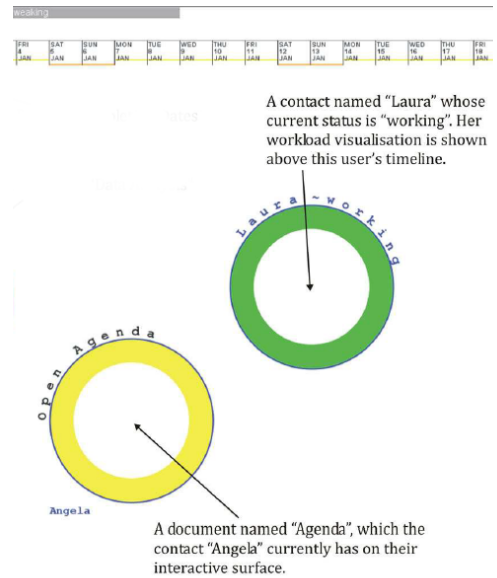


Fig. 8. Interactive Token System with simpler yellow-blue and green halos, adapted from[6]

So how could a typical storyboard for using this system look like? A user takes a look on the token placed on the interactive surface and recognizes that this task has to be finished on the weekend, which seems unlikely considering his remaining workload. So he decides to change the completion date to next week. Therefore the user nudges the token upwards toward it's completion date and finally sets the new completion time with the control knob placed on the right side of the user's keyboard.

3.2.2 Peripheral Embodied Interaction

Hausen and Butz[9] expanded Edge's definition of peripheral tangible interaction to peripheral embodied interaction by considerung further physical capabilities[9]. Based on this, two experimental prototypes were built. Hausen's and Butz's[9] understanding of peripheral embodied interaction is in line with Dourish's, who describes it as the attempt "to move computation and interaction out of the world of abstract cognitive processes and into the same phenomenal worlds as our other sorts of interactions"[9, 5].
Having a closer look on everyday situations, someone can identify many small activities a human being can do in parallel to primary activities, without setting the focus of attention on it. For instance, it is no problem having a conversation with someone while tying shoes. Hausen and Butz[9] use the activity of moving a cup out of the way while talking to somebody as an example for a activity that can be carried out "with a flick of the wrist"[9]. In contradicition to this, while working on a PC, even small activities require a "context switch, precise pointing or exact knowledge about certain key presses". Hausen and Butz[9] argue that "especially simple things, which do not belong to the current primary task (e.g., typing a text), but still matter and require interaction (e.g., setting the status in an instant messenger) will benefit from new forms of embodied interaction"[9]. In general, (peripheral embodied) interaction can be categorized in five different design dimensions, as the following list shows:

- Explicitness: The normal way a user interacts with a computer, e.g. with mouse and keyboard, is titled explicit interaction. In contradiction, implicit interaction means "an action performed by the user that is not primarily aimed to interact with a computerized system but which such a system understand as input"[16].
- Input Mode: Many different input modes can be chosen to provide the system with information: Someone's gaze can be

tracked, speech can be recorded, hand-gestures or gestures with other parts of the body can be detected[9].

- Granularity: Depending on the chosen gesture, a system can encode different numbers of commands: "Glancing at an object encodes two levels - looking or not looking at it"[9] while speech input provides the opportunity to distinguish between an infinite number of commands and therefore has a much finer granularity.

- Privacy: While using gestures to interact with a system, one can easily observe what someone is doing, because this information can be gathered without looking at a display. That is why it is important to keep this in mind when designing such a system, especially for sensitive data.[9].

- Proximity: Depending on the chosen input method it is possible to keep a certain distance from the system while interacting with it (e.g., speech). In constrast, while using gestures it is necessary to stay close to the tangible[9].

The first prototype built by Hausen and Butz[9] is the so called "Ambient Appointment Projection" as shown in the left picture of figure 9. This prototype uses a "spiral visualization of the overall time flow of upcoming appointments"[9], projected on top of the users work desk. Everytime an appointment comes close, the spiral starts pulsating so that the user is reminded of the event. After the spiral starts pulsating, the user has two different options for interacting: If the user is interested in the appointment details, he makes a wiping gesture towards himself and the details are displayed in a balloon tooltip on his display. If he is not interested in the appointment details the user makes a wiping gesture away from him which makes the spiral stop pulsating. The advantage of these systems is that the user gets "not disrupted as forcefully as by state-of-the-art reminder pop-ups"[9]. User do not need to directly focus on these gestures because both wiping towards or away from the user is a quite natural movement.

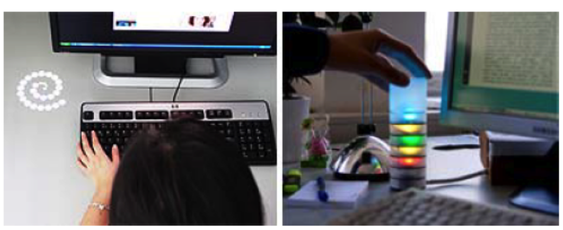


Fig. 9. Ambient Appointment Projection (left), Tangible Presence Indication(right), adapted from[9]

The second prototype built by Hausen and Butz[9] is the so-called "Tangible Presence Indication" as shown in the right picture of figure 9. This prototype consists of different levels which are all build on top of each other. The result is a small, cylindrical object which is connected to Skype. This object shows "presence information about the user (biggest and topmost level) and selected contacts (other levels)"[9]. Besides the normal skype-statuses like "available", "away" and "do not disturb" the object encodes customized statuses of the contacts in someone's contact list. Users are able to change their own statuses via the topmost level: Turning this level changes the status; pressing it down sets the estimated time someone stays in the chosen status. Therefore no context-switching on the screen is necessary, as turning or pushing a button is "very natural in the physical world"[9].
Along with the mentioned five design dimensions, figure 10 classifies the two prototypes presented in this chapter. Both systems have in common, that they use explicit interaction, operate on personal data and need to be nearby for interaction whereas the appointment projection interprets gestures and the presence indication supports interaction through object manipulation[9].

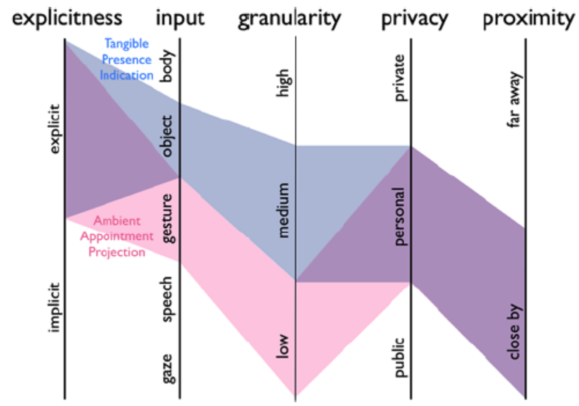


Fig. 10. Classification of both prototypes along the five design dimensions, adapted from[9]

3.2.3 Picture Navigation using an Ambient Display and Implicit Interactions

Ryu et. al[15] used implicit interaction and ambient displays for creating a prototype, where the user can interact with it depending on the distance the user stays away from the ambient display. The ambient display consists of an RFID reader, ultrasonic sensors, a touch panel, a 15" TFT-LCD and LEDs[15]. In general, there are three different zones "corresponding to the proximity of the user to the display"[15]. The closer the user gets to the display, the more detailed and specific the displayed information (in this case photos) gets. While staying outside of all the zones only a frame and black-and-white pictures are displayed.

- Appealing Zone: When the user comes closer (4m -7m), the ultrasonic sensors recognize the user's presence and the TFT-LCD displays coloured pictures. Nevertheless, the display does not demand excessive attention[15]. The user is also able to interact via a mouse-like pointing device when entering the appealing zone. Therefore a high-level menu is displayed.

- Interesting Zone: In the so called "Interesting Zone"(1m - 4m) the ultrasonic sensors and the RFID reader recognizes the user via a small tag the user wears as a necklace which leads to a 3D-Photo as long as 3D information is provided by the picture diplayed on the screen. Depending on the closeness of the user to the display, the picture zooms in or out.

- Communication Zone: While the user stays very close to the display (1m), he can explicitly interact with the photos via touch-screen.

Figure 11 shows the different zones.

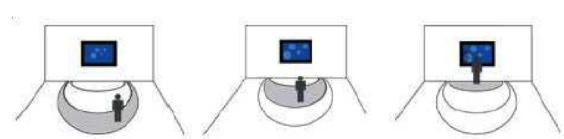


Fig. 11. Appealing Zone, Interesting Zone, Communication Zone, adapted from[15]

The authors see their approach for peripheral interaction as an opportunity for users "who are not familiar with the operation of digital photos"[15].

4 Future possibilities

Nowadays, most communication within of a large company is done via email, which I can confirm with my own experiences. This leads to a

huge amount of emails every employee has to send and to answer on a daily basis. Because of this heavy traffic, many emails remain unanswered beacause not everyone can answer every email. While working at various big and medium-sized companies, it was quite normal that everyone wrote emails, even if the content was unimportant. Many employees left a lot emails unanswered which resulted in an absolutely overcrowded inbox. In this case, the employees lost sight of emails with really important concerns. Also, the most email-programmes do not present the information about new or incoming emails in a calm way as they use pop-ups or sounds which makes the user to focus on it.
In this section I would like to summarize ideas, how a prototype, which allows the user to focus on the primary task and not on the daily "email-flood" may look like. In general, there are different points I would like to consider when trying to handle the mentioned problems.

- How can it be arranged, that users write less or only really important emails?
- How can the workload of users be monitored?
- How does an interaction in the periphery of attention with that system take place?

When addressing these questions the ideas stated in Hausen's and Butz's[9] paper can be considered as a useful basis.
I would like to make use of an object quite similiar to the object Hausen and Butz[9] used for the "Tangible Presence Indication". In contrast to this, the object is not connected to Skype but to the Microsoft Office Communicator. Along with Hausen's and Butz's[9] ideas, the object should, in this case only, consist of only two different levels: The first, topmost level indicates my personal workload. The second level, which is much bigger than the first level, shows the workload of the team I work with. Depending on my personal workload and the workload of my team-members the object shines in a color between green (=nothing to do) and red (=busy). To set the workload color, the user can manualy turn the topmost level. The color of the object switches directly so the user can stay focused on the primary task and "monitor" the correct color in the periphery of attention. Using the default settings of the object and the software behind, the workload status is provided through an algorithm which calculates a certain number depending on the amount of unread emails, upcoming appointments and open tasks, all gathered from Microsoft Outlook. Then, the running software sends the information to the object which directly starts shining in the corresponding color. Once the user manualy selected a workload status he can switch back to the automatic mode through pressing the built-in button on the topmost level. The second level of the object, which represents the workload of the whole team, can not be set manual and instead gathers the workload information from my team members and sets the corresponding color automatically. This team workload level only shows the team's workload while working in an "general environment". As soon as someone chooses a single contact via the address book the color of the object's second-level switches to the workload color of the choosen contact. Thus, there are two different levels which can be displayed through the object's second level:

- Most granular level: The general workload of my team.
- Fine granular level: The personal workload of a single user when choosing his contact from the address book or writing his name in the address bar of my email programme.

While using this object and the shining colors there is also a nice side effect: Colleagues sitting in the same office or next to me can monitor my personal workload directly when the object stands on a good visible place nearby to me so that I can still interact without changing my position. Figure 12 shows the different parts of the mentioned system.
The system should be connected to the Microsoft Office Communicator. A great opportunity for lowering the email traffic is that the

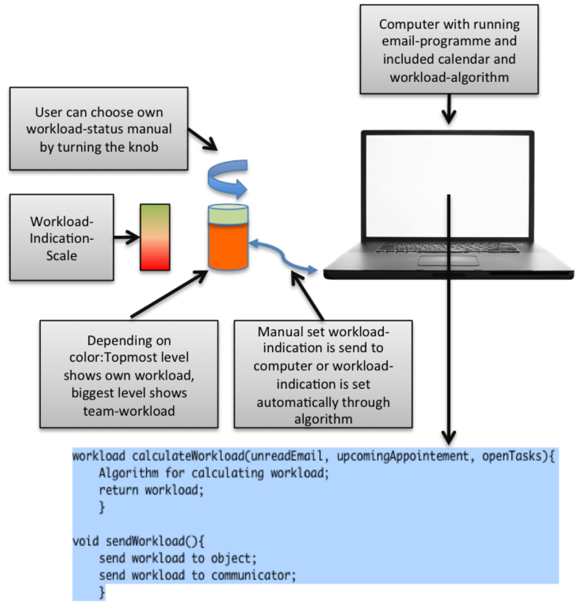


Fig. 12. Ambient Workload System

personal status set in the Microsoft Office Communicator is linked with the personal status in Microsoft Outlook. Giving an example, a colleague who wants to write an email to me has the following options:

1. He recognizes the team workload color of his object in the periphery of his attention and then decides to write an email or not. Someone should be more likely to write an email when the status is not red.
2. He switches to Outlook, types the email and then sees the addressee's workload status Outlook gathered from the Communicator which is displayed directly in Outlook and also on the second level of his colored object. Again the colleague can choose whether to write the email or not.
3. If my personal status is red and my colleague decides to write the email anyway, the object's second level colored in my personal workload-status starts blinking for three times and the user is prompted to confirm the delivery.

In some cases it is better for the user to set his workload status manualy, for example when an appointment got canceled but still is available in the calendar of the participants. Given this situation, the workload status is busy (red-coloured object) though the appointment does not take place, which leads to a wrong workload status. In this case, the user could delete the appointment from his calendar or set the workload status manualy by choosing the status "available" as he turns the topmost level of his object. In both cases, the workload status should change immediately. Under consideration of Hausen's and Butz's five design dimensions[9] the following classification seems to be meaningful:

- Explicitness: The system uses both explicit and implicit interaction.
- Input: The interaction takes place with an object.
- Granularity: The system's granularity is located in the middle area.
- Privacy: The system operates on personal data, nevertheless any member of my team gets informed about my current workload.
- Proximity: The objects need to be nearby a user, therefore the proximity is chosen as "close by"

The next steps for evaluating these ideas are to build such a prototype and to check whether there is the possibility to change the Communicator's status via other software. Also, user experiments in an normal office environment should be arranged and evaluated. Additionally, a good algorithm for calculating the workload status considering a good mixture and prioritization between unread emails, open tasks and upcoming appointments has to be found. In my opinion, these solutions would help users to think about the neccessity to write an email or not, which probably could lower the email-traffic.

5 Conclusion

In this paper, literature on calm technology was reviewed and, for a better understanding of system design, different attention theories and cognitive psychology were presented. Systems, which make use of the mentioned theories were introduced, including ambient systems without any user interaction and interactive systems such as Edge's Token system or Hausen's and Butz's prototypes of peripheral embodied interaction.

In general, the research on the mentionend interactive systems is a quite new area, for which reason there are only few systems which deal with interaction. I also tried to explain another field were interactive systems could make sense and would led the user to focus best on the primary tasks.

References

[1] Treisman's attenuation model. `http://www.simplypsychology.org/attention-models.html`, 2011. visited 11.06.2011.

[2] J. Anderson. *Cognitive Psychology and its Implications, 6th Edition*. Worth Publishers, 2004.

[3] J. Anderson. *Kognitive Psychologie, dt. Ausgabe*. Spektrum, 2007.

[4] S. Bakker, E. van den Hoven, and B. Eggen. Design for the periphery. In *EuroHaptics 2010: Proceedings of the EuroHaptics 2010 Special Symposium*, pages 71–80, 2010.

[5] P. Dourish. *Where the action is: The Foundations of Embodied Interaction*. MIT Press, 2011.

[6] D. Edge. Tangible user interfaces for peripheral interaction, technical report, 2008.

[7] D. Edge and A. F. Blackwell. Peripheral tangible interaction by analytic design. In *Proceedings of the Third Conference on Tangible and Embedded Interaction*, pages 69–76, 2009.

[8] M. W. Eyseneck. *The Blackwell Dictionary of Cognitive Psychology*. Blackwell Reference, 1999.

