I’m Home

Smartphone-enabled Gestural Interaction with Multi-Modal Smart-Home Systems

Diploma Thesis
interactive media.virtual environments (im.ve)
Prof. Dr.-Ing. Steffi Beckhaus
Faculty of Mathematics, Informatics and Natural Sciences
University of Hamburg

submitted by
Tilo Westermann

Supervisor:  Prof. Dr.-Ing. Steffi Beckhaus (Universität Hamburg),
Prof. Dr.-Ing. Sebastian Möller (Deutsche Telekom Laboratories, TU Berlin)

Tilo Westermann
Registration number: 5604919
Eitelstraße 54
10317 Berlin
Abstract

Smart-home environments are a highly active research topic in the field of Ubiquitous Computing. Through technological progress, more and more devices are equipped with networking capabilities and intelligent sensors that allow a merger in a home network. However, the control options for these devices are still the same as in the early 1980s. To curb the rising number of remote controls, so-called universal remote controls were introduced. The user-friendliness, however, remains questionable due to the large number of buttons. Therefore, new approaches for controlling smart-home environments are needed, not only because of the need for usability, but also for reducing the amount of required controls.

This thesis investigates the use of a smartphone-based gestural interface for smart-home control to encounter the emerging problem of an abundance of remote controls that are not necessarily user-friendly, or are replaced through a universal remote control. This approach seems worthwhile, as the mobile phone has become a pervasive part of everyday live. A series of studies for the creation of a set of applicable gestures is introduced and details on the implementation of a gesture recognition system for mobile devices are presented. As an example platform, the Apple iPhone was chosen for evaluating the implementation in a final usability study.

The results from this study show that gestures may be used for controlling devices in a smart-home environment.
Acknowledgments

I would like to express my gratitude to Prof. Dr.-Ing. Steffi Beckhaus and to Prof. Dr.-Ing. Sebastian Möller for their supervision and advice throughout this thesis. It is a pleasure to thank Christine Kühnel, Fabian Hemmert, Sven Kratz and Alexander Müller for constantly supporting me with patience and knowledge. It was a blessing to work within such a friendly and cheerful group of highly engaged researchers. A very special thanks goes to Aubrey-Derrick Schmidt, without whom I would not have learned of this opportunity. He accompanied me from contact linking, over relocating to Berlin to standing by my side with advice during the entire time. Many thanks to Dr. Benjamin Weiss, who has supported me in the statistical analysis of the results of the usability study. Thanks to Sarah Schipper, who took the time to realize the visualization of gestures by taking long exposure photographs. Further thanks goes to Joshua Marr and Christoph Szepp for proofreading this thesis.

I also want to express my appreciation to Daniela for her encouragement, understanding, and patience even during time-consuming times. Finally, I thank my mother for everything. Without her, all this would not have been possible. Thank you for your support in all areas of life.
Contents

4.2.2 Procedure ................................................. 20
4.2.3 Results ................................................. 21
4.2.4 Conclusion ............................................... 25

4.3 Clarity of Gesture Vocabulary .................................. 26
4.4 Memorability of Gestures ..................................... 28
4.5 Type of Feedback ............................................ 29

5 Implementation ................................................... 31
5.1 Choosing an Appropriate Device ............................. 31
5.2 Design ......................................................... 33
5.3 Implementation Details ....................................... 35
5.3.1 Developing for the Device ................................ 37
5.3.2 Recognizing Gestures .................................... 37
5.3.3 Identification of Problematic Areas and Their Solutions 41
5.4 Concurrent Validation ......................................... 46

6 Evaluation .......................................................... 49
6.1 Goal ........................................................... 49
6.2 Hypothesis ..................................................... 49
6.3 Test Design .................................................... 50
6.3.1 Task Design .............................................. 51
6.3.2 Questionnaire ............................................ 52
6.3.3 Participants .............................................. 53
6.4 Results ........................................................ 53
6.4.1 Evaluation of the Gesture Recognition Rate ............... 53
6.4.2 Analysis of Survey Data ................................ 53
6.4.3 Correlation between Recognition Rate and Ratings ....... 55
6.5 Discussion ..................................................... 55

7 Conclusions .......................................................... 59
7.1 Summary ..................................................... 59
7.2 Results ......................................................... 60
7.3 Future Work ................................................... 60

Bibliography .......................................................... 62

Appendices ............................................................. 69
 List of Abbreviations ............................................. 71
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Accelerometer axes of an Apple iPhone</td>
<td>6</td>
</tr>
<tr>
<td>2.2</td>
<td>Camera-based gesture recognition using a multi-colored glove</td>
<td>7</td>
</tr>
<tr>
<td>2.3</td>
<td>CyberGlove II wireless glove</td>
<td>7</td>
</tr>
<tr>
<td>2.4</td>
<td>Electromagnetic tracking system</td>
<td>8</td>
</tr>
<tr>
<td>2.5</td>
<td>A user performing a gesture on an interactive table</td>
<td>9</td>
</tr>
<tr>
<td>2.6</td>
<td>BioSemi FLAT electrodes sticked on the skin</td>
<td>9</td>
</tr>
<tr>
<td>4.1</td>
<td>Agreement for each referent in descending order</td>
<td>23</td>
</tr>
<tr>
<td>4.2</td>
<td>Questionnaire: Assigning referents to gestures</td>
<td>27</td>
</tr>
<tr>
<td>4.3</td>
<td>Illustration of steps 1 and 2 of the memorability study</td>
<td>28</td>
</tr>
<tr>
<td>4.4</td>
<td>Results of the survey on vibration feedback</td>
<td>30</td>
</tr>
<tr>
<td>5.1</td>
<td>Smartphone market share of Apple’s iPhone</td>
<td>32</td>
</tr>
<tr>
<td>5.2</td>
<td>iPhone audience composition - first quarter of 2009</td>
<td>33</td>
</tr>
<tr>
<td>5.3</td>
<td>A use case diagram presenting a graphical overview of the system functionality</td>
<td>34</td>
</tr>
<tr>
<td>5.4</td>
<td>Stages of Graphical User Interface (GUI) design</td>
<td>35</td>
</tr>
<tr>
<td>5.5</td>
<td>Process of adding a new gesture</td>
<td>36</td>
</tr>
<tr>
<td>5.6</td>
<td>State diagram giving an abstract description of the behavior of the system when executing a gesture command</td>
<td>39</td>
</tr>
<tr>
<td>5.7</td>
<td>Example of four different resolutions evaluated during a complete run of FastDTW</td>
<td>44</td>
</tr>
<tr>
<td>5.8</td>
<td>Recognition rates for seven gestures, 50 tries per gesture, single user</td>
<td>47</td>
</tr>
<tr>
<td>5.9</td>
<td>Used gestures, represented by long exposure photos</td>
<td>48</td>
</tr>
<tr>
<td>6.1</td>
<td>Sketch of the experimental room</td>
<td>50</td>
</tr>
<tr>
<td>6.2</td>
<td>Recognition rate of gestures during the usability test</td>
<td>54</td>
</tr>
<tr>
<td>6.3</td>
<td>Correlation between recognition rate and ratings</td>
<td>56</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Referents and their complexity</td>
<td>21</td>
</tr>
<tr>
<td>4.2</td>
<td>Gesture nature and mean and standard deviation of conceptual complexity ($C$), planning time ($t_p$), articulation time ($t_a$) and number of matches with other participants ($N$)</td>
<td>22</td>
</tr>
<tr>
<td>4.3</td>
<td>Comparison of experts and non-experts concerning gesture suitability and articulation time</td>
<td>25</td>
</tr>
<tr>
<td>4.4</td>
<td>Gesture-referent mapping</td>
<td>26</td>
</tr>
<tr>
<td>5.1</td>
<td>Technical specifications for <em>iPhone 3G</em> and <em>3GS</em></td>
<td>45</td>
</tr>
<tr>
<td>5.2</td>
<td>Critical values for Grubb’s test for outliers</td>
<td>46</td>
</tr>
<tr>
<td>6.1</td>
<td>Differences in the frequency of use of modalities</td>
<td>55</td>
</tr>
</tbody>
</table>
1 Introduction

In this chapter, the motivation for this thesis, as well as the contribution to existing research is presented. Finally, the structure of this work is outlined.

1.1 Motivation

Smart-home environments are a highly active research topic in the field of Ubiquitous Computing. Due to technological advancement, home environments are equipped with increasingly more technology. To overcome the possible narrowness of today’s home device controls, novel interfaces could be developed to bridge the gap between the human user and the digital world. Current smart-home environments, such as the Infotainment management with SPEech interaction via REMote-microphones and telephone interfaces (INSPIRE) system at Deutsche Telekom Laboratories1, TU Berlin2, are still mostly an object of research. They are designed to facilitate user interaction with various household devices by using spoken input as a single interface for multiple devices.