[9] D. Hausen and A. Butz. Extending interaction to the periphery. In *Workshop Embodied Interaction: Theory and Practice in HCI. In conjunction with 29th ACM International Conference on Human Factors in Computing Systems*, pages 1–4, 2011.

[10] H. Ishii, C. Wisneski, S. Brave, A. Dahley, M. Gorbet, B. Ullmer, and P. Yarin. ambientroom: Integrating ambient media with architectural space. In *Conference Summary of CHI 1998*, pages 173–174, 1998.

[11] W. James. *The Principles of Psychology*. Dover Publications, 1890.

[12] N. Moray. Attention in dichotic listening: Affective cues and the influence of instructions. *The Quarterly Journal of Experimental Psychology*, 11(1):56–60, 1959.

[13] D. Norman. *Memory and Attention: An Introduction to Human Information Processing*. John Wiley and Sons, 1976.

[14] Z. Pousman and J. Stasko. A taxonomy of ambient information systems:four patterns of design. In *Proceedings of the working conference on Advanced visual interfaces*, pages 67–74, 2006.

[15] H.-S. Ryu, Y.-J. Yoon, M.-E. Lim, C.-Y. Park, S.-J. Park, and S.-M. Choi. Picture navigation using an ambient display and implicit interactions. In *Proceedings of the 2007 conference of the computerhuman interaction special interest group CHISIG of Australia on Computerhuman interaction design activities artifacts and environments OZCHI*, pages 223–226, 2007.

[16] A. Schmidt. Implicit human computer interaction through context. *Personal and Ubiquitous Computing*, 4:191–199, 2000.

[17] R. Solso. *Kognitive Psychologie, dt. Ausgabe*. Springer Medizin Verlag, 2005.

[18] R. J. Sternberg. *Cognitive Psychology*. Itps Thomson Learning, 2005.

[19] A. Treisman. Verbal cues, language and meaning in selective attention. *The American Journal of Psychology*, 77(2):206–219, 1964.

[20] M. Weiser and J. S. Brown. *The Coming Age of Calm Technology*. Springer-Verlag, 1997.

Peripheral Interaction in the Digital World

Lorenz Schauer

Abstract— In the physical world, people are confronted with a ton of information all day long. So it is necessary that they can handle more than one information at the same time by dividing their attention into a centric and a peripheral part. Humans are instinctively able to perform actions in their periphery while concentrating on another task in the center of the attention (e.g. drink a cup of tea while reading a book). Due to the mass of information in the digital world, the same problem is caused. Thus, new forms of user interfaces which are distinct from mouse and keyboard become necessary to support the advantages of human's peripheral interaction. This paper describes the different techniques and ways of how peripheral interaction can be realized in a digital environment. On the one hand, it will be considered how digital information can be sent in a peripheral way. Ambient information will play a huge role in this context. On the other hand, it will focus on those actions which can be performed in the user's periphery and captured by a computer system. Several prototypes which were developed to investigate different possibilities of peripheral interaction will be presented and discussed at the end.

Index Terms—Peripheral interaction, ambient information, attention, tangible interaction

1 Introduction

Information can be found everywhere and every time in our life. We receive information all day long and we interact with it on different ways depending on the importance of that information.
However, sometimes people receive so much information at the same time that they have to prioritize it. In this case, humans' psychology manages instinctively which information attracts the main attention and which one is not that important so it can be handled in a minor way. This is an important fact of humans' psychology and a central point for the theoretical background of this paper why peripheral interaction is possible. Bakker et al. [2] name in this context "the center of the attention" and "the periphery of the attention". Humans instinctively act like this: They concentrate on one thing and are able to interact with another object without losing the main attention. For Bakker [1] this means e.g. that a person is able to read a book with the center of the attention and drinks a cup of tea with the periphery of the attention. This is an important advantage of how humans can interact and handle more than one information at once.

Thus, people in the physical world already interact in a peripheral way. In the digital world a user receives very often more than only one information at the same time (e.g. get a signal of an incoming email while reading a website) and so it rather seemed quite natural and easy if he also could react to it in a peripheral way. But how could this be afforded in a digital context? How is it possible to receive digital information and react on it in a peripheral way while concentrating on a screen and interacting with mouse and keyboard? What kind of actions or possibilities does a user have to interact peripherally with a machine? And what are the advantages of these actions in the periphery? Or are there even disadvantages we have to take into account?
New techniques and systems are required that offer other forms of user interaction, far beyond the typical interfaces like mouse or keyboard. These techniques have their seed in the so called ubiquitous computing in which computers are embedded in the human's natural movements and interactions with his environments [13]. One example in this context are the tangible user interfaces which are a new type of interfaces that connects the digital and the physical world [17].
This paper will give some answers on these research questions. Section 2 will explain the theoretical background about the center and the periphery of the humans' attention. After that digression, section 3 shows both ways of user's interaction which includes a passive part (getting information in subsection 3.1) and an active one (sending information in subsection 3.2).
However, the focus will always be on the active part. Before some general aspects can be discussed in chapter 5, a few prototypes which play a huge role about tangible user interfaces and peripheral interaction will be presented in section 4. The paper will end with a short conclusion and a presentation of some future steps (section 6).

- *Lorenz Schauer is studying Media Informatics at the University of Munich, Germany, E-mail: name@campus.lmu.de*

- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

2 Human's Form of Attention and Interaction

According to the Sternberg's divided attention theory [18] each human being has an amount of attentional resources which he can assign to some potential activities. If an activity needs many resources then there is not many space for other activities left and we speak about an high attentional task e.g. reading a website. If an activity does not allocate many resources, then the human is able to handle more of such low attentional activities at the same time. This is one reason why a person is able to interact in a peripheral way.

2.1 The Center and the Periphery of the Attention

As mentioned in the introduction, Bakker et al. [2] use the Sternberg's divided attention theory and speak about the center and the periphery of the attention. They explain that most of the resources are allocated in the center and in the periphery there are just a few of them. So the main concentration of a human is defined by the center of his attention (e.g. reading a book or driving a car). However, the important part for our purpose is the periphery of the attention where a person can also perform low attentional tasks (e.g. drink a cup of tea or switch a channel of the car radio). The periphery consists of all those potential activities which are not in the center and which do not allocate many resources.
Weiser and Brown [19] use the expression periphery in their paper of "Designing Calm Technology". This work is very important and "the field of Tangible User Interface research has its roots in Weiser's vision of ubiquitous computing" [5]. So Weiser and Brown "use periphery to name what we are attuned to without attending to explicitly" [19]. This statement confirms Bakker's theory of the periphery of the attention.
It becomes clear that human beings are able to receive information in the periphery while focusing on something else. This focus describes the objects which are in the center of the attention. Weiser and Brown explain that the humans' attention can immediately switch between the center and the information which is in the periphery. This happens when some peripheral information appears, disappears or changes. For example when we drive a car, we are concentrated on the road but not on the sound of the engine which is just in our periphery. But when this sound changes or a special noise appears, the attention switches

immediately, telling us that something might be going wrong. So this explains why the periphery of the attention is very important and the borderline in the divided attention theory can be easily passed. However, which conclusions can there be drawn for human interaction and how dose the attention influence its form?

2.2 Interaction in the Periphery

People are able to interact in the center of the attention (e.g. driving a car) and they can also receive information and handle them in the periphery of the attention (e.g. listen to the radio while driving). To be able to interact in the periphery, the action should not require many resources as we can see in the divided attention theory [18]. Many tasks can be performed at once when they do not need a high mental effort. And this mental effort decreases with practice and experience [20]. For example when a child learns to walk it costs a lot of mental effort and this activity allocates many resources. But for an adult who already knows to walk it almost requires no mental efforts and so he is able to do anything else while walking [2]. Small or short movements which are very common like certain gestures (we use everyday) are also low attentional activities and can be used for a peripheral interaction. So an action which can be performed in the periphery has the following characteristics: It has to be a small, short and an intuitive activity which is very common in the human's life [18]. In the car example, a peripheral interaction would be to switch a radio channel with a short, common movement of one finger while driving the car. So there are a lot of possibilities to act in the periphery of the attention. Those which are suitable for a digital environment will be described in the following chapter under subsection 3.2.
The next question is, what does peripheral information mean and how can it be classified in a digital context?

3 Information in the Periphery

Information which is in the periphery of a person "is anything but on the fringe or unimportant" [19]. The data which a human connects to that information can be very important. In the car example a traffic light just switches to red, but the information of this signal is essential for the driver's life. So the information in the periphery can mean a lot for the human although he just recognizes it on the fringe and it does not normally disturb the central attention. The same thing happens with information a person sends in a peripheral way. The action itself is short and intuitive as mentioned in the previous chapter, but the information for a computer in a digital environment can be much more (e.g. a little gesture can signalize that a request will be accepted).

Bakker et al. explain that the periphery is everything which is not in the center of the attention [2]. That means in a digital context, that all information which does not come from the screen, because the screen is in the user's center of the attention [19], is firstly allocated to the user's periphery of the attention. Thus, this is the type of information we want to look at.
To classify this peripheral information in a digital context it is important to consider the direction of the information. If the user receives information he takes a passive part. In this case, ambient information will play a huge role and the next subsection will show what kind of information can be used in this context. When the user sends information he takes an active part. It is one of the research questions to see what kind of actions can be performed in a peripheral way to send digital information. Thus, this part describes the main focus of that paper and it will be examined later in subsection 3.2.

3.1 Getting Information

Information which does not come from the screen of a computer must find a different way to be received by the user. Comparing the other human senses (despite from the visual sense that is approached concerning getting information via a screen), the audio sense seems quite a good alternative.

3.1.1 Audio Information

Although audio or sound signals are one form of the so called ambient information [21], they are named here at first as a special and important type for the communication of peripheral information. Bakker mentions that sound plays a huge role to monitor events in the periphery of the attention and it is often used for peripheral information [1]. Ishii et al. use natural sounds of a rain forest for the communication of information to the periphery of the user's awareness in their so called "ambientROOM" [12]. They say, that sound is one example of an additional source which is monitored in the background. This shows that sound would be an alternative way to get peripheral information from a computer system without using the screen.
But a communication which just consists of sounds to transport information has one disadvantage: In the beginning, the user has to learn all the different noises of a system, connect every single noise with a certain information and memorize it. Like in the "ambientROOM", the user has to know what the sound of a bird or rainfall means to him [12]. Thus, other forms to communicate peripheral information between a computer system and the user are required.

3.1.2 Ambient Information

In the field of ambient information there are some other possibilities to communicate data, not only via sounds. Ambient information has the ability to come to the user's periphery of the attention. It is sent out through an ambient media which uses dynamic changes in light, sound, form or color in a physical space [21]. Ambient Media is one of the ubiquitous computing techniques which brings information into the human's surroundings [14]. Thus, the user will not be disturbed while interacting with mouse and keyboard by these types of information.
Light or color changes plays a huge role in that area. In the "ambientROOM" of Ishii et al. there are illuminated patches projected onto an inner wall to illustrate the human's movement or the number of people present [12]. Thus, via ambient media it is possible to get digital information in a peripheral way. A good solution which hardly not disturbs the user in his center of the attention would be an adequate combination of sound, light and color changes, like in the "ambientROOM" example.

However, this is just one direction in which information can be sent. To interact in a digital environment, it is necessary that the user is able not only to react on the received data but also to send information to a computer in a peripheral way.

3.2 Sending Information

This section is very important for the topic of that paper because it describes the active part of a user in which he performs actions in the periphery of the attention. Via these actions, information is sent to a computer system while the user himself is concentrated on the screen and is almost not disturbed neither by the information nor by his actions. But what kind of actions can a user perform to interact with a machine in a peripheral way?
As already mentioned in chapter two, a peripheral action cannot require many resources. Therefore only small, short and intuitive actions are possible in this context.

3.2.1 Information via Tangibles

Bakker says that most of the peripheral actions "are performed with the hands, on small or wearable objects" [1]. That supports Bakker's choice for tangible interaction in the design process for peripheral interaction. Tangible interaction is possible via a tangible user interface (TUI) which "augments the real physical world by coupling digital information to everyday physical objects and environments"[11]. Thus, a tangible user interface would be one option to perform peripheral actions also because it "makes computing truly ubiquitous and invisible"[11].
Edge [4] wrote an important technical report about this technique of ubiquitous computing for peripheral interaction. He described how

TUIs can support user's low-attention interactions with digitally augmented physical tokens. For that purpose, a prototype of a TUI was designed and implemented (which will be presented in subsection 4.4). Edge evaluated that prototype in an office context and worked out some identifications of the qualities of peripheral interaction of the different ways a user can act with a TUI. He found out four essential qualities of peripheral interaction and named them as "concise implications for design" [4]:

- "Tangibility is about giving users the freedom to project new systems of meaning onto the world, based on the existing meaning that it already has for them"
- "Tangibles are not just physical objects with digital augmentations, they are socially-situated objects of external and distributed cognition"
- "Tangibles exist in the physical world and have the potential to remain meaningful even when outside of the sensing region of their parent interface"
- "Tangibility provides an opportunity for selective, fluid, episodic engagement with information in a more direct manner than WIMP-based interaction"

As we can see, Edge talks about socially-situated objects which have the potential to remain meaningful in the periphery. These tangible objects should already have a meaning for the user and the user should be familiar with the usage of them. Thus, if the user naturally knows how to deal with those object, they can be suitable for peripheral interaction. That is supported by the theory of Wickens [20] who says that a mental effort decreases with practice and experience. A task can only be performed in a peripheral way when it does not need a high mental effort, as we learned in chapter two.
So tangible user interfaces are very powerful and offer a great possibility to perform peripheral actions. But we do not necessarily need a digitally augmented physical token to communicate with a computer in the periphery. There are other forms of human's communication that are very common in the physical world and so they can also be suitable for peripheral interaction in a digital environment.

3.2.2 Information via Hand Gestures

Pavlovic et al. [15] wrote an article about hand gestures for human computer interaction (HCI). In the field of HCI, natural ways of human's communication are analyzed to integrate them in a digital environment. Thus, the HCI wants to use these ways of human-to-human communication from the physical world to offer other possibilities for human computer interaction, away from the typical interfaces of mouse an keyboard. As we learned in previous chapters, natural and common actions can be performed in the periphery of the attention, because they do not allocate many resources. So these alternative ways of human-to-computer communication which the HCI describes serve as potential ways of peripheral interaction in a digital context.
One of these ways are human hand gestures which "are a means of non-verbal interaction among people" [15]. There are simple gestures like pointing at an object or complex ones to express our feelings. Quek mentions that "the chief motivation for using gesture is that humans have the facility and intuition for gestural communication" [16]. We will focus on simple and common human hand gestures because these are predestined for our purpose.
To be able to use gestures in a human-to-computer communication we need a possibility to measure the movements of the hand and to interpret the corresponding information. Pavlovic et al. use a set of video cameras to capture the actions, and computer vision techniques to interpret them. They worked out a "gesture recognition system" [15] which is based on 3 main components:

- Gesture Modeling to model natural human hand gestures
- Gesture Analysis to estimate the parameters of the gesture model using video images
- Gesture Recognition where the data analyzed from the images of gestures is recognized as a specific gesture.

The system is shown in *figure 1* which presents an idea of how human hand gestures can be used to communicate with a computer. This type of communication can be easily used to interact with a computer in a peripheral way by two main reasons:
First, common and natural gestures do not have to be performed in the user's center of the attention because they do not require many resources (compare chapter two). But when a user interacts with keyboard and mouse, we can imagine that he uses his two hands. So he would have to stop with one task to perform an action with a hand gesture. But this would mean that the user is getting disturbed in his central work and so that gesture could never be peripheral.
However, Edge [5] found out during a study in an office context, that only half the time, the user used both hands to interact with mouse and keyboard. When just one hand operated with mouse and keyboard then the dominant hand was employed for twice the duration of the left hand. So the second reason why hand gestures are suitable for peripheral interaction is, that one of the user's hand is unemployed anyway for half of the time. Thus, he would not be disturbed by performing a gesture while using the other hand to interact with mouse or keyboard.

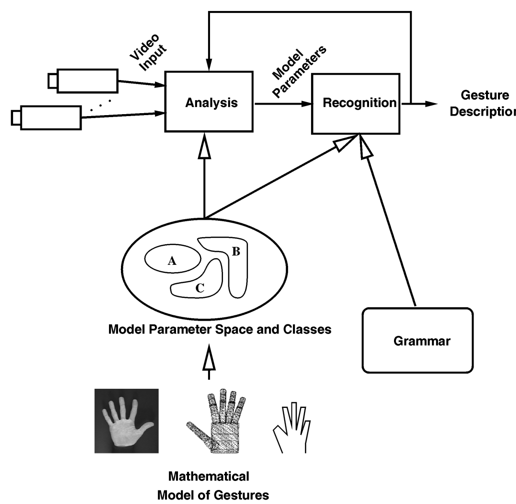


Fig. 1. Gesture recognition system by Pavlovic et al. [15]

3.2.3 Information via Movements

Gestures are just a special form of hand or arm movements because they are expressed by them. According to Pavlovic et al. [15], hand or arm movements can be classified as gestures or unintentional movements. Last ones are those movements that do not transport any useful information, so they are irrelevant for our purpose. Thus, a short and common arm movement can be suitable for peripheral interaction because of the same reasons as for hand gesture which have been discussed in subsection 3.2.2. Many other forms of human motions are imaginable in this case, too. We can think about to nod the head, to shrug or to use the feet (see also [10]) to interact with the computer in a peripheral way.

3.2.4 Information via Audio

Audio signals are not only suitable to transport information from a computer to the user (see subsection 3.1.1). They can also be used in the other way. Users are able to interact with a machine via audio signals which can be captured by a microphone and interpreted by a computer system. By the way, this form of human computer interaction is also an actual research area.
Igarashi and Hughes [9] presented some audio interaction techniques that use non-verbal features in speech to directly control interactive applications. These techniques are called "voice-as-sound techniques".

They mentioned three types of control. Each of them has its advantages in certain use cases and so they are used for different purposes:

- Control by continuous voice, where the user's voice is like an on/off button
- Rate-based parameter control by pitch, where the user adds an additional parameter by pitch
- Discrete control by tonguing for discrete value selection

It becomes clear that there are various possibilities to send audio information for different purposes. So in our case we need audio signals which do not allocate many human resources, so they can be sent in a peripheral way. We learned that small and simple actions meet this condition (see section 2) and thus, simple audio signals are predestined for peripheral interaction in this context. Igarashi and Hughes mention the simplicity of their "voice-as-sound techniques" as one advantage over traditional speech recognition. They say that these "techniques rely on very simple signal processing" and for that reason we can use them as another possibility for peripheral interaction.

Goto et al. used the possibilities of audio interaction and designed a speech interface "that makes full use of the role nonverbal speech information plays in human-human communication" [6]. So they also transferred some standards of communication from the physical to the digital world (like in the case of human hand gestures in subsection 3.2.2).
In the physical world, people can sing, whistle or speak while concentrating on something else. In the car example of subsection 2.2 a person is able to sing a song while driving the car and concentrating on the road. So the audio information is sent in a peripheral way.
Transferring this into a digital environment, we can see that the communication of information happens in the same peripheral way. A user can interact with mouse or keyboard and be concentrated on the screen while singing, speaking, whistling or making some other noise. So the user is able to send information in a peripheral way by creating certain sound signals.

In summary, sending information in a peripheral way to a computer system is possible and can be realized in various ways and via different types of user actions. We learned that peripheral interaction is easy, intuitive and quite common for the user so he would not be disturbed in his center of the attention and can handle more tasks at once. Tangibles, hand gestures, arm or hand movements and audio signals are suitable to perform peripheral actions and so they are already used in various papers about designing prototypes for peripheral interaction. Some of those academical works will be presented in the next chapter.

4 Prototypes in Literature

This section is about peripheral interaction in the actual research area of ubiquitous computing. The following subsections will describe some prototypes which were designed and developed to evaluate techniques for user interactions without mouse and keyboard. The corresponding papers present the current state of research in this area. Each of them shows different aspects and possibilities of human computer interaction and totally they give a good overview about how peripheral interaction can be realized in the digital world.

4.1 Cyber PK - Peripheral Interaction of Ambient Media

Cyber PK is a prototype, designed and developed by Park and Nam [14] from the department of Industrial Design at the Korea Advanced Institute of Science and Technology. It "represents virtual activities in online cyber spaces through changes in movement, light, and graphical images, to make its peripheral interaction possible".
Thus, the focus is more on the passive part of a user (see subsection 3.1) and not on the active one, as in this paper. The possibilities of user interactions in the periphery are not as considered as the forms of getting information in a peripheral way. However, Cyber PK represents an important part of peripheral interaction (the passive one) and describes the possibilities of ambient media so it has to be named in this context.
Park and Nam were inspired by a windbell which normally hangs down from an orient temple and describes the strength of the wind. Thus, here we see the connection between the physical and the digital world as we did in previous chapters (like in subsection 3.2.2 and 3.2.4). For a better imagination, the windbell is shown in *figure 2*.

Fig. 2. A Korean windbell serves as inspiration for Cyber PK [14]

Park and Nam presented dynamic design elements which can be applied to ambient media design in a physical space. Tempo, continuity, intensity and rhythm were identified as such elements and they were classified for peripheral interaction. Park and Nam worked out some characteristics of such dynamic design elements. The following points show a short summary of their findings:

- Stressing dynamic design elements is useful in moving from the user's periphery to its center of the attention.
- Dynamic design elements are conspicuous identifiable.
- Dynamic elements in physical spaces have an intuitive relationship to certain information in virtual spaces
- Tempo, continuity and intensity can be optimized for peripheral interaction because they are sequentially changeable. Rhythm can express different types or levels of information

This theoretical background serves for the design of Cyber PK's peripheral interaction which was worked out using dynamic design elements. The context of that prototype which presents cyber activities, is the Web blog of communities or individuals. After a survey of 18 Web blog users, cyber activities were ordered according to their importance. The result list shows the following order (starting with the most significant cyber activity): logging in/out, the accumulated activities degree, uploading, and scrapping. Each activity has its own information. The importance of this information and the relation between information and expression were considered and dynamic design elements were applied for peripheral interaction.
Information about important cyber activities like logging in/out is mapped to the swings of a poise of Cyber PK because they "can easily attract an inattentive user's attention through sounds and visual movements". However, information about less important cyber activities like uploading or scrapping, is represented by the blinking of the light of Cyber PK to a certain color. Thus, this information is transported in a peripheral way because color changes does not affect so much the user's center of the attention as the movement of the poise does.

The accumulated activities' degree which is a semi-important information, is visualized by graphical animation on the screen of Cyber PK. It moves at a fast or a slow rate according to the amount of cyber activities, just like a heartbeat.
The various information is represented in different ways by Cyber PK: Information about the amount of visitors (users who are logged in) is represented through the intensity of Cyber PK's movements: Many visitors cause a high intensity of movements to engage the user's center of the attention and less visitors decrease it to address the user's periphery of the attention. If a visitor belongs to a certain group, "the tempo of the movement can go up to peripherally attract the user's attention".
When there are not many important cyber activities (like uploading or scrapping) the tempo of changing the light colors and their intensity are low. If the frequency increases the tempo and the intensity go up to address the user's attention.
The tempo of the animation on the screen which shows the accumulated activities for one day is slow when inactive and fast when active. So it changes from the user's periphery (slow) to his center of the attention (fast).
Four different types of Cyber PK for various application scenarios were developed: A hanging-type, a stand-type, a streetlight-type and a mobile type. *Figure 3* shows the stand-type of Cyber PK which is adequate for our purpose. As already mentioned, this prototype focuses on ambient information. This is a very good example how ambient media can be used to transport information to the user's periphery. However, the active part in which a user sends information by performing actions in the periphery is not really considered. Thus, the prototype which will be presented next considers both parts of peripheral interaction and focus on the active part of the user.

Fig. 3. The stand-type of Cyber PK, developed by Park and Nam [14]

4.2 Ambient Appointment Projection

Hausen and Butz from the university of Munich wrote two papers about peripheral interaction and developed two prototypes which will be presented in these two subsections [7, 8].

They described five design dimensions for, how they call it, "peripheral embodied interaction" which are summarized in the following points:

- *Explicitness* which is a dimension ranging from explicit to implicit interaction
- *Input mode* for peripheral embodied interaction like hand gestures or tangible objects
- *Granularity* describes the number of commands that can be encoded for one form of interaction
- *Privacy* which is hardly to realize in peripheral embodied interactions
- *Proximity* which describes the variety of distances between the different form of peripheral embodied interactions

Both of their prototypes have been developed based on these points of peripheral embodied interaction.

The first is called "Ambient appointment projection" which is shown in *figure 4*. It projects a spiral visualization on the users' desk which indicates the overall time flow of upcoming appointments. The spiral starts pulsating when an event is coming close. Thus, via this ambient information the user will be reminded about the appointment in a peripheral way because of dynamic light changes (see subsection 3.1.2).
The user can now react on that information by a wiping hand gesture which is captured by a camera. If the user wipes towards himself he will get more details about the next appointment, shown as a balloon tooltip. If he wipes away from himself then the pulsating of the spiral will stop immidiately. Thus, this action is performed in the user's periphery because natural and common gestures are used (see subsection 3.2.2) which "meet the metaphor of fetching wanted or pushing away unwanted things".
The prototype was evaluated in a user study by twelve participants who had to type a given text as fast as possible while not missing any appointments. The result of the study was that the ambient appointment projection offers a smoother handling of upcoming appointments than other state of the art reminder (e.g. the one of Outlook) do. So this indicates one of the advantages of peripheral interaction which offer a better and more continuous workflow (see also section 5).

In summary, this prototype represents an adequate solution for our purpose because it considers both directions of human computer interaction. The user gets and sends information and this is realized in a peripheral way.

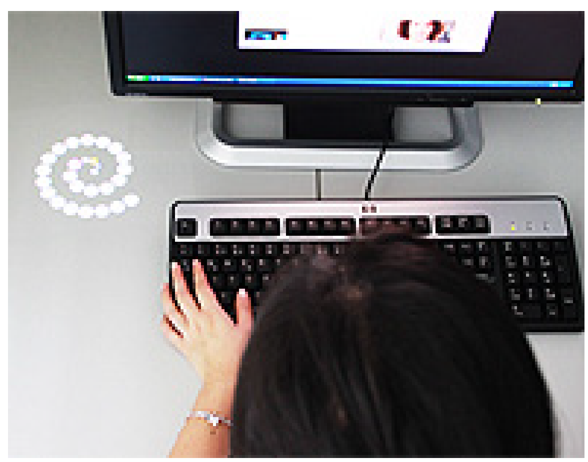


Fig. 4. The prototype of ambient appointment projection, developed by Hausen and Butz [7, 8]

4.3 Tangible Presence Indication

Hausen and Butz developed a second prototype using the notion of peripheral embodied interaction [7, 8]. It is called the "tangible presence indication" which is shown in *figure 5*. This prototype is designed as a cylindrical object consisting of several levels. It indicates information about the user's presence through the biggest and highest level and shows his selected contacts via the other levels. It is connected to Skype and encodes user statuses in a color-coded way. Also other customized statuses than the standard ones ("available","away" and

“do not disturb”) like “in a meeting” can be set.
If anyone who uses the tangible presence indication changes his status, the corresponding light of each cylinder switches to another color according to the new status of that person. In this case, ambient information is sent to the other users by dynamic changes of color and that means (consider subsection 3.1.2) that this information is transported in a peripheral way.
If a user wants to change his status he can do that by turning the highest level of the tangible presence indication. This level integrates a button and when a user pushes it he can set the time he wants to be in a selected status. Anyway, pushing or turning a button is “very natural in the physical world” (e.g. managing the radio in the car example of subsection 2.2). Natural and short arm or hand movements are a suitable way to perform actions in the periphery (see subsection 3.2.3). Thus, the user acts in a peripheral way to change his status by using this prototype.

Summing up, the tangible presence indication also describes a prototype which supports a complete peripheral interaction in both ways. So it is another adequate example (next to the ambient appointment projection, described in subsection 4.2) for our purpose and it represents a good solution of how peripheral interaction becomes possible in a digital environment.

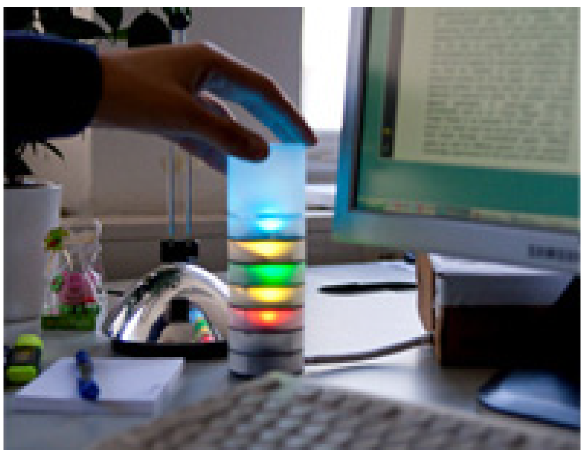


Fig. 5. The prototype of the tangible presence indication, developed by Hausen and Butz [7, 8]

4.4 Group Task Management

Edge, an important researcher in the field of tangible user interfaces (TUI), developed a prototype for peripheral task management, based on the analytic design process for TUIs. There are several papers about this prototype which shows the possibility of peripheral interaction through the usage of TUIs [3, 4, 5]. In cooperation with Alan F. Blackwell [5] he describes the analytic design process for TUIs which consists the following four stages:

- Context analysis which “identifies the activities in a context that could benefit from TUI support”
- Activity analysis which “describes the properties of a TUI that would appropriately support these activities”
- Mapping analysis which “generates the physical-digital mappings of a TUI structure with these properties”
- Meaning analysis which “provides these mappings with meaning that users can understand and adapt”

Edge and Blackwell used Weiser’s vision of calm technology [19] to design a TUI “based on tangible objects that could drift between the focus and periphery of a user’s attention according to the momentary demands of their activity”. Consisting the analytic design process, a tangible interaction system was developed as a prototype for peripheral task management.

The designed system is located in an office context where open tasks should be coordinated within a team. It contains a knob and an interactive surface on which a user can place tokens. A camera above the surface captures their position and identity. The tokens represent items of common interest (tasks, documents and contacts) within a work group. They are visually augmented by display halos, displaying their attributes (e.g. name, estimated time remaining) when they are on the surface. Each kind of token is augmented by a certain halo color: Tasks are yellow-and-red, documents yellow-and-blue and contacts are green-and-blue. Thus, the user connects a certain information with the color of the tokens. For better understanding, the interactive surface is illustrated in *figure 6*. A timeline, located on the top of the surface, shows the actual days of the week or month. The illuminated areas between the tokens and the timeline indicate how long a task is active and can be performed within the team.