As a second topic in the field of Ubiquitous Computing, Mobile (Human-Computer-) Interaction is of high activity. According to a forecast by Stordahl et al. (2005), over 90% of the population in Western Europe will use a mobile phone by the year 2010. As with home environments, there is a significant progress in mobile phone technology, enabling further usage possibilities beyond phone calls and short messaging. Therefore, moving mobile devices into the context of stationary home environments seems worthwhile since mobile devices have become a pervasive part of everyday life. Existing research in this area has shown promising results, such as controlling multimedia players using Near Field Communication (NFC)-enabled mobile phones (Sánchez et al., 2007). Current mobile phones are equipped with a multitude of sensors such as touch screens supporting multi-touch, accelerometers, magnetometers or even Infrared (IR) distance sensors and are capable of connecting to other devices via Infrared (IR), Bluetooth or Wireless Local Area Network (WLAN).

1 http://www.laboratories.telekom.com
2 http://www.qu.tu-berlin.de
Introduction

Since speech recognition as a common input modality is subject to a high error rate, its advantages (e.g. hands-free device control) do not outnumber those of traditional controls.

Besides voice, gestures seem to be a natural way of interaction for human beings. Therefore, adding a gesture interface renders the home environment even more natural in its style, potentially gaining a plus in flexibility and robustness.

1.2 Contribution

This thesis investigates the topic of gestural interaction in multi-modal smart-home environments. In this topic, it appears important to clarify the question to which extent gestures are adequate to the task of controlling home appliances and whether gestural interaction can keep up with alternate modalities such as the conventional remote control or voice input.

This thesis contributes to the existing body of knowledge in this area, as it sheds light on how gestures as an input modality may be brought into the context of smart-home environments. In order to be able to examine the applicability of gestural interaction, a method for the development of gestures is introduced in one of the first steps. Later, the resulting gestures as well as the implemented gesture recognition algorithm are evaluated in a usability study with 27 participants and compared to speech as an alternative input modality for home appliances.

1.3 Outline of the Thesis

In this chapter, the topic of this thesis was introduced and the contribution to existing research was presented. The following chapter provides the necessary background on which this work is based on. It surveys related work on gesture recognition and fields of application using gesture recognition. Finally, it provides a brief overview of the use of gestures and mobile devices in smart-home environments and leads to the problem analysis. Chapter outlines the problem of an abundance of remote controls for home appliances that are not necessarily user-friendly, or are replaced with a universal remote control. The subsequent approach addresses this problem and identifies key criteria for an alternative input modality. In Chapter preliminary studies are described. Theses studies were carried out to illuminate issues that arose during the conceptual design. In addition, a set of gestures for controlling appliances in a smart-home environment was created by involving users in the design process. Chapter provides details on the
implementation phase. Besides the choice of the device, it highlights the design process and gives an insight into the gesture recognition algorithm. Finally, a verification test is run to prove operability of the developed system. The evaluation of the implemented approach is described in Chapter 6. This chapter deals with a usability test to evaluate the previous work and provides details on the test design and execution. Subsequent, the results of this study are presented and discussed. Finally, Chapter 7 is devoted to the conclusions that can be drawn from the results of the previous chapters. It discusses stated goals of this work and provides an outlook on future work.
2 Related Work

In order to understand the upcoming issues, it is important to reflect on the following: status quo of gesture recognition techniques and their fields of application. Gesture recognition is the interpretation of a human gesture, typically originating from the hand or face (mimic; facial expression). Following the trend of designing more natural ways of interacting with computer systems, this is a highly active topic in computer science and other disciplines as well.

In the field of Human Computer Interaction (HCI), many attempts were made to recognize gestures automatically. Determining performed gestures may be achieved by various techniques, which will be introduced in this section. Also, their suitability in the space of smart-home environments is examined.

2.1 Gesture Recognition Techniques

Gesture recognition techniques are widely diversified in consequence of their broad range of application. This subsection gives an overview of common methods and sensors employed. The presented techniques differ in terms of precision in which they can recognize gestures. In the following this will be referred to as granularity.

2.1.1 Accelerometers

An accelerometer measures changes in velocity along a linear axis. Devices using accelerometers are most often equipped with a Micro Electro-Mechanical Systems (MEMS)-based 3D-accelerometer to measure changes along each of the primary axes in three-dimensional space; that is, right/left (X), up/down (Y), and front/back (Z). Figure 2.1 shows the accelerometer axes of an Apple iPhone in a right-handed coordinate system.

Approaches to recognizing gestures using accelerometers reach from utilizing Nintendo’s Wii Remote [Schlömer et al., 2008] [Kratz et al., 2007] over employing Electromyography (EMG) in combination with accelerometers [Zhang et al., 2009].
Related Work

Figure 2.1: Accelerometer axes of an Apple iPhone
(Source: http://developer.apple.com)

to build custom devices (Park et al., 2008b; Oakley and Park, 2007). Current mobile phones are increasingly equipped with motion sensors, which makes this technique particular interesting for this work.

The typical granularity allows for recognition of palm gestures, as the user usually holds a mobile device in his hand. Evaluation of these approaches concerned mostly technical performance, e.g. Schlömer et al. (2008) achieved an average recognition rate of 90% for five gestures.

2.1.2 Computer Vision

The use of cameras provides the user with a natural interface, since he does not need additional equipment. However, he needs to stay in the viewport of the camera(s), thus limiting the range of application to a stationary context.

Among others, Richarz et al. (2008) contributed to research in this area by using deictic gestures in a smart-home environment. Lee and Kim (1999) provided a speaker with increased expressive power by using gesture recognition to control a PowerPoint presentation.

Besides bare-handed tracking, a couple of other methods utilizing cameras have been subject to research. Wang and Popović (2009) make use of a glove with a color pattern (see Figure 2.2) resulting in a simple, inexpensive tracking system. Based on the data glove metaphor, Pamplona et al. (2008) built an image-based data glove. A camera attached to the user’s hand tracks visual markers at the finger tips; using inverse kinematics, the position of each finger joint is estimated, enabling the recreation of finger
2.1 Gesture Recognition Techniques

Figure 2.2: Camera-based gesture recognition using a multi-colored glove (Wang and Popovic, 2009)

movement in the virtual world. Alongside traditional cameras, cameras with infrared-pass filters are used in combination with IR Light-Emitting Diodes (LED) to achieve reflection of e.g. the user’s hands (Starner et al., 2000).

The granularity of camera-based gesture recognition ranges from the hand as a whole (as seen in Lee and Kim (1999)) up to individual fingers (as seen in Pamplona et al. (2008); Wang and Popovic (2009)). The recognition rates are also promising, for example, Richarz et al. (2008) reported a recognition rate of 82.5%. A usability analysis of the image-based data glove (Pamplona et al., 2008) showed, however, that the system provides only limited comfort for the user.

2.1.3 Glove devices

Figure 2.3: CyberGlove II wireless glove (Source: http://www.vrealities.com/cyber.html)

Glove-based gestural interfaces require the user to wear a electronically-equipped glove, which is either connected to a computer via cables or attached to a wireless unit, as seen in Figure 2.3. Applications using data gloves (Zimmerman et al., 1987; Park et al., 2008a) may take advantage of inverse kinematics to recognize sign language and gestures. Though such gloves are accurate in reporting hand states, it is still cumbersome to put on the glove; using wired gloves (as shown by the example of Baudel and
Related Work

Beaudouin-lafon (1993)), the user is potentially limited in his freedom of movement. One advantage over camera-based tracking is that glove technology does not rely on line-of-sight observation.

Using inverse kinematics, it is possible to scan every single finger movement, which allows for the recognition of very detailed gestures. A user study by Park et al. (2008a) produced a recognition rate of 94% for 17 gestures.

2.1.4 Electromagnetic Trackers

![Figure 2.4: Setup of an electromagnetic tracking system for gesture recognition (Craven et al., 2000)](image)

Electromagnetic tracking devices measure magnetic fields generated by sending current through units consisting of three wire coils (which are oriented perpendicular to one another). These units can be attached to the object of interest, e.g. the user’s hands (Craven et al., 2000), as shown in Figure 2.4. Sequential activation of each of the wires enables the determination of 3-space positions and orientation of the sensor.

Besides the limited range (up to 4.6 meters for a Polhemus Liberty¹), the main disadvantage of such systems is the sensitivity to metallic objects (e.g. office furniture) that can disturb the magnetic field. Under laboratory conditions, Craven et al. (2000) reported a recognition rate between 82-96%.

2.1.5 Interactive Surfaces

Limiting the space to two dimensions, surfaces may be used for pen or finger-based stroke gesture control. These are mainly touch-sensitive surfaces (Dietz and Leigh

¹http://www.polhemus.com/?page=Motion_Liberty
2.1 Gesture Recognition Techniques

Figure 2.5: A user performing a gesture on an interactive table (Wobbrock et al., 2009)

Wobbrock et al. (2009) or translucent surfaces (Lorenz et al., 2009), where a camera scans the contact points, as shown in Figure 2.5. Wobbrock et al. (2009) also presented a very promising approach to designing tabletop gestures that will be revisited later in this work.

2.1.6 Electromyography (EMG)

Figure 2.6: BioSemi FLAT electrodes stucked on the skin (Source: http://www.biosemi.com/active_electrode.htm)

Sensing muscle activity with EMG is frequently used in clinical settings, but also more recently the use of sensing the surface of the skin for HCI applications is examined. Existing research in the area of gesture recognition using EMG has shown promising results, as Saponas et al. (2009) used forearm muscle-sensing (as shown in
Related Work

Figure 2.6) to develop a gesture set to enable interaction, yielding a recognition rate between 79-88%.

2.2 Fields of Application Using Gesture Recognition

*Everything is best for something and worst for something else.*

---

WILLIAM A. S. BUXTON

The use of gesture recognition is explored and used in many areas. This section provides a survey of application ranges in a mobile and stationary context.

2.2.1 Mobile Systems

The use of gesture recognition in mobile systems is not confined to a specific setting and the user is not restricted in his movements. The examples shown in this section are categorized according to their type of communication.

**Human-to-Human**

The ability to use web-based services on mobile devices supports the sustained growth of social networks. Thus, for example, Scheible et al. (2008) used mobile devices equipped with accelerometers to share multimedia-art by ‘throwing’ captured pictures.

**Human-to-Computer**

Current smartphones are increasingly equipped with touch-sensitive displays, replacing the traditional keypad and expanding the possibilities for interaction, including surface gestures. Besides these already commercially available products, other techniques are subject to research. Applications range from basic menu-driven interaction using accelerometers (Oakley and Park, 2007), over recognizing Arabic numerals (Zhang et al., 2008) to text entry using a joystick on the back or front of a mobile phone (Wobbrock et al., 2007). Saponas et al. (2009) examined the use of [EMG]-based gesture interaction to control a portable music player.
2.2 Fields of Application Using Gesture Recognition

**Computer-to-Human**

Interacting with robots in a non-conventional manner could be useful, for instance, in supporting disabled people in specific tasks. Bonato et al. (2004) embedded cameras in a mobile robot and used gesture recognition for control.

**2.2.2 Stationary Systems**

Stationary systems offer a wider technical range of options for gesture recognition, often restricting the user to some extent in his freedom of movement. This section presents areas of research in a stationary context.

**Collaborative Work**

Today, collaborative work is considered very important. Therefore, it is easy to comprehend that researchers are trying to push forward methods facilitating cooperation. Existing research in this area has shown that tabletop interaction, particularly on multi-touch surfaces, may be an appropriate means. Examples were provided by Dietz and Leigh (2001); Wu and Balakrishnan (2003); Tse et al. (2006).

**Interaction with Home Environments**

Another emerging area that makes use of gesture recognition is the remote control of various devices. For example, Cheng and Pulo (2003) are aiming for interaction with display systems by means of IR laser tracking devices. Wilson and Shafer (2003) and Park et al. (2008b) took a different path by utilizing accelerometers to interact with intelligent environments.

**Alternative Computer Interfaces**

Gestures may also be used as a substitute for interaction via mouse and keyboard under certain circumstances. Farella et al. (2007) used accelerometer-based gesture recognition for navigation in virtual worlds. Schlömer et al. (2008) also used accelerometers for photo browsing, among other applications.

**Manipulation of Virtual Objects**

Research shows that glove-based input is particular well suited for applications that aim to manipulate virtual objects, especially when employing tactile feedback to create an
Related Work

immersive experience. Examples based on camera tracking were provided by Pamplona et al. (2008) using an image-based data glove and Wang and Popovic (2009) using a glove with a color pattern. Earlier research using wired gloves has shown promising results, as of Zimmerman et al. (1987); Baudel and Beaudouin-lafon (1993); Kavakli et al. (2007).

Immersive Gaming

In the field of computer games, researchers are trying to make the player’s experience more interactive and immersive. Kang et al. (2004) applied camera-based tracking to create an interface between a video game and its user. Keir et al. (2006) and Kratz et al. (2007) are using 3D accelerometers for game play input; Zhang et al. (2009) go even further by adding EMG sensors.

Pedagogical and Medical Field

Gesture recognition also finds use in the pedagogical and medical field, too. Considering new interaction modalities potentially serves disabled people better than traditional modalities. Thus, for example, Starner et al. (2000) employed camera-based tracking for home automation control and medical monitoring. Craven et al. (2000) addressed this topic by using an electro-magnetic tracking system. Amft et al. (2005) examined the use of body-worn inertial sensors, namely accelerometers and gyroscopes, to detect eating and drinking gestures. These are intended for automatic dietary monitoring in the domain of behavioral medicine.

An educational application was presented by Bevilacqua et al. (2007), using accelerometer-based gesture recognition for music pedagogy.

2.3 Use of Gestures and Mobile Devices in Smart-Home Environments

In addition to gestural interfaces that were briefly addressed in the previous section, there are approaches using mobile devices to control home environments - some of them using motion sensors. This section presents some approaches in this area, as they are of particular interest for this work.

Koskela and Väänänenen-Vainio-Mattila (2004) conducted an ethnographic study of smart-home usability and evaluated three User Interfaces (UIs), including a mobile
2.3 Use of Gestures and Mobile Devices in Smart-Home Environments

phone that was capable of controlling home appliances via a GUI. One of their findings is that “the mobile phone could become the primary centralized remote control”. Myers (2005) followed a similar approach, using a handheld for remote control of personal computers and appliances. Roduner et al. (2007) examined strengths and limits of a mobile phone acting as a universal interaction device. Like Koskela and Väänänen-Vainio-Mattila (2004), they used a GUI to select options. Their user study that compared a universal appliance controller to traditional, physical appliance interfaces showed that participants were faster at solving exceptional tasks, but slower when performing everyday tasks. As already noted in the previous section, Park et al. (2008b) presented a promising approach using a handheld, equipped with a 3-axes accelerometer and an IR module, to control home appliances. With a limited number of gestures they achieved a recognition rate between 74.3% and 91.5%.

There are also systems allowing for multimodal interaction using mobile devices and speech. Johnston et al. (2007) suggested a system that allowed users to access movies using speech, pen, remote control and dynamic combinations of these modalities. They conducted a user study to compare the modalities and found that 55% of the participating users preferred GUI based input over speech and handwriting. Gieselmann and Denecke (2003) combined speech and pointing gestures, recorded by a stereo camera, to control devices in an intelligent room. Control is done by saying a command and simultaneously pointing at a device.
3 Problem Analysis and Approach

The scientific work presented in the previous chapter offers promising approaches, however, there still remains room for further work. This chapter analyzes the identified problem and offers an approach to solving it.

3.1 Problem Analysis

Home environments are increasingly populated with sophisticated technical devices. However, interaction techniques, mostly in the form of remote controls, remain the same as in the early 1980s. While the traditional remote control may be an appropriate means for simple tasks on a single device, controlling multiple devices at a time tends to be cumbersome. Current electronic devices for the use in living rooms are mostly shipped with an extra remote control, cluttering up the coffee table. An existing approach addressing this problem is combining multiple remote controls to one ‘universal’ remote control. Although eliminating one issue, these devices can be equipped with as many as 60 buttons\(^1\), limiting the user-friendliness.

Usability of devices in home environments is of crucial importance. Research in recent years has shown promising basic approaches in designing ‘smart’ appliances, yet not bridging the gap between the human user and the digital world. Speech and gestures seem to be natural ways of expressing oneself for human beings.

Numerous approaches examined the use of voice-based interfaces for home automation, facing not only technical obstacles. The need to interrupt a conversation to perform system commands might constitute a social barrier. On the technical side, noise can have a disturbing effect on recognition systems, possibly resulting in high error rates. A possible approach to avoid unintended commands might be a physical switch to toggle the recognition system (‘push-to-talk’), or adding a keyword to activate recognition, which could, on the other hand, have a deterrent effect on the user.