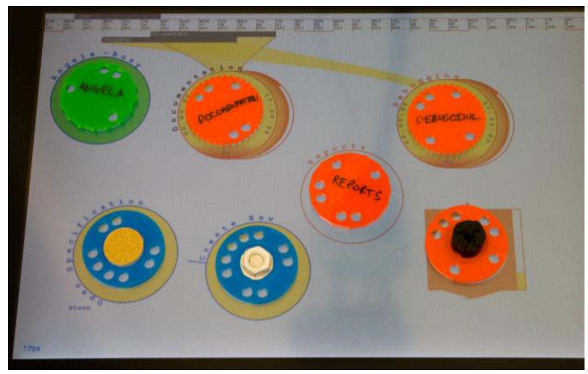


Fig. 6. The interactive surface of Edge’s prototype for a peripheral task management [5]

The knob of the system serves to adjust a certain attribute of a token by turning it. Tasks can be coordinated within the work team by changing the different attributes of the tokens. If a worker needs more time for a certain task, he nudges the corresponding token to select attributes to change and then he turns the knob. Pushing or turning a button is very natural and a common movement in the physical world, as we have seen in subsection 4.3. Nudging a token is a small movement, so this indicates one reason why the user can perform peripheral actions in this context. The other workers will be informed about the changes of the attributes by changes of the display halos on their surfaces and by a pop-up dialog box on their screens. Thus, a user is getting informed in a peripheral way because of dynamic changes of lights (see subsection 3.1.2) of the user’s surface. *Figure 7* shows the complete prototype for the peripheral task management. For further details about this prototype a closer look into the work of Edge is recommended [4]. We will now focus on its style of peripheral interaction.
We learned that the dominant hand is used twice for the duration of the other hand (see subsection 3.2.2). For that reason, the surface is located to the left (if the person is right-handed) of the user’s keyboard because this is not the dominant side and so the user can interact in a peripheral way while his left hand is unemployed.
However, the knob of the system is on the right side. The idea is that the TUI is used bimanual because the usage of both hands is very common in this context (mouse and keyboard are also used with both hands). Intentional and natural actions do not allocate many resources (see section 2). Therefore, the user can switch fast and easily from the workstation (used for main tasks) to the TUI which is used for auxiliary tasks. Thus, fluid switching is possible between the center and the periphery of the attention. Turning the button and changing attributes of the tokens are both actions which can be performed

peripherally, as we have seen above. So there are two main reasons why peripheral interaction is supported in this context:
First, the workflow is not interrupted because switching is possible in a rapid and intentional manner. And second, the user is able to interact with the TUI while he continues concentrating on the main task because his actions do not allocate many resources. Therefore, we can say that the TUI meets all requirements to offer peripheral interaction.

This prototype is a little complexer than the others, presented in this chapter. It should become clear that a TUI like this offers great possibilities for our purpose (compare subsection 3.2.1).
All in all, the presented papers and prototypes give a good overview about the techniques of peripheral interaction in the digital world. The differences and the advantages of these techniques will be discussed in the following section.

Fig. 7. The prototype of a peripheral task management, developed by Edge [3]

5 Discussion

Concluding the facts of the techniques of peripheral interaction, the results shall now be critically contemplated. We will firstly name some general aspects before we can discuss the differences of the prototypes which we presented in section 4.

5.1 General Aspects of Peripheral Interaction

According to the physical world, we can see that peripheral interaction offers the possibility to handle more information in the same time (e.g. driving the car, look at the road and listen to the music of the radio). In a digital environment, this means that users can perform more important tasks and handle more information because they will not be disturbed by the huge amount of little infos (like incoming mails, upcoming appointments, system messages) which normally appear in a computer session. This is a great advantage especially for economical and social reasons. People perform those natural actions, which they perform all day long. So they feel comfortable, familiar and free while interacting with a computer system. Satisfied worker normally do a better job and so we can say that natural and common ways of human's attention are used to improve the workflow. That can help to save a lot of time and money.
One problem is, that peripheral interaction cannot be realized through mouse and keyboards. New types of user interfaces become necessary which support interactions in the user's periphery. This means high investments in investigation and development of such interfaces and huge costs of implementation and infrastructures will follow. Therefore a simple solution for peripheral interaction is required which do not need a huge amount of new hardware and software.

5.2 Comparison Between the Presented Prototypes

Peripheral interaction is an actual research area as we have seen in chapter 4. Many possibilities are discussed of how an information can be received (e.g. audio, color or light changes) and sent (e.g. tangible interaction, speech, hand gestures) by the user in his periphery of the attention. Each way of interaction has its own advantages and disadvantages in comparison to the others. Thus, different systems and situations require different ways of peripheral interaction in different user scenarios.
The presented prototypes of the previous section give a good overview about various user cases and different forms of peripheral interaction. Thus, in this part of the paper we want to discuss these differences and pick out some advantages of each prototype.

As already mentioned in subsection 4.1, Cyber PK [14] focus on ambient media and its form of peripheral information but only in one direction. The user has no possibility to interact actively, he just receives information in his periphery. The advantage of this concept is that the user gets many different information by different peripheral signals. These will not disturb the user in his center of the attention although he receives many information. Thus, the user will be informed very well about the activities in the Web blog without losing the attention on his main task.
On the opposite, the prototype of peripheral task management by Edge [4] supports more the active part of a user than the passive one. A complex system was developed to discover the possibilities of TUIs for peripheral interaction. The usage of the hands play a huge role and the interaction in the user's periphery is supported by a rapid and intuitive switch between the main and the auxiliary task. Small movements of hands or arms and a short push on a little button describe all the actions a user has to perform. However, the information a user gets from the surface in his periphery is just that something changed. Thus, we can say that this prototype focus on user's active peripheral interaction. The advantage is that through the complexity of this prototype many aspects of user interactions can be discovered (e.g. the usage of the hands during a computer session). Considering the papers of Edge [3, 4, 5] we get a good idea about the potential usage of tangible user interfaces for peripheral interaction in the digital world.
The two prototypes of Hausen and Butz [7, 8] can be put between the Cyber PK (supporing the passive part) and the Group task management (supporting the active part). They support both directions of peripheral interaction in the same manner. The ambient appointment projection presents peripheral information about upcoming appointments by dynamic changes of light of the spiral (passive direction) and the user can react on it in a peripheral way, by a wiping hand gesture (active direction). The advantage of this prototype in comparison to the others is the simplicity of the interactions. Users are not disturbed in their central attention by a simple hand gestures and they can continue their main tasks while receiving information about the next appointment. The simplicity of that prototype also shows that peripheral interaction is easy to realize and do not need a huge amount of new hardware components or system changes.
The tangible presence indication supports the same advantages. Peripheral information is given by various colors of different levels. The variation is, that this prototype uses object manipulation for peripheral interaction and so it supports a similar possibility of user actions like the group task management. However, this prototype is much simpler. Just one hand is used and a short movement (e.g. push a button, turn a level) is sufficient to manage the user's presence. Thus, the user does not have to switch completely from one interface to the other (like with the Group task management) and so he can continue with his main activity while using the prototype. This could be seen as a better form of real peripheral interaction.

6 Conclusions and Future Work

This paper described in details the possibilities and the techniques of peripheral interaction in a digital environment. According to Sternberg's divided attention theory [18], we learned how humans act in the center and the periphery of the attention. These psychological findings were used to determine different ways of how a user can send and get information in a peripheral way. We discovered that ambient information via dynamic changes of light, sound, form or color is suitable for such information a user can receive in his periphery.

We named audio signals as a special and adequate form for getting peripheral information.
The user can react on it via tangibles, hand gestures, short movements or audio signals. Each way of peripheral interaction has its own properties and advantages. Thus, various use cases need different types of interactions. Therefore, several prototypes from the literature which cover different styles of peripheral user interactions were presented and discussed. We explained Cyber PK [14] as a good example for ambient information. The prototypes of the ambient appointment projection and the tangible presence indication [7, 8] showed adequate ways to make peripheral interaction possible in a simple and natural way. The group task management [4] described a complexer form. It used a tangible user interface to switch very fast between the center and the periphery of the attention.

In summary, we answered the research questions of how peripheral interaction can be realized in a digital environment and we discussed the advantages of these possibilities in general. However, social and economical aspects of peripheral interaction were not considered in details. We have to consider the adaptation of interfaces which support this kind of user interaction in our daily life. Do people really want to perform hand gestures or make noises to communicate with their computer? Is it necessary that we have more than the common interfaces like mouse and keyboard on our desk? And what about the costs of these new user interfaces? These and other questions have to be answered before integrating peripheral interaction in our digital environment. Aspects of security and privacy must also play a huge role in this context. Considering Hausen and Butz, we can see that "peripheral embodied interaction can be observed much more easily" by a third person than keyboard or mouse interactions [7, 8]. Thus, there are problems of privacy and (in case of sensitive data) security which have to be solved in the future. Complex user studies and evaluation of other prototypes are required to get a comprehensive knowledge about the problems, the potentials and the possibilities of peripheral interaction in the digital world.

REFERENCES

[1] S. Bakker. Talking tangibles: design for peripheral interaction. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, pages 427–428. ACM, 2011.

[2] S. Bakker, E. van den Hoven, and B. Eggen. Design for the periphery. *EuroHaptics 2010*, pages 71–80, 2010.

[3] D. Edge. A tangible user interface for peripheral task management. In *Visual Languages and Human-Centric Computing, 2006. VL/HCC 2006. IEEE Symposium on*, pages 228–229. IEEE, 2006.

[4] D. Edge. Tangible user interfaces for peripheral interaction. Technical report, university of cambridge, 15 JJ Thomson Avenue, Cambridge, UK, december 2008.

[5] D. Edge and A. Blackwell. Peripheral tangible interaction by analytic design. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pages 69–76. ACM, 2009.

[6] M. Goto, K. Itou, T. Akiba, and S. Hayamizu. Speech completion: New speech interface with on-demand completion assistance. In *Proceedings of HCI International 2001*. Citeseer, 2001.

[7] D. Hausen and A. Butz. Entangling ambient information with peripheral interaction. In *Workshop Work-in-Progress. 5th ACM Conference on Tangible, Embedded and Embodied Interaction (TEI 2011)*. ACM, 2011.

[8] D. Hausen and A. Butz. Extending interaction to the periphery. In *Workshop Embodied Interaction: Theory and Practice in HCI. In conjunction with 29th ACM International Conference on Human Factors in Computing Systems (CHI 2011)*, pages 1–4. ACM, 2011.

[9] T. Igarashi and J. Hughes. Voice as sound: using non-verbal voice input for interactive control. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 155–156. ACM, 2001.

[10] H. P. Institute. Multitoe, `http://www.hpi.uni-potsdam.de/baudisch/projects/multitoe.html`. last visited: 07-05-2011.

[11] H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 234–241. ACM, 1997.

[12] H. Ishii, C. Wisneski, S. Brave, A. Dahley, M. Gorbet, B. Ullmer, and P. Yarin. ambientroom: integrating ambient media with architectural space. In *CHI 98 conference summary on Human factors in computing systems*, pages 173–174. ACM, 1998.

[13] K. Lyytinen and Y. Yoo. Ubiquitous computing. *COMMUNICATIONS OF THE ACM*, pages 63–65, 2002.

[14] J.-Y. Park and T.-J. Nam. Dynamic design elements for the peripheral interaction of ambient media. In *CHI '08 extended abstracts on Human factors in computing systems*, pages 3717–3722. ACM, 2008.

[15] V. Pavlovic, R. Sharma, and T. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, pages 677–695, 1997.

[16] F. Quek. Toward a vision-based hand gesture interface. In *Virtual reality software & technology: proceedings of the VRST'94 Conference, 23-26 August 1994, Singapore*, pages 17–31. World Scientific Pub Co Inc, 1994.

[17] O. Shaer and E. Hornecker. Tangible user interfaces. 2010.

[18] R. Sternberg, J. Mio, and J. Mio. *Cognitive psychology*. Wadsworth Pub Co, 2009.

[19] M. Weiser and J. Brown. Designing calm technology. *PowerGrid Journal*, pages 75–85, 1996.

[20] C. Wickens and J. McCarley. *Applied attention theory*. CRC, 2008.

[21] C. Wisneski, H. Ishii, A. Dahley, M. Gorbet, S. Brave, B. Ullmer, and P. Yarin. Ambient displays: Turning architectural space into an interface between people and digital information. *Cooperative buildings: Integrating information, organization, and architecture*, pages 22–32, 1998.

Evaluation of Ambient Information Systems

Simon Mang

Abstract— This work concentrates on the evaluation of ambient information systems. These systems provide information while blending into the user's environment. Therefore the devices are designed as – or even embedded in – every-day objects. This makes evaluation a difficult, expensive and time-consuming task. On the one hand evaluating in an unnatural laboratory setting includes the risk that the device is pulled back to the user's main focus. On the other hand study settings in the user's real environment improve the ambient situation but are costly and time intensive. Therefore researchers are looking for new and alternative methods how to evaluate ambient information systems best. In this paper we describe three different study designs and compare their weaknesses and strengths. For example a lab study is compared with a field evaluation. Furthermore we address a study working with a modified set of Nielsen heuristics and compare the results with a Wizard-of-Oz study. This paper can be seen as a summary of study designs and methods for evaluation ambient information systems.

Index Terms— Ambient information, peripheral displays, evaluation methods, ubiquitous computing

1 Introduction

Ambient information systems (AIS), often simply called ambient displays, become more and more interesting. These displays are hidden in lamps, in cables hanging from the ceiling or in art. All of them kind of "hide" the information that is displayed. In that way a picture which looks at a first glance like a modern art picture can be a device displaying bus departure times. Despite the raise of interest in these systems it is amazing that "the majority of ambient displays that have been published have not been evaluated" [11]. The research community agrees upon the fact that evaluation is existential to the design process of new implementations and applications. Evaluating a newly designed technology leads to the discovery of issues and problems and therefore guarantees proceeding improvement.

The nature of AIS is to blend to the periphery of the user. They are designed to be not obtrusive, nor distracting and neither disturbing. That is why ambient systems are often embedded into everyday objects and thus automatically move away from a person's focus. This great feature of these technologies offers the possibility to easily ignore them if more important tasks require the user's full concentration. This of all things puzzles researchers evaluating AIS. Conducting user studies with a device whose overall goal is not to attract attention is a difficult task. Of course there is the possibility to run long-term *in-situ* studies that allow users to take the display to an environment they are used to (desk at work, living room, etc. . .) but these studies are money and time consuming. Whereas lab studies may falsify the results because every probing will pull the ambient device from the user's periphery back to her main focus. This destroys the ambient character of the device. The difficulty lies in developing a method to collect data without heavily influencing the relationship between the user and her peripheral awareness [6].

After giving a brief overview of existing work about evaluating AIS, this paper concentrates on three main aspects. We have a detailed look at heuristic, laboratory and field evaluation of ambient systems. Each method is examined based on a case study. Outlining the results of these studies shows up advantages and disadvantages of each method. In a comparing section we oppose the different methods and evaluate their strengths and weaknesses. Finally we conclude with a summary and brief outlook on future work.

• *Simon Mang is studying Media Informatics at the University of Munich, Germany, E-mail: mangs@cip.ifi.lmu.de*

• *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

2 Background

In 1995 Weiser and Brown [19] presented one of the first ambient displays. They referred to it as "calm technology" back then. Their work does not include any evaluation but illustrates how users benefit when information is moved to the periphery. One challenge they encounter is the problem of how to design ambient technology. To capture the space of ambient information systems Pousman and Stasko [16] developed four design dimensions. They examined 19 existing systems and ranked each on one of the following axes: information capacity, notification level, representational fidelity and aesthetic emphasis. Depending on which area system designers like to focus they may emphasize one or the other dimension. Whereas *information capacity* refers to the quantity of information elements a system is able to display, *notification level* stands for the level of "ambiance" of a given system, i.e. to which extend the system is obtrusive or not. *Representational fidelity* is considered to be the measure how direct information is displayed, i.e. the abstraction of a system and finally *aesthetic emphasis* represents the importance given to aesthetic design of the system. By clustering the 19 systems over each dimension Pousman and Stasko [16] uncovered four archetypes (Symbolic Sculptural Displays, Multiple Information Consolidators, Information Monitor Displays, High Throughput Textual Displays) of ambient information systems. These archetypes serve as a basis for system developers who are modeling new systems.

In general there are two ways how to evaluate ambient displays and peripheral monitors. Studies may be conducted in a laboratory setting or "out in the wild", so called *in-situ* studies. [6] and [7] mention that lab studies especially serve for determining the usability and functionality of the implementations. Even if a study is designed to cover a long period of time subjects sense the artificial setting and any probing forces the system to move from the periphery back to the focus of the subject. Due to this reason there is many work on field evaluation [4, 6, 10, 12, 18].