The use of gesture recognition was also investigated in this area. Starner et al. (2000) designed a pendant that the user wears, which was equipped with a camera, tracking

\(^1\)e.g. Sony’s Universal Remote
the user’s hands. Lighting in living rooms tends to be less than optimal for camera tracking, possibly limiting the practicability of camera-based approaches under real life conditions. As highlighted in the previous chapter, research investigated the use of other methods. It seems worthwhile to investigate the combination of mobile devices and home control, as both seem to converge and possibly do so even more in the future. Kela et al. (2005) utilized a SoapBox (Sensing, Operating and Activating Peripheral Box): a small box equipped with a wireless unit and a three-axis acceleration sensor to control a design environment. Park et al. (2008b) employed a customized Personal Digital Assistant (PDA) to control home appliances using gestures. Gestures were recognized using accelerometers, system commands were sent through IR. This approach sounds promising but should be explored in further depth.

In summary, one can say that so far there is no adequate substitute for the usual remote control. Either the solutions presented are too specific for everyday use, or they limit the user in his freedom. More often than not, systems were evaluated concerning technical performance, e.g. recognition rate. The evaluation of subjective experiences was rarely the case. Although technical performance is an important factor, the inclusion of users in the design process may lead to better results. This problem - the missing user-friendly interaction method for home environments - brings up the following task: seeking a more natural way of interaction that seamlessly integrates into the user’s everyday life.

3.2 Approach

As mentioned above, gestures seem to be a natural communication medium for human beings. While this is likely to be true for mimics and full body gestures as subcultural signs in an interpersonal dialogue, this does not include gestures for device control. Gestures of any kind - as well as language - have to be learned in the course of development and the vocabulary of gestures can later be extended. Nevertheless, the transfer of gestures into a new domain means merely an adjustment or enlargement of the gesture vocabulary. For this reason, this work examines the use of gestural input in home environments that seamlessly integrates into the user’s everyday life. Lorenz et al. (2009) investigating the usage of a PDA for controlling a remote cursor for text input on displays in home environments, clarified the required attributes for remote interaction devices as followed:

- The device is not associated with a fixed location.
3.2 Approach

- The device is equipped with communication capabilities for local-area networking (Bluetooth, IR or WLAN).
- The device offers powerful components available for user-device interaction
- The device is popular to customers and its operating is quite common to a large community

Those requirements that are consistent with those in our area of application, limit the presented approaches to accelerometer-based gesture recognition and the usage of handheld devices, e.g. PDAs or smartphones. The device and its functions should also be immediately accessible, which may not always be the case with conventional remote controls.

We assume that gesture control may not be well suited for all possible tasks and that the functionality of a common remote control is quite fitting in some cases. With these considerations it seems reasonable to use a commercially available smartphone with built-in acceleration sensors. On one hand, one could derive a benefit from both technologies (motion sensors and smartphone features), on the other hand it is an always accessible pocket-carried device.
4 Preliminary Studies

This chapter describes five studies that were conducted prior to or simultaneously with the actual implementation. The first survey polled for participants’ opinions on gesture-controlled home appliances. The following three studies are concerned with the design and evaluation of a gesture set for smart-home control. The final survey was targeted to the question of whether a vibration feedback of a cell phone induces negative or positive associations.

4.1 Initial Survey

To obtain a first impression of whether or not a gesture-based control modality has any appeal to possible end users, and which home appliances people would like to control, a first study was conducted, which is described below.

We interviewed 17 students from different disciplines, aged between 18 and 28 \( (Mean = 24.82, SD = 2.46) \). Among them were eight women and nine men. Twelve participants had experience with the iPhone or Wii, that is to say, they owned one of the devices or at least used it regularly. Eleven of the 17 participants could well imagine to control home appliances via gestures. Interestingly, the remaining six participants, apart from one, were those who seldom or never used any of both devices. While four of them could hardly imagine using a gesture control, two participants rejected this idea completely.

When asked about equipment, 10 participants named the TV, lamps were mentioned 8 times, hi-fi equipment 6 times, and blinds 3 times. Corresponding functions were turning the device on and off, adjusting sound volume and lighting conditions, changing the TV channel or lowering/raising the blinds.

The final question targeted on whether there should be a pre-defined set of gestures. Interestingly, six participants wanted to have a predetermined set that they can extend or modify according to their wishes. Four preferred a static gesture set, while three would like to define their own gestures; the remaining four participants gave no answer.
Preliminary Studies

In summary, the results indicate that users are generally interested in controlling home appliances via gestures. Users with prior experience with motion-controlled devices were more open-minded towards this alternative input modality. Since we are experiencing an uprise in such interaction techniques, users may show growing interest in gesture-based interfaces in the near future. Our findings suggest that users might prefer a pre-defined set of gestures for less complex tasks, while they would like to define own gestures for more complex tasks, for which no gesture is immediately apparent.

4.2 Finding a Gesture Vocabulary

Following the promising approach of Wobbrock et al. (2009), the first step considers the search for a gesture vocabulary. Due to the different domain and conditions, adjustments were made where necessary.

4.2.1 Participants

Eighteen participants volunteered for the study. Among them were six women and twelve men with an average age of 26.5 (SD = 4.99). Two of them were left-handed. Four participants owned an Apple iPhone or iPod, five regularly played with the Nintendo Wii. The majority of participants were recruited from the research facility with backgrounds in communication studies, computer science, design research or psychology.

4.2.2 Procedure

The study was conducted in a fully furnished living room (see Figure 6.1). Participants were seated on a sofa, having all devices in the field of view. The experimenter triggered the functions of devices (e.g. drawing the blinds) via a GUI and described them afterwards (“draw the blinds”). The participants were then asked to come up with an appropriate gesture for this function, while thinking aloud. A total of 23 functions – referred to as ‘referents’ in the following – were presented in a predefined order and divided into seven groups (devices).

During the experiment, all movements were recorded on video and then annotated to calculate planning and calculation time of each gesture. After each task, the participant evaluated the newly elected gesture using two 7-point Likert scales (-3 = strongly disagree, 0 = undecided, +3 = strongly agree). The first asked for the suitability of the
4.2 Finding a Gesture Vocabulary

### Table 4.1: Referents and their complexity (1 = simple, 5 = complex)

<table>
<thead>
<tr>
<th>Referents by device</th>
<th>Complexity by device</th>
<th>Complexity by device</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Blinds</strong></td>
<td></td>
<td><strong>Electronic Program Guide (EPG)</strong></td>
</tr>
<tr>
<td>DOWN</td>
<td>1.00 0.00</td>
<td>SHOW</td>
</tr>
<tr>
<td>UP</td>
<td>1.00 0.00</td>
<td>RECORD MOVIE</td>
</tr>
<tr>
<td>STOP</td>
<td>2.50 0.50</td>
<td>SET REMINDER</td>
</tr>
<tr>
<td><strong>Lamps</strong></td>
<td></td>
<td><strong>Video Recorder</strong></td>
</tr>
<tr>
<td>TURN ON</td>
<td>1.25 0.43</td>
<td>SHOW LIST</td>
</tr>
<tr>
<td>DIM</td>
<td>1.25 0.50</td>
<td>DELETE MOVIE</td>
</tr>
<tr>
<td>TURN OFF</td>
<td>2.50 0.50</td>
<td>PLAY MOVIE</td>
</tr>
<tr>
<td><strong>TV</strong></td>
<td></td>
<td><strong>Answering Machine</strong></td>
</tr>
<tr>
<td>TURN ON</td>
<td>1.50 0.87</td>
<td>PLAY MESSAGE</td>
</tr>
<tr>
<td>NEXT CHANNEL</td>
<td>1.25 0.43</td>
<td>NEXT MESSAGE</td>
</tr>
<tr>
<td>PREVIOUS CHANNEL</td>
<td>1.25 0.43</td>
<td>PREVIOUS MESSAGE</td>
</tr>
<tr>
<td>VOLUME DOWN</td>
<td>2.50 0.50</td>
<td></td>
</tr>
<tr>
<td>VOLUME UP</td>
<td>1.00 0.00</td>
<td></td>
</tr>
<tr>
<td>TURN OFF</td>
<td>1.50 0.87</td>
<td></td>
</tr>
<tr>
<td><strong>General</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HELP</td>
<td>2.25 1.09</td>
<td></td>
</tr>
<tr>
<td>ABORT</td>
<td>2.00 0.00</td>
<td></td>
</tr>
</tbody>
</table>

The study produced a number of 414 recorded gestures (18 participants, each with 23 proposed gestures), of which 174 were of different types. The complexity of each referent was rated by four experts, as shown in Table 4.1. This rating yields an average complexity of 1.97 \((SD = 1.03)\).

### Classification of Gestures

Wobbrock et al. (2009) proposed a taxonomy of gestures that classifies the gestures along four dimensions: **form**, **nature**, **binding** and **flow**. While Wobbrock et al. (2009) studied two-dimensional surface gestures for desktop and table top interaction, this approach makes use of a smartphone as an input device for smart-home interaction. In the following, the four dimensions will be applied to this area of application using wizard-style ‘wand’ gestures.

The **form** dimension distinguishes between ‘static’ and ‘dynamic’ ‘pose’ and ‘path’ of one hand. In this case, ‘static pose’ can be interpreted as the orientation of the device, whereas a ‘dynamic pose’ is a transition from one orientation to another. A movement of the device while maintaining the orientation is referred to as ‘static path’, a simultaneous change in orientation is classified as a ‘dynamic path’. Most of the performed gestures
Preliminary Studies

<table>
<thead>
<tr>
<th>Nature</th>
<th>Occurrences</th>
<th>C (Mean, SD)</th>
<th>t_p (Mean, SD)</th>
<th>t_a (Mean, SD)</th>
<th>N (Mean, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical</td>
<td>34 (8.2%)</td>
<td>1.22 (0.54)</td>
<td>2.38 (3.74)</td>
<td>0.89 (0.27)</td>
<td>11.71 (5.05)</td>
</tr>
<tr>
<td>Metaphorical</td>
<td>221 (53.4%)</td>
<td>1.55 (0.70)</td>
<td>6.00 (9.93)</td>
<td>0.97 (1.23)</td>
<td>5.68 (4.30)</td>
</tr>
<tr>
<td>Abstract</td>
<td>115 (27.8%)</td>
<td>2.63 (1.04)</td>
<td>16.21 (19.53)</td>
<td>1.56 (1.04)</td>
<td>1.02 (1.89)</td>
</tr>
<tr>
<td>Symbolic</td>
<td>44 (10.6%)</td>
<td>2.93 (0.90)</td>
<td>13.43 (14.18)</td>
<td>1.50 (0.64)</td>
<td>2.41 (2.70)</td>
</tr>
<tr>
<td>Total</td>
<td>414 (100%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.2: Gesture nature and mean and standard deviation of conceptual complexity (C), planning time (t_p), articulation time (t_a) and number of matches with other participants (N)

(310 of 414 = 74.9%) can be described as ‘static path’. Only one repeatedly appearing gesture could be identified as a ‘static pose’: 58 (14%) times the participants pointed at a device with the smartphone in their hand and pressed an imaginary button. In 8.9% (=37) of cases, the gesture could be described as a ‘dynamic pose’, such as for referents like VOLUME UP and VOLUME DOWN (roll right/left). Even fewer gestures (9 of 414 = 2.2%) were categorized as ‘dynamic path’.

The nature dimension again classifies gestures into four categories: symbolic (e.g. writing a question mark into the air), physical (e.g. pulling down the blinds), metaphorical (e.g. tapping an imaginary button) or abstract (in case of arbitrary mappings). This dimension can be directly adapted from the taxonomy of [Wobbrock et al. (2009)](Wobbrock2009). Their finding that the conceptual complexity C_r of the referents, as rated by experts, significantly influenced the nature of gestures could be repeated (F(3, 410) = 73.41, p < .0001), with more complex referents resulting in symbolic gestures. Furthermore the nature of gestures has a significant influence on gesture planning time (F(3, 410) = 18.91, p < .0001), gesture articulation time (F(3, 187) = 4.60, p < .01) and number on matches with other participants (F(3, 410) = 89.08, p < .0001). Gestures classified as physical were found to have the highest agreement between participants and abstract gestures were preceded by the longest planning time (cf. Table 4.2).

The binding dimension describes, whether gestures only require information about the object they affect, or whether they refer to the world (environment). The gestures studied in this area are by definition acting on devices (or device-independent for the referents HELP and ABORT), thus a transfer of this dimension is not easy.

A gesture’s flow can be discrete or continuous. Discrete refers to gestures that are recognized as an event (e.g. a swing to the right as ‘next’), whereas flow is continuous if ongoing recognition is required (e.g. turning the device to dim the light). Among the
4.2 Finding a Gesture Vocabulary

performed gestures only a small proportion (11.8%) could be classified as continuous and they all related to the referents DIM LIGHT, VOLUME DOWN and VOLUME UP.

Analysis of Gesture Vocabulary

Apart from counting the number of matches with other participants for each gesture, we computed the agreement score $A$ for each referent. The agreement score, as adopted from Wobbrock et al. (2009), delivers the degree of consensus among participants:

$$A_r = \sum_{P_i \subseteq P_r} \left( \frac{|P_i|}{|P_r|} \right)^2$$

$A_r$ is the agreement score for one referent $r \in [1, \ldots, 23]$. $P_r$ represents the complete set of 18 proposed gestures for referent $r$, and $P_i$ is the subset of identical gestures from $P_r$. A compilation of all agreement scores can be found in Figure 4.1. Consistent with the findings of Wobbrock et al. (2009), the conceptual complexity of referents is negatively correlated with the agreement between participants with a Pearson’s $r = -0.67$, significant at the 0.001 level. That is to say, the more complex the referent, the fewer participants came up with the same gesture.

The analysis of the previously calculated planning and articulation time and user ratings confirms most of the findings of Wobbrock et al. (2009):
Preliminary Studies

- The average planning time for a gesture correlated significantly with the conceptual complexity of the referent ($r = .41, p < .0001$). In general, the more complex the referent, the longer it took participants to conceive a gesture. Simple referents (BLINDS DOWN/UP, VOLUME UP and PREVIOUS/NEXT MESSAGE) took on average 1.52 seconds ($SD = 4.12$). The referent rated most complex (SHOW EPG) took on average 26 seconds ($SD = 19.30$).

- The referents’ conceptual complexity did not correlate with gestural articulation time.

- The higher the number of matches of one gesture with gestures from other participants, the higher the suitability rating of this gesture ($F(4, 404) = 12.78, p < .0001$).

- Suitability ratings of the gestures correlated strongly with referents’ conceptual complexity ($r = -.923, p < .0001$). Gestures associated with complex referents were poorly rated ($Mean = -1.41, SD = 1.28$), while simple referents led to better ratings ($Mean = 1.64, SD = 1.63$).

- In contrast to Wobbrock et al. (2009), a small correlation of a referent’s conceptual complexity with average ratings of gesture ease was found ($r = -.174, p < .0001$). This finding indicates that more complex referents elicited gestures that were more difficult to perform.

- Planning time has a significant influence on the suitability rating ($F(251, 162) = 1.48, p < .01$), as well as on the gesture ease rating ($F(251, 162) = 1.51, p < .01$). As planning time increased, ratings decreased.

- Articulation time did not affect suitability ratings, but it did affect gesture ease ratings negatively ($F(60, 130) = 2.07, p < .001$). By contrast, Wobbrock et al. (2009) found that gestures with longer articulation time were rated better.

Since we are going to explicitly choose the device to control, the gestures do not need to be unique across referents. For this reason, the gesture vocabulary can be reduced to a smaller subset, as participants actually tended to reuse gestures.

Mental Model Observations

Dichotomous Referents, Reversible Gestures As previously found by Wobbrock et al. (2009), participants employed reversible gestures for most dichotomous
4.2 Finding a Gesture Vocabulary

<table>
<thead>
<tr>
<th></th>
<th>Experts</th>
<th></th>
<th>Non-Experts</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Suitability</td>
<td>1.48</td>
<td>0.44</td>
<td>0.51</td>
<td>0.86</td>
</tr>
<tr>
<td>Articulation time</td>
<td>0.86</td>
<td>0.62</td>
<td>1.51</td>
<td>1.35</td>
</tr>
</tbody>
</table>

Table 4.3: Comparison of experts and non-experts concerning gesture suitability and articulation time

Experts Non-Experts

Table 4.3: Comparison of experts and non-experts concerning gesture suitability and articulation time

Referents (BLINDS UP / DOWN, NEXT / PREVIOUS CHANNEL, VOLUME UP / DOWN, NEXT / PREVIOUS MESSAGE). Turning on/off a device (TV, light) most often led to identical gestures (e.g. 15 of 18 gestures for the referent TURN ON / OFF TV), possibly based on the metaphor of an on/off button. In 30 of 34 cases, a movement to the right meant NEXT (CHANNEL or MESSAGE) and in 15 of 18 cases, a movement downwards meant VOLUME DOWN. This seems to be influenced by the arrows of a keyboard rather than by a typical remote control: only two participants used an upwards movement to switch to the next channel and a movement to the right to raise the volume (as found on a remote control for TV).