An approach to meet the growing social concern of supporting aging adults while leaving them in their own houses rather than moving them to some sort of nursing home is presented by Mynatt et al. [12]. Their digital family portrait consists of an icon populated frame and a family picture. While the frame displays the condition at the remote location (like health, activity or events), the picture remains untouched. As it was an successive study first there was an interview of adult children and their parents about what needs to be sensed. Second the device – designed due to the findings of the first step – was evaluated in a nine day field trial with a grandmother and her grandchildren. A slightly different version of the family portrait is the *CareNet Display* presented in [4]. The device consists of a display showing a portrait picture and a wooden frame (see Figure 4). A more detailed description can be found in Section 6.2. In difference to the family portrait – where the elders received a display, too – in this deployment only the caregivers received displays. These are updated with information gathered by daily phone interviews with the elders. The main result is that ambient displays can support care coordination and information sharing. This in home evaluation uncovered at least one design problem: participants are disturbed by the glowing display in the dark.

[10] presents a completely different ambient information system. *Breakaway* is designed as a sculpture sitting on the office desk, reminding workers how long they are sitting at their desk and changing their behavior in a way that they take a break and get up. Information is collected automatically by sensors integrated in the seat. After sitting for a period of 60 to 90 minutes the sculpture slouches, while getting up and being absence causes the sculpture to move back to the upright position. A two week *in-situ* study with one participant showed a relationship between the slouching pose and the break times. Besides the advantage of ambient systems being ignorable in busy times was attested. These results were easily gained by analyzing the collected data and a debriefing interview.

Another deployment with a no-monitor display was presented by Hazlewood et al. [6]. They provided an information channel for students allowing them to give feedback about their courses. The university instructors received an ambient orb (i.e. a small frosted glass sphere) that displayed the feedback information by shifting between a red, yellow or green state. During a two month period the students were sent an email after each class that contained a form to rate their confidence regarding their understanding of the class. Due to this rating the orb color of their instructors changed. The researchers intended that a change in display color results in higher access rates to a web page that provided detailed information. An analysis of the server logs led to the insight that little use of the informal website and inconsistent participation on the part of the students makes strong inferences impossible.

To address the problems and possibilities of the different forms of evaluation we examine in the following the work of [4], [9] and [11]. We focus on three different research approaches:

1. evaluation without the user
2. evaluation in the laboratory
3. evaluation in the field

The work "Heuristic evaluation of ambient displays" by Mankoff et al. and Consolvo's CareNet approach serve as a base for 1 while Hsieh's comparison of peripheral displays is used intensified but not exclusively for points 2 and 3.

3 Heuristic evaluation of AIS

Heuristic evaluation is a "discount usability engineering method for quick, cheap, and easy evaluation" [14] to find usability problems in an user interface. Instead of many inexperienced test users a small set of evaluators (which may be display designers or usability experts) examines the UI. Nielsen found that 3-5 evaluators find already 60-75% of all usability problems, when 15 evaluators find 90% [13]. The evaluators inspect the interface by comparing it and it's components to a list of heuristics (taken from [15]): 1) Visibility of system status 2) Match between system and the real world 3) User control and freedom 4) Consistency and standards 5) Error prevention 6) Recognition rather than recall 7) Flexibility and efficiency of use 8) Aesthetic and minimalist design 9) Help users recognize, diagnose and recover from errors 10) Help and documentation.

These heuristics are meant to evaluate traditional displays that require the user to *use* them. As ambient displays are *perceived* by the user some of Nielsen's heuristics do not fit. Mankoff et al. [11] conclude that if the heuristics where modified they could be applied to AIS, too. So they asked separate research members of their group for two independent reviews. Based on those they removed heuristics (3,4,5,7,9 and 10) being not applicable to AIS. Additionally they altered some heuristic titles and definitions and added again 5 ambient display specific heuristics. Finally they made a survey with 7 experts to check the "ambient" heuristics. The result shown in Table 1 is taken from their paper ([11]) and presented here for sake of completeness.

1	**Useful and relevant information** The information should be useful and relevant to the users in the intended setting.
2	**"Peripherality" of display** The display should be unobtrusive and remain so unless it requires the user's attention. Users should be able to easily monitor the display.
3	**Match between design of ambient display and environments** One should notice an ambient display because of a change in the data it is presenting and not because its design clashes with its environment.
4	**Sufficient information design** The display should be designed to convey "just enough" information. Too much information cramps the display, and too little makes the display less useful.
5	**Consistent and intuitive mapping** Ambient displays should add minimal cognitive load. Cognitive load may be higher when users must remember what states or changes in the display mean. The display should be intuitive.
6	**Easy transition to more in-depth information** If the display offers multi-leveled information, the display should make it easy and quick for users to find out more detailed information.
7	**Visibility of state** An ambient display should make the states of the system noticeable. The transition from one state to another should be easily perceptible.
8	**Aesthetic and Pleasing Design** The display should be pleasing when it is placed in the intended setting.

Table 1: Nielsen heuristics modified in a way that they are applicable to ambient displays [11].

Original Nielsen heuristics versus "ambient" heuristics

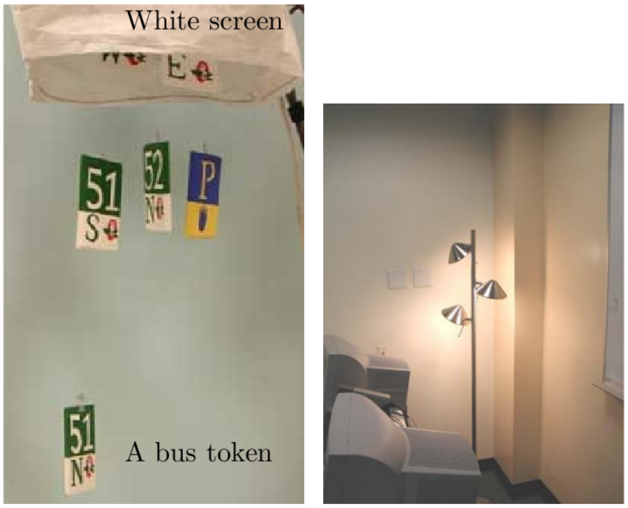
To be able to compare the both heuristic sets Mankoff et al. developed two different ambient displays (see Figure 1). These displays were designed in an user-centered approach. In a first step undergraduate students were interviewed what information would be interesting to them. The displays were build afterwards in a second step.

BusMobile The first display is similar to a mobile. It consists of tokens representing a bus line that moves up and down depending on how far the bus is away from the station in front of the building. The actual time of the day and the bus schedules are used to determine the distance. If a bus is less than 25 minutes away the token hides at the top of the mobile. When it approaches the token moves to the lowest possible depth and from there on continues to raise. E.g. Figure 1a shows that Bus 51S is nearer to the bus stop than Bus 51N.

Daylight Display The second display provides information about the light condition outside (e.g. if it is dusky, light or dark outside). Therefore it serves as an approximated clock and helps concerned students to walk home before night. It is a lamp (Figure 1b) that slowly fades from light/bright to dark/dim at the evening. At dawn it does the opposite.

The research team conducted a study with 60 students over a two week period to receive feedback on the design of the two displays (especially on the usefulness and the problems in the design). They found that the displays are useful to users that are interested in the displayed information. The paper is *not* focused on usability testing of the displays but on evaluating the "ambient" heuristics. Nevertheless it was important to test the displays to be able to guarantee at least a minimum of appropriateness. Mankoff et al. compared the effectiveness of Nielsen's heuristics to their heuristics in a further study using a formal heuristic methodology. They shaped three hypotheses:

1. The number of issues found in the ambient condition will be greater and the issues will be more severe than those found in



(a) BusMobile (b) Daylight Display

Fig. 1: Ambient Displays designed for "ambient" heuristic evaluation [11].

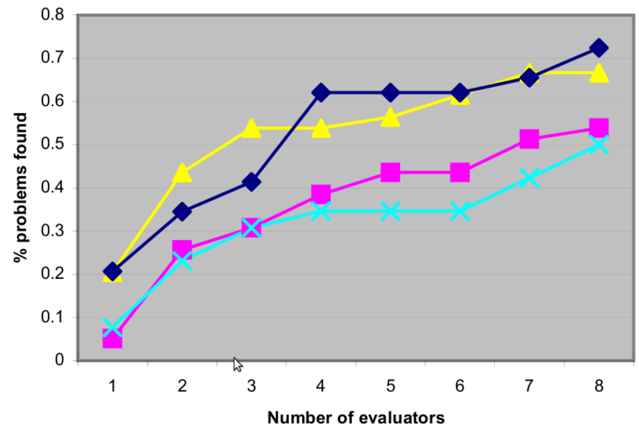


Fig. 2: Results of applying the different heuristics to both displays. The two upper graphs show the good performance of the "ambient" heuristics. The diamond graph shows the percentage of issues found by evaluators by applying "ambient" heuristics to the bus display (triangles for the Daylight display). Evaluating with Nielsen's heuristics results in the squares- (busMobile) and x-graph (daylight display) [11].

the Nielsen condition.

2. The percentage of known issues found in the ambient condition will be higher than the percentage of known issues found in the Nielsen condition.

3. The ambient heuristics will be more useful to evaluators than Nielsen's heuristics. A heuristic that finds many, severe problems is more useful than a heuristic that finds fewer problems with lower severity.

The 16 participants (all of them with half a decade of evaluation experience) were split into two groups. A number of 5 persons typically is seen as sufficient but as there is criticism towards this claim Mankoff et al. included 8 participants per group. One group evaluated the displays using Nielsen's heuristics whereas the other group used the modified heuristics. Each participant evaluated the BusMobile and the Daylight Display individually using a web form. Both displays were described on a separate web page. After reading the description the evaluator created a list of issues. I.e. if a problem was found the related heuristic and a severity rating (1-5 scale, minor:1-2, severe: 4-5) was given. In a local expert review Mankoff et al. created a master list of issues and severities.

Results All in all the results of the evaluation can be interpreted as an evidence that the "ambient" heuristics are an improvement on Nielsen's in the field of ambient displays. As Figure 2 shows a single evaluator discovers averagely 22% of known problems while 8 evaluators even find 70% of all problems. With Nielsen's heuristics a single evaluator finds 13% in average and 8 evaluators find half of all problems (50%). For an effective evaluation it is necessary that a small number of evaluators discovers a large number of usability problems. Using the "ambient" heuristic set the evaluators find 24 out of 26 problems with the Daylight Display and 35 out of 39 with the BusMobile. However 4 of the missed problems are not severe and 2 where severe. In none of the both conditions Mankoff et al. found a significant difference in the average number of problems found across the heuristics. That may be due to the fact that Nielsen heuristics are rather general, and issues can be addressed with a heuristic that is not a perfect fit. Also there were some problems that were never found by an evaluator using the "ambient" heuristics. The researchers came to the decision to merge their set of heuristics with Nielsen's heuristics to cover *all* issues. The titles of the final set of heuristics are as follows ("ambient" heuristics are in italics): 1) *Sufficient information design* 2) *Consistent and intuitive mapping* 3) Match between system and real world 4) *Visibility of state* 5) *Aesthetic and pleasing design* 6) *Useful and relevant information* 7) Visibility of system status 8) User control and freedom 9) *Easy transition to more in-depth information* 10) *"Peripherality" of display* 11) Error prevention 12) Flexibility and efficiency of use

In an early stage of ambient display design this set of heuristic can be used as an evaluation guideline. And therefore provides a inexpensive design technique without intense time commitment and costs. But one has to keep in mind when evaluating with heuristics the researcher evaluates a system without user interaction. Most often experts or at least people with experience in ambient system design and evaluation are recruited as evaluators. Therefore it is necessary to conduct user studies to test the device with people not involved in the design process nor specialized in the topic. This user-driven testing may be set in a laboratory (see following section) or in an *in-situ* deployment (see Section 5), i.e. the environment where the system is likely to be used (e.g. the user's home).

4 Evaluation in the Lab

In experimental research – like in communication science or HCI – laboratory experiments are common practice. Experiments in laboratories are that popular because they are easy to set up, therefore are rather cheap, and the researcher has good control of the experimental situation. The influence by confounding variables is kept small and study results are well interpretable. Furthermore laboratory experiments provide a good possibility to manipulate the independent variables. There are different kinds of experiments like interviews, surveys or paper-prototypes. The latter "typically require little effort, time, and money, to be applied" [2]. As their name says paper-prototypes are a very low-level stage in design process. In the field of ambient information systems that fact leads to the question how applicable they are for evaluating peripheral systems. What if users just look at the display because they wonder how a system made out of tins and carton works instead of looking at it because of the displayed information?

Especially when evaluating ambient information systems the advantage of good control over variables and situations turns out to be disadvantageous at the same time. Even adapting the laboratory environment to a real-world situation (e.g. designing a room to look like a living room) the extern validity stays small and transferring the results to the reality remains difficult. Hazlewood et al. [6] state that in a lab setting users have to be introduced to the AIS for comprehension reasons and then are distracted from the system again by doing some sort of task. When running the study over a extended time this method pushes the system to the user's periphery. But any probing of the researchers pulls it back into focus again. One may agree upon the fact that obscuring the information channel (the ambient display) by distracting the user is not the same as as the same channel existing within the periphery. Due to the limits of what can be learned from laboratory settings – especially with AIS – the benefit of evaluation results with ambient and peripheral implementations are questionable [7].

Lab Study

As in Section 3 this subsection picks up a case study and outlines the topic, the settings and the results. Hsieh and Mankoff conducted a study comparing two peripheral displays for monitoring emails [9]. Nowadays emails are an important part of business communication. But with an increasing number of received emails the number of irrelevant emails increases, too. A tool is needed to distinguish between important and relevant emails or needless and less-priority ones. According to their paper there are email clients that display a small envelop icon in the task bar (sometimes accompanied by sound) to make the user aware of incoming emails. This technique by itself is not interruptive, the awareness is high but the usability is low. A system not providing enough information whether or not an email is relevant forces users to interrupt – even if the technique itself does not – their current task and to switch applications. An even higher usability could be reached if the information is displayed in a way that it can be accessed without looking at the monitor.

Hsieh and Mankoff state that evaluating usability is basically important but for ambient displays especially awareness and distraction evaluation is necessary. Awareness means how easily the presented information can be observed by the user while distraction is the level an user gets distracted from his primary task by the presented information. Awareness and distraction data is gathered with questionnaires, self-reports or interviews. As there are no standard questions about awareness Hsieh and Mankoff quote that interviews in field studies may help to better understand awareness issues. Whereas distraction data is often gathered in lab settings where properties like response time can be observed more easily. To measure usability a researcher can fall back on traditional HCI techniques. Hsieh and Mankoff designed two peripheral displays (see Figure 3) capable of email monitoring without forcing the user to switch tasks to measure the former mentioned variables. They used pre-existing devices like a Ticker (3a) display and an Ambient Orb (3b). Both were modified to display specific information about email from people of interest.

The target group for which the displays were designed were administrative assistants who receive a large amount of emails. In interviews Hsieh and Mankoff came to know that persons from the target group check email frequently and that it is important to them to check who send the message. With this information in mind Hsieh and his team developed two peripheral displays notifying about incoming emails. Figure 3a shows the Ticker, a on-screen display showing scrolling text. The other display was an Ambient Orb sitting on the user's desk. The LEDs under the frosted surface are able to change the color in order to submit specific information (Figure 3b). The design was reconsidered after a heuristic pilot study where the ambient heuristics mentioned in Section 3 were used. Especially the distracting animations like flickering and blinking where removed.

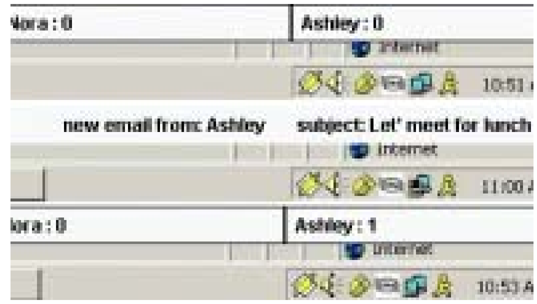
Following the pilot study Hsieh and Mankoff conducted a dual task study. Twenty-six students aged from 18 to 23 and used to use email participated. Their primary task was to sort a fake inbox containing 1500 emails. As a secondary task the participants had to monitor the ambient displays because they were told that they would be questioned about it after finishing the task. The participants were split equally in two groups, with each group working with a different display. In this between subject design both groups were situated in front of personal computers with either the Ticker display located at the bottom of the screen or the Ambient Orb put beside the monitor in user's focal vision.

The participants were asked to imagine that they are a famous CEO receiving a lot of spam, but also important emails from her/his three employees. Before the study started all participants had to pass a simple memory test to ensure that they could remember the names of their employees and their visual appearance on the display. If a new incoming email was send by one of the employees or one of ten familiar celebrities it had to be saved or removed otherwise. New emails were not visible in the inbox but only on the display because the sorting in the email client was from least recent to most recent. All in all 15 new emails arrived in random intervals (but previously determined, so that the conditions are the same for all users) during the whole study.

The study took 15 minutes of time. Three minutes were used to gather baseline data. During the left 12 minutes the participants should remember as much information about the ambient displays as possible. Because afterwards they would be asked questions. In a first step they were asked to self-report on awareness. Hereby question were similar to "how often did you look at the display" or "how much attention did you pay to the peripheral display?". In a second step the researchers asked objective questions about the information an user gathered from the display. Question of this sort were for instance "How many new emails did you receive from James" or "who did you receive the most emails from during the first half of the study?". The order of questions was chosen like this because specific content questions may influence on general self-reporting.

The results of the lab study were reached on the base of measures of awareness, distraction and usability. Objective performance records (like reaction time after arriving of a new email or total processing time) on the primary task and questionnaires about the peripheral displays gathered data about usability. Results of awareness and distraction were mainly based on self-reporting methods. One finding of the lab study was that self-reporting and knowledge questions were weakly correlated. Also participants preferred the display that was ranked most consistently high on awareness. According to Hshieh and Mankoff these two facts put together may be an indicator that self-reporting is a reasonable, low-cost technique for initial feedback. Note that this is just an brief excerpt of the total results of the study which will be analyzed more deeply in Section 6.

Of course Hsieh and Mankoff knew about the weaknesses of laboratory settings regarding ambient information systems but as they state "there is no consensus about the best approach for evaluating peripheral displays" they performed different evaluations in the field *and* in the lab. This is an important step towards exploring new evaluation



(a) Ticker

(b) Ambient Orb

Fig. 3: The Ticker and Orb displays. This figure illustrates the different states of each device. The ticker shows no unread email (top), then notifies about one new message by displaying the sender and the subject (middle) and last returns to his first state but indicates one unread email. Same for the Orb display but from left to right [9].

methods that consider the special requirements of ambient systems. Hazlewood et al. mention that these new methods will focus on long-term *in-situ* studies. They see the only way for ambient systems to become truly ambient lies in arrangements that conduct studies over a long period of time and "in the wild" [7].

5 Evaluation in the field

Field evaluation is a popular method of in today's research. Especially research concerned with evaluation of ambient information systems conjectures that field evaluation is the only way to find out how people interact with ambient systems in the real world [7]. In the last 10 years there have been many arrangements gaining their results in field studies [6, 12, 10, 18, 9]. Field evaluation is more expensive, more time consuming and more difficult to organize and survey than lab evaluation. But because of its high external validity (i.e. the relationship how well results of the study may be transfered to real world) it is crucial to AIS evaluation. There is some antagonism between internal and external validity. If the external validity is high the internal decreases and the other way round. Conducting an experiment in the participants environment (little transfer to reality necessary, high external validity) results in a situation where researchers are limited in their possibilities to control the situation (variation of variables and raising specific events is difficult, low internal validity). That is why field studies have to run over an extended period of time because the conductors often have to wait for a specific event to happen or a distinct situation to change.

Field Study

This paragraph summarizes the field study by Hshieh and Mankoff [9] conducted in the context of their comparison of two peripheral displays for email monitoring. Like in Section 4 we describe the setting and outline the results of the study. A detailed discussion of the results can be found in Section 6. Hsieh and Mankoff set up a field study with four participants. These were recruited based on their need of monitoring emails from a limited group of people. Furthermore it was important that their job did not exclusively focus on email management but also on other desktop work. Hsieh and his team elected administrators from their university department. Two of them were given an Ambient Orb and the other two used the Ticker application. The two displays were the same as described in the previous section.

Data was gathered over four weeks with two weeks with the display present wrapped by one baseline week at the beginning and the end of the study. Six times a day a window popped up on the participant's screen in random intervals which contained nine questions on email and display awareness. As results should be comparable these questions were similar to the ones asked in the questionnaire of the lab participants. In advance the Orb display was turned black and the Ticker became blank. The displays were shut down during the appearance of the pop-up because the participants should remember their use of the displays without being influenced when responding. If a participant decided for what reason ever to ignore the questionnaire it disappeared within one minute.

For better understanding of display usage Hsieh and Mankoff did a short interview after the participants had used the display for one week entirely. After two weeks of display usage each participant was interviewed in more detail and each one was asked to fill out the questionnaire which was formerly used in the lab. Furthermore the participants were questioned on usability. The questions were based on the heuristics mentioned in Section 3.

6 Comparison of evaluation methods

After giving an overview of the different studies and after outlining the several evaluation methods we compare in this section the findings of the former mentioned studies. We focus on a comparison of lab and field evaluation (Section 6.1) based on the results found by Hsieh and Mankoff [11]. Furthermore we compare an evaluation approach with "ambient heuristics" with a Wizard-of-Oz deployment, which is a form of field evaluation conducted by Consolvo and Towle [5].

6.1 Lab study compared to field evaluation

The results of Hsieh and Mankoff's study [9] briefly outlined in Section 4 and 5 are discussed in this subsection in more detail. Data was gathered about awareness, distraction and usability. Participants answered self-reporting and objective knowledge questions in order to provide measurements for awareness. Thus information about how much attention they payed to a display and how much information they retained from a display was collected. Distraction was measured as well with self-reporting questions and data concerning accuracy and speed from the primary task. Interviews and a heuristic rating with the participants measured usability.

Results Originally Hsieh and Mankoff formulated four hypothesis, in this paper we will examine only two of them (names and descriptions are taken from the original paper):

- **D1** The Orb is less distracting, and supports a higher level of awareness, than would the Ticker.
- **T2** The level of awareness supported by a display in the lab correlates with the same in the filed. Similarly, the level of distraction will correlate.

D1: As mentioned before awareness was measured on the one hand by self-reports. Lab participants were slightly less aware of the Orb than of the Ticker due to their own statements. Participants using the Orb in the field told they were very aware of the system. The same was reported by the field users of the Ticker who were only slightly less aware of their display. On the other hand participants were asked knowledge questions on how many emails had arrived. In the lab[1] the Ticker users were averagely rated with 3 at maximum of 5 points whereas Orb users reached 2. In the field a flaw in the study design caused limited data. The participants answered in most cases that no emails had arrived because they were asked for the last 15 minutes (the study ran over four weeks).

That the Orb is less distracting could not be proved. In the lab setting participants using the Orb as well as those using the Ticker self-reported the same value. In the field all participants (whatever display they used) reported that they were not at all distracted. One of the Ticker users even would have liked something more distractive like some flashing animation instead of simply scrolling text.

Somehow the results on awareness are surprising as one would assume that the forced exposure to the displays in the lab situation results in a high awareness. However it speaks in favor of ambient displays when they obtain high awareness ratings in field evaluations. The second interesting finding has some society aspect. Are we that accustomed to flickering and blinking signs and pop-ups that a scrolling text is no more distracting enough to catch our attention?

T2: Contrarily to *D1* this hypothesis is of a more technical nature where *D1* was more design concerned. Based on the results of the different evaluations Hsieh and Mankoff quote this conjecture to be true in most cases. The level of awareness and distraction of a display correlates in between lab and field study. The measure of awareness for the Orb display in the field was different. The participants stated that they could monitor it even when not being at the desk and therefore gave high ratings on awareness. The participants in the lab, who had to sit in front of the computer and who interacted with it most of the time, logically where not able to experience that.

6.2 Heuristic evaluation compared with a Wizard-of-Oz deployment

This subsection refers to a system called "CareNet" which was developed by Consolvo, Roessler and Shelton in 2003 [4]. The CareNet Display can somehow be seen as a successor of the Digital Family Portrait [12]. It is designed for network members in an elder's care network. While the Digital Family Portrait was a bi-directional display (i.e. the elder and the opposite part obtained a display), CareNet

[1] We assume that unfortunately there is a mistake in the paper of Hsieh and Mankoff. They report two times about the results in the field. Throughout their result report about hypothesis D1 they compare between lab and field results.

is a single direction approach (i.e. only the caregivers have a display). Furthermore the CareNet display is designed to be both ambient and interactive. The display itself is embedded in a normal, wooden picture frame. Being in the ambient mode it displays a portrait or picture of the elder and a set of icons representing the elder's condition. To get enhanced information about a specific icon or condition the user can use the display interactively by touching the screen.

The ambient mode surrounds the picture with up to seven types of icons. They stand for medication, outings. activities, mood, falls and calendar. They can be varied e.g. to display that everything is okay, something unexpected happened or the system is not working. The display also is capable to show more than one icon of one type in order to display information about e.g. breakfast, dinner and lunch. If caregivers like to know more about a specific icon (i.e. a specific condition or event) they can interact with the display in a way that they touch the icon of interest. Then the picture of the elder is replaced by detailed information about e.g. *when* the elder had breakfast, *what* she or he ate or what the unexpected event was in detail. In addition a five-day trend can be viewed (see Figure 4). The displayed data is collected by sensors or people. That may be the elders themselves or network members (e.g. a nurse, the food delivery man or a family member). However sensors could track when the elder took which medication. In their deployment Consolvo et al. only used people not sensors but they state that the system was designed to collect data with sensors, too.

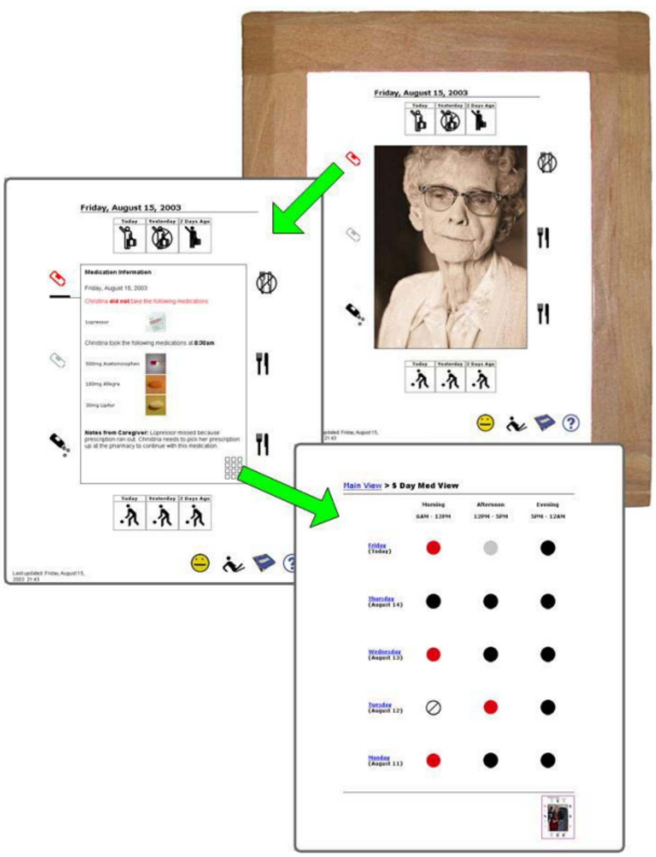


Fig. 4: *Top right*: main screen of CareNet Display prototype. *Left*: The overall picture of the elder's condition switches to an event detail when users interact with the display. *Bottom right*: Users can monitor a 5-day trend view for each event [4].

To check how successful the heuristics proposed by Mankoff et al. [11] meet the criteria for heuristic evaluation of ambient displays Consolvo an Towle [5] conducted two evaluations of the CareNet display and compared the results. One was an evaluation with said heuristics and the other was an *in-situ* Wizard-of-Oz user study. The heuristic evaluation was conducted similarly to the approach in Section 3. Eight evaluators were asked to interact with an online prototype of the CareNet display and had to create a list of heuristic violations. A detailed description with pictures of the display was provided on the world wide web, too. The Wizard-of-Oz evaluation ran over three weeks with 13 elder participants. Per elder two or three family caregivers participated, all of them were given a CareNet display. All participants took place in pre- and post-study interviews. The researchers updated the displays remotely with data from multiple phone calls per day [4].

Results In this study three to five evaluators found 39%-55% of known issues (this meets Nielsen's definition of a successful heuristic evaluation where 3-5 evaluators find 40%-60% of known issues). Known issues are issues found in the *in-situ* evaluation. A single evaluator found one to three of the eight usability problems. Seven evaluators revealed the same issues while three of them identified a problem none of the others found. When aggregating these findings the evaluators found 75% of all known usability issues. That means that two of the eight problems were missed in the heuristic evaluation. But the evaluators mentioned 60 violations that were not discovered in the *in-situ* evaluation.

The *in-situ* evaluation for example taught the researchers that the icon for "falls" could be replaced by "household needs" (e.g. the bathroom's light-bulb has to be changed). An initial design study that should determine the icons used in the display showed instead that the "falls" icon ranked at the top and the "household needs" icon ranked last. After testing and living with the display the users felt information about falls being still very important but suggested that household needs would be better suited [3]. Another issue that was not discovered by the heuristic evaluation was the most severe one. In fact the display glowed brightly in low-light conditions. This stopped the display from being ambient because it was that distracting. Heuristic 10 *"Peripherality" of display* would have matched perfectly but none of the evaluators reported a violation [5].

Furthermore Consolvo et al. mention that of the 60 reported issues in the heuristic evaluation only some would slip in to a redesign of the display. Despite most of these violations were concerned with heuristics 6 and 10 (*Useful and relevant information* and *"Peripherality" of display*) comments of the *in-situ* participants refuted the issues reported by the evaluators, in some cases. Because the evaluators were asked to rate all the violations from a aggregated violation list it becomes clear that many evaluators did not agree upon the 60 issues. So the huge set of reported issues could be shrunk to a list of issues upon which at least 2 or more evaluators agree.

We can summarize the results as follows. 5 evaluators found up to 75% of known issues. I.e. heuristic evaluation provides a discount but efficient methodology. But at least one very crucial issue (the glowing in the dark one) was missed. Additionally not all of the usability issues (as we see that may be a quite large set) should be considered in a redesign of a display. Therefore heuristic evaluation can not replace *in-situ* studies but can be useful in an iterative design process to eliminate problems in an early step. Afterwards the results of an *in-situ* evaluation may be more effective by potentially revealing more or at least unknown and new issues.

7 Conclusion

In this paper we present an overview on evaluation methods for ambient information systems. Based on different case studies we examine the pros and cons of heuristic, laboratory and field evaluation. We find that heuristic evaluation is a discount, little time intense and an organizational easy to conduct approach. Furthermore there is evidence that it reveals most crucial usability issues. But it is not a dead proof method to find *all* issues in an ambient display design. Lab settings may result in organization effort, are more time consuming and more expensive than heuristic evaluations. They provide good control over the evaluated display and the surrounding environment but especially for AIS it is difficult to map the gained results to a real world application. Studies in the wild, as field studies are often called, are able to reveal some issues that are missed by the heuristic evaluation. But one has to keep in mind the difficulty of conducting such studies. The amount of time which is necessary to e.g. call participants several times a day over various weeks (that may include weekends, nights and holidays) is an important issue [3].

There is no consensus in the ambient system research community about which approach, lab or field study, is the best. Hsieh and Mankoff therefore conduct both settings [9]. As many researchers prefer an *in-situ* approach for evaluating ambient displays (e.g. [6]) because it is "nearer to reality" it is questionable how long lab studies are conducted any more. Abowd et al. [1] provided a living laboratory concept that merges lab and field settings. That is a step towards the consensus in current literature that new, better suited methods are needed for the evaluation of ambient information systems. One of these methods could be the framework provided by Holmquist [8] who states that most HCI experiments with ambient displays require that the user already knows *that* and *what* information is displayed by a certain device. Is recognizing that some ambient device displays information really that easy? And if an user finds out that information is displayed the next step may be even harder (see Figure 5): what is displayed[2]?