**Gestural Form and Flow**

There was no obvious distinction between dynamic pose (movement of the mobile device) and static path (movement of arm) gestures. Participants chose gestures depending on their tendency to perform small or large movements. This is especially apparent for the dichotomous referents named above. In general, continuous gestures took longer to perform than discrete gestures. Apart from this, articulation time was influenced by the nature of the gesture and by participant’s disposition.

**Influence of Expertise**

About half of the participants had prior experience with motion-controlled devices (see 4.2.1) and were therefore considered as experts. We found that experts, compared to non-experts, were more confident in choosing an appropriate gesture and faster in performing the gesture (see Table 4.3).

4.2.4 Conclusion

As hypothesized by Wobbrock et al. (2009), our findings underline the presumption that the “wisdom of crowds” may prove very useful in designing a coherent set of gestures for HCI. Our study showed that a small group of experts is able to estimate the complexity of proposed referents. Since participants from the previous study showed an interest
Preliminary Studies

<table>
<thead>
<tr>
<th>Gesture</th>
<th>Description</th>
<th>Referents (number of correct mappings are indicated in brackets)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>up</td>
<td>swing up</td>
<td>BLINDS UP (11), VOLUME UP (8), BRIGHTEN LIGHT (7)</td>
</tr>
<tr>
<td>down</td>
<td>swing down</td>
<td>BLINDS DOWN (13), VOLUME DOWN (9), DIM LIGHT (8)</td>
</tr>
<tr>
<td>pull cord</td>
<td>move down and up (like a lamp cord)</td>
<td>TURN ON LAMP (8), TURN OFF LAMP (8)</td>
</tr>
<tr>
<td>Metaphorical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>right</td>
<td>swing to the right</td>
<td>NEXT CHANNEL (15), NEXT MESSAGE (14)</td>
</tr>
<tr>
<td>swipe</td>
<td>horizontal movement, directionless</td>
<td>STOP BLINDS</td>
</tr>
<tr>
<td>left</td>
<td>swing to the left</td>
<td>PREVIOUS CHANNEL (10), PREVIOUS MESSAGE (11)</td>
</tr>
<tr>
<td>circle</td>
<td>draw vertical clockwise circle</td>
<td>TURN ON LAMP (3), BRIGHTEN LIGHT (7)</td>
</tr>
<tr>
<td>anti-circle</td>
<td>draw vertical counter-clockwise circle</td>
<td>TURN OFF LAMP (3), DIM LIGHT (8)</td>
</tr>
<tr>
<td>remote control</td>
<td>pointing forward</td>
<td>TURN ON TV (10), TURN OFF TV (5), TURN ON LAMP (6), TURN OFF LAMP (3)</td>
</tr>
<tr>
<td>alarm clock</td>
<td>shake mobile phone quickly next to ear</td>
<td>SET REMINDER (5), PLAY MESSAGE (5)</td>
</tr>
<tr>
<td>phone</td>
<td>move mobile phone towards ear</td>
<td></td>
</tr>
<tr>
<td>Symbolic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>draw ‘record’-symbol</td>
<td>RECORD MOVIE (4)</td>
</tr>
<tr>
<td>L</td>
<td>draw ‘L’</td>
<td>SHOW LIST (6)</td>
</tr>
<tr>
<td>&gt;</td>
<td>draw ‘play’-symbol</td>
<td>PLAY MOVIE (3)</td>
</tr>
<tr>
<td>X</td>
<td>draw ‘X’</td>
<td>DELETE MOVIE (13)</td>
</tr>
<tr>
<td>?</td>
<td>draw question mark</td>
<td>HELP (14)</td>
</tr>
<tr>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>landscape</td>
<td>tilt mobile phone</td>
<td>SHOW EPG (5)</td>
</tr>
<tr>
<td>remind me</td>
<td>point to myself</td>
<td>SET REMINDER (2)</td>
</tr>
<tr>
<td>point downwards</td>
<td>point towards floor</td>
<td>ABORT (5)</td>
</tr>
<tr>
<td>wipe</td>
<td>horizontal shake</td>
<td>ABORT (15), DELETE MOVIE (11)</td>
</tr>
<tr>
<td>wave</td>
<td>vertical shake</td>
<td>ABORT (6), DELETE MOVIE (6)</td>
</tr>
</tbody>
</table>

Table 4.4: Gesture-referent mapping

in both a pre-defined and a custom-trained gesture set, this finding may be helpful, as the distinction can be made by experts.

4.3 Clarity of Gesture Vocabulary

Based on the results of the previous study, we selected 26 out of the 174 gestures to compose a first gesture vocabulary. A description of the gestures and their associated referents are listed in Table 4.4. Some gestures (left, right, up, down, circle) were included twice, either as dynamic pose / static path or small / wide in form.

We conducted a second study, in which participants were asked to assign referents to the previously selected gestures. 22 adults who had not previously participated in any of the two studies were shown a short video for every single gesture. For each gesture, they should assign one or more suitable referents. These were taken from the previous study and extended to an additional referent, namely BRIGHTEN LIGHTS. Gestures were shown in randomized order, whereas the sequence of referent blocks (one block for each device) was randomized once for each participant. The design of the questionnaire is illustrated in Figure 4.2.
4.3 Clarity of Gesture Vocabulary

According to our gesture-referent mapping (see Table 4.4), some gestures could be used for more than one referent ($Mean = 2.26, SD = 1.53$), and vice versa ($Mean = 2.54, SD = 1.53$). Particular with regard to metaphorical gestures for less complex referents, gestures were mapped to up to five referents (e.g. a swing to the right). On the part of symbolic and abstract gestures (e.g. drawing a question mark), however, a mapping to only one referent was the normal case.

The survey yielded a total of 1570 assigned referents to gestures, with an average of 56.07 ($SD = 7.97$) for each gesture. Given the number of 22 participants, 2.55 referents were assigned to one gesture on average. The least checkmarks (39) were set for the gesture ‘alarm clock’: seven participants selected ABORT, while only five picked SET REMINDER. The gesture ‘right’ got the highest number of assignments (79): most frequently selected referents were NEXT CHANNEL (14) and NEXT MESSAGE (13).

As in the previous study, we calculated the agreement score $A_r$ (c.f. Equation 4.1) for each referent. The agreement for gestures with referents with low complexity was expected to take a higher score than for those gestures associated with referents with high complexity. This results in a negative correlation between the conceptual complexity and the agreement score ($r = -0.490, p < .05$). The agreement measure calculated for the previous study correlates with the agreement found in this one ($r = 0.457, p < .05$).

The results of this second experiment confirmed the findings of the previous study. As simpler referents yielded more intuitive gestures in the previous study, these could be mapped back to their original referents more successfully in this study. Due to the findings of this study, we further reduced the gesture-referent mapping.
Preliminary Studies

4.4 Memorability of Gestures

Inspired by the work of Nielsen et al. (2004), we conducted a third study that should examine the memorability of the elected gestures. We invited ten participants (five women and five men), who had not previously participated in one of the experiments. The participants were mostly students of diverse disciplines, aged between 22 and 35 ($Mean = 26$, $SD = 4$). The study was divided into two parts:

**Step 1** (repeated for each gesture)

1. A video of each gesture and the associated referents (one to three) are shown to the participant (see Figure 4.3a).

2. The participant (with the mobile device in his hand) repeats the gesture five times while the video is shown repeatedly.

3. The participant is asked to rate the suitability of the gesture for the intended purpose (one rating for each assigned referent) on a 7-point Likert scale. In addition, the subjectively experienced effort for each gesture is queried on a SEA scale (Eilers et al., 1986).

**Step 2** (continued until all corresponding gestures were correctly performed)

1. One by one, all referents are displayed on a screen, each for five seconds (see Figure 4.3b). Referents are arranged by device, the order of devices is random. The participant is asked to perform the corresponding gesture within these five seconds.

2. a) If the participant made the right gesture, the presentation of referents continues.
4.5 Type of Feedback

b) If the gesture could not be recalled within time, the presentation is stopped. The participant is again shown the correct gesture associated with the presented referent. After adding the referent to the end of the list, the presentation continues.