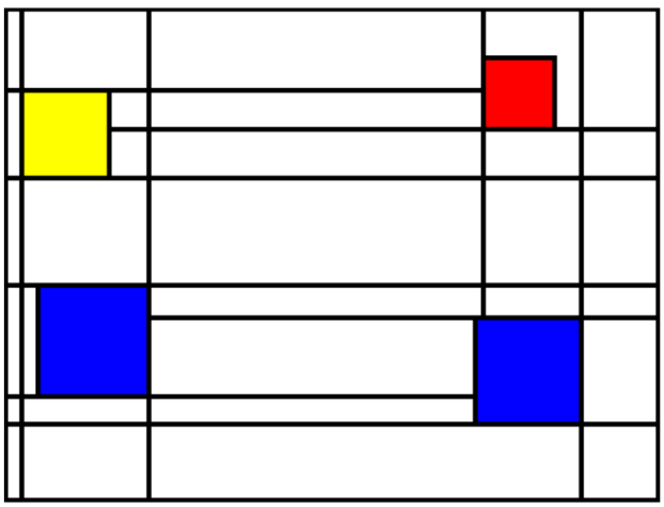


Fig. 5: Intelligent Art. A picture taken form Skog's work about Aesthetics and Utility: Designing Ambient Information Visualizations [17]

References

[1] G. D. Abowd, C. G. Atkeson, A. F. Bobick, I. A. Essa, B. MacIntyre, E. D. Mynatt, and T. E. Starner. Living laboratories: the future computing environments group at the georgia institute of technology. CHI EA '00, pages 215–216. ACM, 2000.

[2] S. Carter and J. Mankoff. Challenges for ubicomp evaluation. Technical Report UCB/CSD-04-1331, EECS Department, University of California, Berkeley, 2004.

[3] S. Consolvo, B. Harrison, I. Smith, M. Chen, K. Everitt, J. Froehlich, and J. Landay. Conducting in situ evaluations for and with ubiquitous computing technologies. pages 103–118. Taylor & Francis, 2007.

[4] S. Consolvo, P. Roessler, and B. Shelton. The carenet display: Lessons learned from an in home evaluation of an ambient display. volume 3205 of *Lecture Notes in Computer Science*, pages 1–17. Springer Berlin / Heidelberg, 2004.

[5] S. Consolvo and J. Towle. Evaluating an ambient display for the home. CHI EA '05, pages 1304–1307. ACM, 2005.

[2]The picture displays bus departure times. Each bus is modeled as an square in the color blue, yellow and red. A smaller square means that less time remains until the bus leaves the bus stop. The color supports this information. If the square is blue you have plenty of time, yellow indicates to get prepared and start walking to the bus stop and red means you have to hurry to catch the bus [17]

[6] W. R. Hazlewood, K. Connelly, K. Makice, and Y.-k. Lim. Exploring evaluation methods for ambient information systems. CHI EA '08, pages 2973–2978. ACM, 2008.

[7] W. R. Hazlewood, E. Stolterman, and K. Connelly. Issues in evaluating ambient displays in the wild: two case studies. CHI '11, pages 877–886. ACM, 2011.

[8] L. E. Holmquist. Evaluating the comprehension of ambient displays. CHI EA '04, pages 1545–1545. ACM, 2004.

[9] G. Hsieh and J. Mankoff. A comparison of two peripheral displays for monitoring email: Measuring usability, awareness, and distraction. Technical Report UCB/CSD-03-1286, EECS Department, University of California, Berkeley, 2003.

[10] N. Jafarinaimi, J. Forlizzi, A. Hurst, and J. Zimmerman. Breakaway: an ambient display designed to change human behavior. CHI EA '05, pages 1945–1948. ACM, 2005.

[11] J. Mankoff, A. K. Dey, G. Hsieh, J. Kientz, S. Lederer, and M. Ames. Heuristic evaluation of ambient displays. CHI '03, pages 169–176. ACM, 2003.

[12] E. D. Mynatt, J. Rowan, S. Craighill, and A. Jacobs. Digital family portraits: supporting peace of mind for extended family members. CHI '01, pages 333–340. ACM, 2001.

[13] J. Nielsen. Determining the number of evaluators. `http://www.useit.com/papers/heuristic/heuristic_evaluation.html#evaluatornumber`. [Online; accessed 15-June-2011].

[14] J. Nielsen. Heuristic evaluation. `http://www.useit.com/papers/heuristic/`. [Online; accessed 15-June-2011].

[15] J. Nielsen. Ten usability heuristics. `http://www.useit.com/papers/heuristic/heuristic_list.html`. [Online; accessed 15-June-2011].

[16] Z. Pousman and J. Stasko. A taxonomy of ambient information systems: four patterns of design. AVI '06, pages 67–74. ACM, 2006.

[17] T. Skog, S. Ljungblad, and L. E. Holmquist. Between aesthetics and utility: Designing ambient information visualizations. pages 30–31. IEEE Computer Society, 2003.

[18] J. Stasko, D. McColgin, T. Miller, C. Plaue, and Z. Pousman. Evaluating the infocanvas peripheral awareness system: A longitudinal, in situ study. Technical Report GIT-GVU-05-08, College of Computing/GVU Center Georgia Institute of Technology, 2005.

[19] M. Weiser and J. Brown. Designing calm technology. *PowerGrid Journal*, pages 75–85, 1996.

Improvements for Ambient Information System Evaluations: An Assessment Based on Current Systems and Their Evaluations

Simon Eumes

Abstract— In this paper, we present an overview over current evaluation techniques typically applied to Ambient Information Systems. This includes a modified version of Nielsen's heuristics as well as the Wizard of Oz technique. On six cases studies we present the usage of those methods in real life evaluations and also discuss the shortcomings in those specific circumstances. Finally a four step framework is presented, introducing a structured process through the idea, development and evaluation phases. This approach also includes the justified question whether or not the proposed system is in accordance of the ambient rule. By running through that process in an ordered manner, it will strengthen believe in the idea and also raise the confidence in the system.

Index Terms—Ambient Information Systems, Evaluation, Methods, Use Cases, Peripheral Vision, Heuristics, In Situ, Wizard of Oz

1 Introduction

Ambient Information Systems are becoming more and more popular and we see a new kinds of ambient systems emerge every day. But what most of those proposals do not include is a comprehensive evaluation of the system, mostly because the lack of knowledge about evaluating ambient systems. But to understand the difficulties we first have to understand what ambient system are. Pousman et al. [10] propose the following definition for Ambient Information Systems:

- The ambient display is intended to convey non-crucial but otherwise important information.
- The ambient system resides typically in the periphery but may become more prominent for a small fraction of time .
- The focus of the system lies on the appropriate tangible setting within the environment
- The ambient system should provide updates on their conveyed information by subtle changes of their state (and therefore not to be distracting but noticable).
- The ambient system should be adapted to the environment it is being placed in. Furthermore an appropriate design must be present.

Even though that such a definition (and also other similar) exists, we can find a lot of displays claiming to be ambient without satisfying the underlying definitions. Often either the system conveys critical information or the information is not perceivable with a glance. Due to the fact that Ambient Information Systems try to present information without distracting the user, but at the same time be as informative as possible, it is difficult to find a suitable evaluation framework. The following sections try to give an overview over current methods and technique. Later on six cases studies are presented and their approach to evaluating the ambient system is discussed.

In chapter 2 we will give an overview of evaluation methods that are suitable for ambient systems. In the following chapter 3 we are presenting case studies covering current evaluations of ambient systems. Chapter 4 introduces a proposal for a structured approach to evaluate Ambient Information Systems and chapter 5 finishes the paper with a conclusion.

- *Simon Eumes is studying Media Informatics at the University of Munich, Germany, E-mail: eumes@cip.ifi.lmu.de*

- *This research paper was written for the Media Informatics Advanced Seminar on Ubiquitous Computing, 2011*

2 Overview of Evaluation Methods for Ambient Information Systems

The following section introduces methods, types and techniques already used in past user studies targeting Ambient Information Systems. The evaluation methods describe the techniques used to conduct the study; additionally the attention is shifted to the location, discussing whether a laboratory study is still suitable for ambient systems. Holmquist [6] mentions a valid concern in point out that most evaluations/experiments should start way ahead of the final evaluation at the end of a development process of an Ambient Information System. He proposes a three step framework for evaluating Ambient Information Systems. These three levels of comprehension are:

- To gasp **that** something is being visualized
- To realize **what** information is displayed
- To comprehend **how** information is prepared

Normally evaluations would take place at the third level, testing only the outcome of project. Many projects might not even make it through the first two levels and applying this framework will often save a lot of effort and time. If a user isn't able to distinguish what information an Ambient Information System is displaying it is nearly impossible to answer/evaluate the how. And if we go a step further and argue that a user is not even aware that information is displayed the whole set-up is rendered useless (regardless of the how and what).

2.1 Heuristic Evaluation

The idea of heuristic evaluation dates back to Nielsen [9]. He presents heuristics as an ”informal method of usability analysis”. Based on those initial heuristics Mankoff et al. [8] propose a customized variant of the Nielsen heuristics targeted at Ambient Information Systems. The proposed heuristics are:

Sufficient information design Limit the amount of information transported by the system to a minimal. Only information really needed should be displayed in order to keep the interface simple and glanceable.

Consistent and intuitive mapping In order to keep the system informative and easy to use ”add minimal cognitive load”. Also use representations familiar to the user and try to minimize the mental work to be done by the user.

Match between system and real world Try to match the users level of knowledge and use a language understandable and comprehendible by the user. This is especially true for different age groups and domains of knowledge.

Visibility of state The states of a systems should be easily distinguishable by the user. Try not to visualize different states with nearby representations (e.g. with similar colors). It is also important that a transition between different states must be perceptible by the user.

Aesthetic and pleasing design When placed in the intended scenery, the system should be appealing but at the same time not outstanding. It might even be a good idea to integrate the system/display into common object found in the setting.

Useful and relevant information Have a purpose for the system or display and convey only information necessary and relevant to the user. Also constrain your information coverage to an acceptable limit.

Visibility of system status There should be an easy to perceive information loop, informing the user about the current system status and what is being done. This feedback should be updated and respond in a timely manner.

User control and freedom Anticipate user input mistakes or unwanted interactions. Present an ”emergency exit” to quickly recover to the default state without the hassle of navigation through a multitude of dialogs. For some systems it might also be necessary to provide redo and undo actions for the user.

Easy transition to more in-depth information In case your Ambient Information System features different information levels it should be easy to reach them. The transition should be as simple as possible without having to go through multiple steps of interaction.

”Peripherality” of display To maintain the concept of an Ambient Information System, the display should remain discreet until a change is happening and the user needs to be notified. The user should nevertheless be able to monitor the system from a glance without any troubles.

Error prevention The system should be unbreakable with regard to user interaction and also anticipate and prevent possible errors and error sources. Try to design your system in a way that errors are not doable.

Flexibility and efficiency of use Try to accommodate both, the novice and also the experienced user, on the Ambient Information System. Provide an option to speed up certain tasks and also to modify interaction patterns.

The typical way to use those heuristics would be to create a list of heuristics tailored to your system and provide them together with either a detailed description or the real prototype to usability and user interface experts for review. The experts then evaluate the given prototype based on the heuristics and report back to the developer. Normally one would ask six to eight experts to evaluate a system to make the most out of this evaluation technique.

2.2 Wizard of Oz

The name relates to the book The Wonderful Wizard of Oz by L. Frank Baum [1] where an ordinary man pretends to be a wizard solely by trickery. The method was first described by Kelley [7] in 1983 and is still a viable evaluation technique, especially in the early stages of a development process. The basic idea behind this method is that the system is still a mock up/prototype and a person (the wizard) is controlling/mimicking all actions and user interactions. This allows to conduct early evaluations without having to build the complete system and also react to unforeseen actions during the trial more flexible. But it also presents a disadvantage because the interactions are most definitely different from when one would have a complete autonomous system running during the evaluation.

This is especially true with regard to the evaluation of Ambient Information Systems. While remote control of an otherwise finished prototype might still be feasible (e.g. via a network connection), presenting a puppet play type of Wizard of Oz evaluation will definitely distract the users and contradict the ambient aspect. The prototype most probably will not blend with the environment (e.g. due to its size) and conducting a puppet play over a longer period of time and always reacting in the same manner is not applicable.

2.3 In Situ vs. Laboratory Studies

In classic usability evaluations one often has the choice of selecting a fitting environment for the study. To monitor a participant in a controlled environment often a laboratory or laboratory like environment was chosen. This allowed the instructor to have the same influences over a longer period of evaluations.

To evaluate Ambient Information Systems another type of study comes to the fore. Because an ambient system is per definition only a passive and secondary information source, a controlled environment (like in a lab) would contradict its intended purpose. And this is where an in situ study is the first choice. It’s a study that is taking place right at the intended context and normally covers a longer period of time (Consolvo et al. [3] decided on a three weeks period for their CareNet display, Strasko et al. [11] choose a two weeks time frame for their InfoCanvas evaluation). This shifts the attention away from the evaluated system and allows the user to interact as intended with the system. This also allows observing long term effect and the impact over time on the user.

2.4 Discussion of Methods

Neither the heuristic nor the Wizard of Oz method is probably a perfect evaluation method for ambient systems. But they are a good starting point from where to begin the evaluation and maybe derive a new method/framework for Ambient Information System evaluation. But was becomes more and more clear is that independent of the respective method the evaluation must take place in situ. Without the proper surrounding setting merely all evaluations methods miss some crucial usability problem and hinder the participants to interact with the system in a normal way.

3 Case Studies of Current Assessments on Ambient Information Systems

There have been a lot of publications on Ambient Information Systems but only a small fraction of them cover the topic of evaluation. In addition, if a new system was introduces only few of them included a proper evaluation. The following six cases all feature an evaluation of the system and are discussed and opinions are shared.

3.1 CareNet

The CareNet Ambient Information System aims at helping caretakers of elderly people to gasp the current status of their entrusted. It was developed by Consolvo et al. [2, 3] for the purpose of testing different evaluation methods. Here specifically the heuristic evaluation as well as an in situ Wizard of Oz was used. The frame can be seen in figure 1.

As mentioned the system is targeted at persons taking care of their elderly family or friends. It provides an overview of the elderly person’s state throughout the day, including meals, medication and activities. The information is displayed on a customized picture frame, housing a touch tablet behind a wooden bezel. The caretaker is supposed to place this at their home and use the provided information as an informal reassurance that the cared person is okay. On the display a picture of the elderly person as well as some icons informing you about the state can be found.

In contrast to the class Ambient Information System the display can be used interactively offering additional information after navigation through the system. This more or less defeats the purpose of ambient displays and goes more into the direction of a general interaction system, but for the sake of argument we regard it as an Ambient Information System.

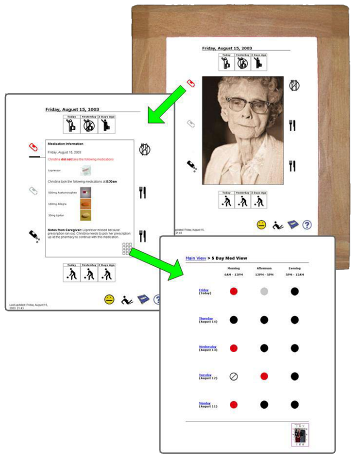


Fig. 1. A mockup of the user interface of the CareNet Ambient Information System [2]. A 15 inch monitor serves as the picture frame used in the prototype.

For evaluation purposes they relied on two different techniques, namely a heuristic evaluation as well as a Wizard of Oz experiment. For the heuristic evaluation, they incorporated the adapted heuristics for ambient system by Mankoff et al. [8]. They used all twelve heuristics proposed and thus covered topics including visibility of state and sufficient information design. The in situ Wizard of Oz evaluations took place over a period of three weeks and were conducted in the respective home of the person cared for. The display was updated remotely via a researcher after communicating to the elder people via telephone.

Even though the heuristic evaluation found a multitude of violations and usability issues, only the in situ evaluation was able to expose major issues. Those included the disturbing bright light of the display in low-light conditions, and in addition the viewing angles of the display were insufficient.

Due to the size of the display and the multitude of information conveyed, we personally do not consider this to be an ambient system. You would not be able to gasp the information provided just by glancing at the screen. You have to actively interact with the system (even if it is only going to stand directly in from of the display) to perceive the updates. This contradicts the basic Idea of an Ambient Information System. Also the ability to display crucial information (has the elderly person fallen on the ground) to the caretaker might influence the participants to not only see the display as an ambient system rather than as a normal information screen.

3.2 BusMobile

Mankoff et al. [8] created the BusMobile Ambient Information System for the purpose of evaluating their proposal for ambient evaluation heuristics. The purpose of the system is to inform about the departure times of nearby bus lines at a nearby stop. An exemplary assembly can be seen in figure 2.

On the top of the system resides an umbrella, hiding the bus tokens if the associated bus is further away than 25 minutes. If a bus closes in on the station the token initially advances to the bottom of the apparatus and closes in to the umbrella according to the distance to the targeted bus stop. Thus the bus token nearest to the umbrella is the one closes.

This systems was deployed in a university lounge to inform students and staff about the arrival of buses at the terminal located next to the university. Even though the system is build in an easy manner, it depicts a Ambient Information Systems that is informative and serves a real purpose. It would be interesting to compare this type of display to an ordinary numerical time-to-depart display.

Together with the Student Feedback Orb (Hazlewood et al. [4, 5], see section 3.3) this system resembles an ambient system closest ad-

Fig. 2. Image of the BusMobile Ambient Information System [8]. It displays bus departure times from a nearby terminal. The time frame roughly corresponds with the distance to the top.

hering to the definition of ambient systems. It transports non critical information and is more or less unobtrusive in with its integration into the environment. The information can be perceived at first glance and the familiarization is more or less instantaneous.

3.3 Student Feedback Orb

In an effort to create a simple Ambient Information System for the purpose of exploring evaluation methods, Hazlewood et al. [4, 5] created the Student Feedback Orb. The ambient display is represented by a light orb with changing light colors. The system is targeted at university constructors with the purpose of displaying real time evaluation results of their lectures. Students are able to rate individual lectures via an email request after each session. The median grading is then pushed to the orb who then displays a certain color associated with the grade (red to green).

The intent was to create an unobtrusive display that would trigger an interest by the constructors who then in turn would request more specific information about the ratings on a dedicated website. The main components can be seen in figure 3.

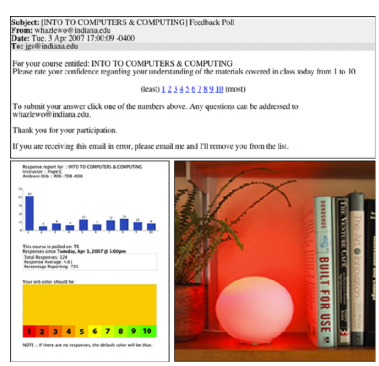


Fig. 3. The orb as well as the more detailed information website of the Student Feedback Orb Ambient Information System [5]. It changes color from green to red based on the evaluation of teaching classes by students.

To evaluate the system they intentionally set up a two month long in situ study without interruptions during this period (e.g. by an interview). During this time it was difficult to gasp the usefulness of the system because for most of the time the grading was more or less consistent. This in turn let the orb keep a certain color over a long period of time and discouraged the participants to check the detailed website. And this is exactly the difficulty in evaluation Ambient Information Systems. Because of their informal character it is difficult to differentiate between intentional and unintentional oversight.

Even though the system seems to be more minor, this is exactly what we understand as an ambient system. It has a dedicated purpose and is blending in fine with the environment. One could argue that the constant glowing of the orb would distract the user - and this might be a valid objection but in general this is only a minor issue. The size and the intensity could be changed easily.

3.4 InfoCanvas

Stasko et al.[11] present an Ambient information System displaying an arbitrary number if information within an interactive picture frame (build from a plain 15 inch monitor) with a frame as a bezel. The user has the freedom to specify the information to be displayed (e.g. stock values or the weather) and the visual element it is mapped to. For example an animal lover could use a zoo as an overall theme and place different kinds of animals in it depending on the information state. Another example would be a landscape image changing the sky color or the number of people in it. Two example themes can be seen in figure 4.

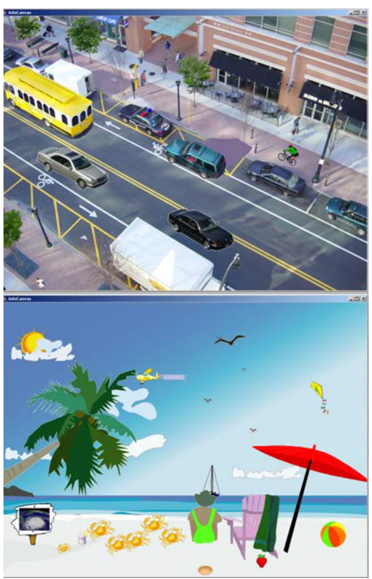


Fig. 4. Two examples for themes in the InfoCanvas framework [11]. Users can place different objects on a canvas representing previously defined states or information.

They offer an open framework to specify the information source and the respective influence on the theme. This guarantees that the user has full control over the environment and also can tightly integrate the system in their familiar surroundings.

The evaluation type was an in situ study over the period of one month, allowing the participants to get familiar with the display. During these four weeks a number of interactive sessions were conducted to monitor the progress and to complete survey forms. This wraps up the used method, being a more general approach, suitable for a lot of different systems. They prepared a catalog of objectives including usefulness, personalization and flexibility, aesthetics, distraction novelty and fun as well as a summary impression.

As a conclusion they have six lessons learned from their experiments. These include among others the need for personalization is important. For users to accept an Ambient Information Systems it must allow for customization by the user, and thus being able to be fully integrated into the residing environment. Additionally the consolidation aspect is significant as well. Allowing the integration of different information into one entity helps the user to comprehend information with only a glance.

At first glance this system might not be perceived as an ambient display but on closer examination it fulfils most of the ambient system definitions. The only thing we would argue is the aesthetically pleasing aspect. While we know that this is only a prototype, one would expect to have a more appealing interface to work with, especially in the phase of an in situ study.

Fig. 5. The arrangement of the clouds in The Clouds Ambient Information System [5]. They present the ratio between persons taking the elevator and persons taking the stairway within the building.

3.5 The Clouds & Follow the Lights

In contrast to the other use cases presented before the two cases presented here have a special property: They were introduced and launched at a university without telling the users the original purpose of the system. Theses ambient systems were developed by Hazlewood et al. [5] and deployed over a three month period in a shared place. The intent was to alter a simple behavior without giving the behavior away.

As seen in figure 5 the The Clouds ambient display consists of clouds hanging from the ceiling. There are two different sets of cloud present, each representing one category. The orange clouds represent the usage of the elevator to navigate through the building, whereas the grey clods map the usage of the stairs. The height at which each cloud cluster is hanging, represent the usage statistics of the stairs and elevators respectively.

The Follow the Lights display (figure 6) makes use of LEDs to lure the users to use the stairs instead of the elevator. It presents the users a guidance system by letting the light run towards the stairs. If a user is not responding to the initial request (e.g. keeps walking towards the elevators), the LEDs start pulsating in a red light to mimic disappointment and rage.

The evaluation was conducted in situ over a period eight weeks. After two weeks of just remote observations some interviews were carried out with 25 randomly selected persons in the building. After four weeks everybody in the building received an invitation to an online survey.

Both systems more or less violate Holmquist's [6] three levels of comprehension. Neither the **that** nor the **what** is being regarded. But in this case this was intended by the author. The goal was to evaluate the perception of different systems by the users. The survey showed

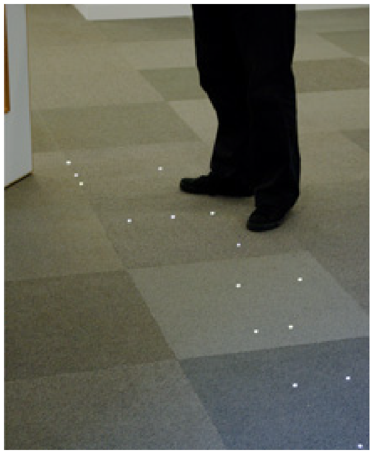


Fig. 6. Integrated lights into the floor in the Follow the Light Ambient Information System [5]. They are intended to subconsciously guide people to use the stairway instead of the elevator.

that the intent of the Follow the Light ambient display was the easiest to understand, while the The Clouds ambient systems was more cryptic. Most of the users had figured out the relationship between stairs and elevators but the exact mapping was unknown. With respect to interrupting the in situ study with surveys they state that a tradeoff must be made between keeping the study clean and the quality of data collected during the study.

We argue that neither one of the systems represent a good example for an Ambient Information System. The Follow the Lights system is definitely intrusive and distracts the user and the The Clouds ambient system leaves the user in uncertainty or even in disbelieve that the system has any purpose.

3.6 Remarks

We have seen six completely different Ambient Information Systems in the past sections. They all used different method for evaluating their studies but merely all of them used an in situ type of evaluation. This is the most common denominator thus far and is probably influenced by the fact that Ambient Information Systems are best reviewed in their normal surroundings.

It is in their nature to be unobtrusive and the information displayed is not crucial so oversights are surely intended by the system. And exactly this presents the difficulties in evaluating these systems.

Additionally all presented case studies tend to violate the basic rules of an ambient system to some extent, either by being obtrusive towards the user or not blending in with the environment. This is particular frustrating because the underlying ideas are very interesting.

4 A Proposal for Evaluation Guidelines for Ambient Information Systems

We have seen that evaluations based on heuristics help to drastically diminish the number of issues with a prototype. But this is in no way a replacement for an in situ evaluation. In our opinion, an in situ study is mandatory for evaluating Ambient Information Systems. Many errors have not been found by utilizing classic evaluations and relied on an in place evaluation over a longer period of time instead. There surely is the possibility to evaluate the general usability and aesthetics without in situ placement but to fully assess an Ambient Information System an in situ study is required. The following is our guideline recommendation for further Ambient Information System evaluations:

1. Is your system complying with the ambient rule?
2. Create video prototypes to allow for early feasibility and acceptance testing
3. Use heuristics to evaluate the general usability of the system
4. Conduct a exhaustive in situ study without interruption during the trial

4.1 Is the System Complying with the Ambient Rule?

We deem Ambient Information Systems to be

1. purely passive
2. non interactive
3. highly integrated into the environment

While exceptions for the passive and interactive part might be valid (e.g. by providing a simple touch interface to acknowledge the forwarding of information or reset the device) the other rule (highly integrated) is not to be tampered with.

If complex user interactions are required the fundamental ambient idea is rendered ad absurdum. In addition the system should be as generic and unobtrusive as possible and respect the glanceable aspect of such a system. This means that information must be perceivable on a large scale and not only by specifically targeting the system.

In addition the developed system or idea must also comply with the definition of an ambient system by Pousman et al. [10]. This guarantees that the basic idea of an ambient system is still preserved.

4.2 Create Video Prototypes to Allow for early Feasibility and Acceptance Testing

In order to get a fast feedback about the usefulness of a planned system a video prototype is an optimal solution. Without having to build the actual prototype one is able to test the core elements in the targeted surroundings. This allows for others to give immediate feedback and might event prevent you from invest time and resources in a concept that might not even work in the end.

As a bonus, the prototype allows one to quickly inform others about the idea and get other people hyped about the system. One is able to go through a lot of iterations very fast to achieve a solid foundation for the actual prototype that will be built.

4.3 Use Heuristics to Evaluate the General Usability of the System

After having the idea tested with a video prototype and after applying the needed customizations it is time to test the prototype with a heuristic evaluation. One could either rely on some predefined heuristics (e.g. the one established by Mankoff et al. [8]) or create a set of individual rating scale. We encourage everyone to use well established heuristics as a starting point but readjust them if the Ambient Information System needs it.

It is also crucial to focus on the purpose of the system. If no user interaction is required, a heuristic covering exactly this option should be disregarded (e.g. the "User control and freedom" heuristic by Mankoff et al. [8]). Our approach to Ambient Information Systems would also one to try to minimize the required interactions to the point of none at all. Only an interaction to reset or acknowledge should be present.

Using a heuristic evaluation before going into in situ trials will save a lot of time in finding issues early and indirectly help the in situ study be eliminating generic issues.

4.4 Conduct an Exhaustive In Situ Study without Interruption During the Trial

For an Ambient Information System it is mandatory to conduct an in situ study at the end. This allows to observe the system in it natural habitat and identify the remaining issues. While conducting the in situ trial it is crucial not to interfere in any way during the study (e.g. by having interviews or other kinds of visits). This will influence the participants and shift the focus back to the system and thus tamper with the in situ aspect. By having tested the system within its final environment one can be sure to catch the remaining issues and is able to fix them ahead of launch. Most of the time the issues found in the in situ study would not have been found by any other evaluation method.

The in situ study should be carried out with a functioning prototype and not by using a Wizard of Oz style or similar techniques. The user will most definitely perceive this fact and have another reaction to the system in general.

5 Conclusion

By looking at those six use cases we can see that a unified evaluation process for Ambient Information System is not in place right now. Everyone uses a technique they deem the most appropriate or the easiest one. By doing so we are basically eliminate the possibility to compare different systems on the same level.

But one must not forget that this field of research is still relatively young and only a few persons have dedicated their work the evaluation processes of Ambient Information Systems. We think that new evaluation frameworks and techniques will present themselves in the near future, also because the developed displays become more and more sophisticated and more and more researchers recognize the important of the evaluations.

The proposed four step framework will help one to go through the idea, development and evaluation phases. They build confidence in the system as it advances and allows one to model the display as ambient as possible. It also reminds you to adhere to the ambient idea and lets you already customize the system accordingly in an early stage of development.

References

[1] L. F. Baum. *The Wonderful Wizard of Oz*. George M. Hill Company, 1900.

[2] S. Consolvo, P. Roessler, and B. Shelton. The carenet display: Lessons learned from an in home evaluation of an ambient display. pages 1–17. Springer Berlin / Heidelberg, 2004.

[3] S. Consolvo and J. Towle. Evaluating an ambient display for the home. CHI EA '05, pages 1304–1307, New York, NY, USA, 2005. ACM.

[4] W. R. Hazlewood, K. Connelly, K. Makice, and Y.-k. Lim. Exploring evaluation methods for ambient information systems. CHI EA '08, pages 2973–2978, New York, NY, USA, 2008. ACM.

[5] W. R. Hazlewood, E. Stolterman, and K. Connelly. Issues in evaluating ambient displays in the wild: Two case studies. 2011.

[6] L. E. Holmquist. Evaluating the comprehension of ambient displays. CHI EA '04, pages 1545–1545, New York, NY, USA, 2004. ACM.

[7] J. F. Kelley. An empirical methodology for writing user-friendly natural language computer applications. CHI '83, pages 193–196, New York, NY, USA, 1983. ACM.

[8] J. Mankoff, A. K. Dey, G. Hsieh, J. Kientz, S. Lederer, and M. Ames. Heuristic evaluation of ambient displays. CHI '03, pages 169–176, New York, NY, USA, 2003. ACM.

[9] J. Nielsen and R. Molich. Heuristic evaluation of user interfaces. CHI '90, pages 249–256, New York, NY, USA, 1990. ACM.

[10] Z. Pousman and J. Stasko. A taxonomy of ambient information systems: four patterns of design. AVI '06, pages 67–74, New York, NY, USA, 2006. ACM.

[11] J. T. Stasko, D. McColgin, T. Miller, C. M. Plaue, and Z. L. Pousman. Evaluating the infocanvas peripheral awareness system: A longitudinal, in situ study. 2005.