As a measure of memorability of a gesture, we drew on the number of attempts that a participant needed: the fewer repetitions, the more memorable the gesture seemed for the corresponding referent. As a result, memorability was negatively correlated with gesture suitability ($r = -0.678, p < 0.0001$). That is to say, gestures rated unsuitable for a certain referent were also more difficult to remember. Gesture suitability was correlated with the suitability ratings achieved in the first study ($r = 0.631, p < 0.01$) and negatively correlated with referents’ conceptual complexity ($r = -0.647, p < 0.01$).

Confirming the results of the previous studies, the memorability test as well as the associated questionnaires showed that users are able to distinguish intuitive gesture-referent mappings from less intuitive ones.

4.5 Type of Feedback

When choosing the kind of feedback for gestural commands, vibration feedback seemed to be very common, as it is provided by most mobile phones. However, opinions were divided on whether vibration reports the success or failure of an operation. We asked 31 adults about their expectations; since we suspected that there might be a difference between iPhone users (eight participants) and non-iPhone users (23 participants), we distinguished between the two groups of users.

Figure 4.4 illustrates the results of this survey. Four out of eight iPhone users associated a successfully completed operation with a vibration feedback. Among the other participants, ten out of 23 were of the same opinion. By contrast, three iPhone users as well as seven non-iPhone users regarded a vibration feedback as an error. Interestingly, a vibration feedback was associated both negative and positive for seven participants (22.58% of all participants).
Figure 4.4: Results of the survey on vibration feedback: (i) negatively associated; (ii) both negative and positive; (iii) positively associated
5 Implementation

This chapter deals with the implementation of the previously presented solution approach. It highlights relevant decisions and gives insight into the workings. A test of the implemented application closes the chapter.

5.1 Choosing an Appropriate Device

Before we could reach the actual implementation phase, it was necessary to find a suitable mobile device for our needs. Besides technical requirements, it also had to meet social ones – both are listed below.

**Popularity**  To reach the widest possible range of users, it was important that the device was widely accepted.

**Sensors**  For the detection of gestures, the device should have built-in acceleration sensors. These sensors will be used to record the user’s motion for controlling home appliances.

**Input modality**  The GUI should allow for touch-based interaction using a touch screen. A customizable interface would be beneficial

**Connectivity**  To be able to connect to the smart-home system, the device needs to have Bluetooth or Wi-Fi capabilities.

Due to the criteria above, it was possible to reduce the eligible devices to a few. For example, the Wii could be eliminated for lack of a GUI as well as others presented in chapter 2. At the present market, there were two suitable devices, namely the Apple iPhone and one of the smartphones running the Android\(^1\) Operating System (OS). We opted for the iPhone, which has the following characteristics:

\(^1\)http://www.android.com
Implementation

**Accelerometer** The *iPhone* responds to motion using a built-in accelerometer. As the device moves, linear acceleration changes along the primary axes in three-dimensional space are reported (see also Figure 2.1). Reporting intervals may be set to 10ms, which corresponds to a 100 Hz update rate.

**Multi-Touch Display** The iPhone features a 3.5-inch (diagonal) widescreen multi-touch screen (480-by-320-pixel resolution at 163 ppi). The concept of the UI is based on direct manipulation using multi-touch gestures. Available gestures are swiping, tapping, pinching, and reverse pinching. Additionally, applications may use the built-in accelerometer to alter the screen’s orientation when changing from portrait- to landscape-view.

**Digital Compass** A magnetometer allows the determination of the direction in which the device is pointing.

**Wi-Fi** Within range of a wireless network, it is possible to connect to the Internet or communicate with other connected devices.

![Figure 5.1: Smartphone market share of Apple’s iPhone](http://www.itunes.com/appstore/)

As can be seen from Figure 5.1, the *iPhone* shows a steady growth on the smartphone market, reaching 30% in September 2009. According to a study by The Nielsen Company (2009b), the *iPhone* is currently the number one mobile phone in the United States, holding 4% of the mobile phone market. Furthermore, the *iPhone* audience is age-diverse (see Figure 5.2) and not limited to a small subset. These findings confirm our choice of the device. And yet there is a downside: applications can only be purchased through the *iTunes App Store* (unless one uses a so-called *Jailbreak*) and this is an ever-growing thorn in the side of developers. The approval process appears very much arbitrary and may extend over several months (see for example Wayner (2009)). In the worst case, this could mean a drop in the popularity of the *iPhone*.

---

2http://www.itunes.com/appstore/
5.2 Design

Besides pure gesture recognition capabilities, the device should also offer an intuitive user interface. As mentioned before, we assume that remote control-like functions may be well suited for specific tasks, e.g. selecting a list entry. Figure 5.3 presents a graphical overview of the functionality that the mobile system was expected to provide. In general, the user has three options: control home appliances via gestures or the GUI, modify gestures (includes adding and deleting gestures), or access the help menu.

During the design process, we were not able to determine the direction in which the device was pointing and thus it was not possible to control a home appliance by just pointing at it. For this reason, we had to think about how we can grant access to the controllable devices. Figure 5.4a shows different approaches to the design of the GUI. Figure 5.4a represents the very first draft, containing only four buttons to control devices and a fifth to access an options menu. Figure 5.4b illustrates the next step, which offers an additional place for displaying feedback. Finally, Figure 5.4c forms the current GUI implementation, which is based on predefined elements provided by the iPhone SDK.

The view shown in Figure 5.4c is the entry point of the application and also the view that provides the remote control-like functionality. The four white buttons in the upper half represent the available devices. Pressing one of these buttons activates gesture recognition for the particular device. The blue bar below provides access to list-based menus for recorded videos and the MP3 collection. The tab bar at the bottom of the window allows for selecting different views.
Figure 5.3: A use case diagram presenting a graphical overview of the system functionality
5.3 Implementation Details

Figure 5.4: Stages of GUI design

Figure 5.5 illustrates the process of adding a new gesture. Selecting the *Gestures* tab brings up a list of all stored gestures (zero in Figure 5.5a). To add a new one, the + button in the upper left is pressed. First, the new gesture is described (Figure 5.5b, 5.5c). This is a way to recall the shape of the gesture later on. Upon completion, the user repeats the gesture five times; during the execution, a button is pressed (Figure 5.5d). The progress is indicated by a progress bar at the top. The last step is adding new functions to the previously recorded gesture (Figure 5.5e, 5.5f).

In return, the procedure to control devices is designed to require significantly fewer steps. As previously noted, pressing a button activates gesture recognition for the selected device. Letting go stops recognition and, if recognition succeeded, a short vibration indicates the success of the operation.

5.3 Implementation Details

The choice of the device is linked to some conditions, which will be presented in this section. Further on, used techniques will be emphasized and associated emerging problems discussed.
**Implementation**

![Images of the process of adding a new gesture]

(a) List of available gestures (0)  (b) Prompt for adding a description of the gesture  (c) Description complete

(d) Screen for recording samples  (e) Menu to add a new function to the gesture  (f) Function selected

(g) Process finishing  (h) List of available gestures (1)

**Figure 5.5**: Process of adding a new gesture
5.3 Implementation Details

5.3.1 Developing for the Device

Developing iPhone applications is similar to creating applications for Mac OS X; both use the same tools and many of the same libraries. The usual programming language used is Objective-C. The development itself is restricted to a MAC OS X Leopard operating system, meaning that one needs to run an Intel-based Macintosh computer in order to write iPhone applications.

The iPhone Software Development Kit (SDK) is a suite of tools that forms the basis of the development environment. It includes the in-house Integrated Development Environment (IDE) XCode. However, distributing applications for iPhone (and iPod touch) requires an iPhone Developer Program\(^3\) enrollment.

The underlying system architecture to iPhone OS is similar to the one found in Mac OS X. On top of the kernel, there are four service layers used to implement applications on the platform. The lower layers of the system, namely Core OS and Core Services are fundamental services and reliance of all applications, while the higher-level layers, namely Media and Cocoa Touch, contain more sophisticated services and technologies, providing object-oriented abstractions for lower-level constructs.

5.3.2 Recognizing Gestures

Accelerometer-based gesture recognition includes the recording of data supplied by sensors, as well as the analysis of this data. The remainder of this section describes the entire process, from capturing a gesture to evaluating it.

Recording Gestures

At the outset, the question arose whether the gesture recognition should run continuously or be triggered by the user. The second option was chosen, as the probability of incidental recognitions seemed too high. Apart from this, the battery would be discharged by the high load very quickly.

The iPhone reports acceleration-related data from the onboard hardware as a device moves. The installed accelerometer is a 3-axis [MEMS] motion sensor limited to a measurement range of \(\pm 2g\) (actually \(\pm 2.3g\))\(^4\). Three steps are required in order to receive these events:

\(^3\)http://developer.apple.com/iphone/program
Implementation

1. Specify the `UIAccelerometerDelegate` protocol for a class. This tells the application that this class is going to respond to acceleration and it may have methods associated with this protocol.

2. Add the event listener by setting this class as a delegate for accelerometer events.

3. Create a method to respond to incoming acceleration events.

The raw data forms a 3D acceleration vector that indicates the direction of gravity. We applied a low-pass filter to the stream of acceleration values in order to reduce noise. Listing 5.1 shows a simplified low-pass filter, as provided by Apple. It uses 10% of the unfiltered acceleration and 90% of the previous filtered value (member variables `accelX`, `accelY`, `accelZ`).

```c
#define kFilteringFactor 0.1

- (void)accelerometer:(UIAccelerometer *)accelerometer didAccelerate:(UIAcceleration *)acceleration {
  if (recognitionEnabled) {
    accelX = (acceleration.x * kFilteringFactor) + (accelX * (1.0 - kFilteringFactor));
    accelY = (acceleration.y * kFilteringFactor) + (accelY * (1.0 - kFilteringFactor));
    accelZ = (acceleration.z * kFilteringFactor) + (accelZ * (1.0 - kFilteringFactor));
  }
  // Use the acceleration data.
}
```

Listing 5.1: Isolating the effects of gravity from accelerometer data

Figure 5.6 illustrates the system behavior for the case of controlling a device via gestures. Once recognition has been enabled by touching a device’s button on the screen, incoming acceleration data is filtered and written into an array. Letting go stops recording, and the application continues by identifying stored gestures, which have been assigned functions for this device. In case there are no assigned gestures, a list of all available functions is displayed from which the user may choose the desired one. Otherwise, the system proceeds by matching the recorded gesture with those that were previously assigned to the device’s functions. If the gesture could not be matched to the stored ones, the function list for the device will be shown as described above. A positive match or a selection of an entry of the list terminates the procedure by sending the appropriate command to the smart-home system.
5.3 Implementation Details

Figure 5.6: State diagram giving an abstract description of the behavior of the system when executing a gesture command
Implementation

The previously mentioned ‘matching’, which is the gesture recognition algorithm, will be introduced in the following chapter.

Recognition Algorithm

Dynamic Time Warping (DTW) is a technique that was introduced by Sakoe and Chiba (1990) in the field of speech recognition. It is able to find the optimal alignment of two given data sequences, where one of them is stretched or shrunk along the time axis.

Given two time series, \( X = (x_1, x_2, \ldots, x_m), m \in \mathbb{N} \) and \( Y = (y_1, y_2, \ldots, y_n), n \in \mathbb{N} \), represented by the sequences of acceleration values, the DTW algorithm provides an optimal alignment in \( O(n^2) \) (if \( N = |X| = |Y| \)) time and space complexity.

Since samples are taken from a feature space \( \Phi \), where each sample is represented by a value in three-dimensional space, it is necessary to apply a local distance measurement function in order to compare the sequences \( X, Y \in \Phi \):

\[
d : \Phi \times \Phi \to \mathbb{R} \geq 0
\]

(5.1)

Similar sequences lead to a small value for \( d \), whereas dissimilar ones lead to a large value. For our application, we decided to use the Mahalanobis distance function introduced by Mahalanobis (1936), since, in contrast to Euclidean distance, it takes correlations of the values into account.

\[
D_{m,n} = \begin{pmatrix}
d_{1,m} & \cdots & d_{m,n} \\
\vdots & \ddots & \vdots \\
d_{1,1} & \cdots & d_{n,1}
\end{pmatrix}
\]

(5.2)

To find the optimal alignment, which is the minimized arrangement of all sequence points, the DTW algorithm starts with creating a distance matrix \( D \in \mathbb{R}^{m \times n} \). The matrix, as illustrated in Equation 5.2, contains all paired distances (costs) of \( X \) and \( Y \):

\[
D \in \mathbb{R}^{m \times n} : d_{i,j} = \|x_i - y_j\| \quad i \in [1, \ldots, m], j \in [1, \ldots, n]
\]

(5.3)

Using a dynamic programming approach, the cost matrix is filled column by column, beginning at \( d_{1,1} \) and ending at \( d_{m,n} \), as depicted in Algorithm 1.

After filling the matrix, the algorithm proceeds with a greedy search to determine the alignment path. Actually, the path is calculated in reverse order, from \( d_{m,n} \) to \( d_{1,1} \) (c.f. Algorithm 2). The path itself is a sequence of points \( p = (p_1, p_2, \ldots, p_k) \) with
5.3 Implementation Details

**Algorithm 1**: Filling the cost matrix $D$ for sequences $X,Y$

```
begin
    $m \leftarrow |X|$
    $n \leftarrow |Y|$
    for $i \leftarrow 1$ to $m$ do
        for $j \leftarrow 1$ to $n$ do
            $d_1, d_2, d_3 \leftarrow \infty$
            $dist \leftarrow \text{mahalanobis}(i,j)$
            if $i = 1$ and $j = 1$ then $d_1, d_2, d_3 \leftarrow 0$
            else if $i > 1$ and $j = 1$ then
                $d_1 \leftarrow D(i-1,1)$
            else if $i = 1$ and $j > 1$ then
                $d_2 \leftarrow D(1,j-1)$
            else if $i > 1$ and $j > 1$ then
                $d_1 \leftarrow D(i-1,1)$
                $d_2 \leftarrow D(1,j-1)$
                $d_3 \leftarrow D(i-1,j-1)$
            $D(i,j) \leftarrow dist + \min(d_1, d_2, d_3)$
        end
    end
end
```

$p_l = p(i,j), i \in [1,\ldots,m], j \in [1,\ldots,n], l \in [1,\ldots,k]$. The following conditions have to be met:

**Path length** $\max(|X|,|Y|) \leq k \leq |X| + |Y|$

**Boundaries** $p_1 = (1,1), p_k = (m,n)$

**Monotonicity** $p_l = (i,j), p_{l+1} = (i',j') \quad i \leq i' \leq i+1, j \leq j' \leq j+1$

The optimal warp path is the one with minimal cost, which is the summation of all cost values along the warp path:

$$c_p(X,Y) = \sum_{l=1}^{k} d(p_l)$$  \hfill (5.4)

### 5.3.3 Identification of Problematic Areas and Their Solutions

Of course, the implementation did not go completely smoothly. Sometimes, obstacles arose that were not initially apparent. These are discussed in this chapter, and possible solutions are presented.
Algorithm 2: Find an optimal path through matrix $D_{m\times n}$

begin
\begin{align*}
  i &\leftarrow m \\
  j &\leftarrow n \\
  \text{path}[] &\leftarrow \text{new Array} \\
  \text{cost} &\leftarrow \text{D}(i, j) \\
  \text{while } i > 1 \text{ or } j > 1 \text{ do} \\
  \quad \text{down, left, diagonal} &\leftarrow \infty \\
  \quad \text{if } i > 1 \text{ and } j > 1 \text{ then} \\
  \quad \quad \text{down} &\leftarrow \text{D}(i - 1, j) \\
  \quad \quad \text{left} &\leftarrow \text{D}(i, j - 1) \\
  \quad \quad \text{diagonal} &\leftarrow \text{D}(i - 1, j - 1) \\
  \quad \text{else if } i = 1 \text{ then} \\
  \quad \quad \text{down} &\leftarrow \text{D}(1, j - 1) \\
  \quad \text{else if } j = 1 \text{ then} \\
  \quad \quad \text{left} &\leftarrow \text{D}(i - 1, 1) \\
  \quad \text{next} &\leftarrow \min(\text{down, left, diagonal}) \\
  \quad \text{cost} &\leftarrow \text{cost} + \text{next} \\
  \text{switch next do} \\
  \quad \text{case down} \\
  \quad \quad i &\leftarrow i - 1 \\
  \quad \text{case left} \\
  \quad \quad j &\leftarrow j - 1 \\
  \quad \text{case diagonal} \\
  \quad \quad i &\leftarrow i - 1 \\
  \quad \quad j &\leftarrow j - 1 \\
  \quad \text{path.add}(i, j) \\
\end{align*}
end
5.3 Implementation Details

**Slowness of Gesture Recognition Algorithm**

Prerequisite for acceptance by potential users of this alternative input modality was processing speed, which could come close to that of a common remote control. However, since we worked with a mobile device with limited processing power and the amount of data could quickly take on a large scale, the above-described approach came up against its limits.

Gestures that need one second to perform, deliver 100 acceleration values at an update rate of 100Hz. The comparison of these gestures constructs a matrix with 10,000 values, which need to be calculated.

To overcome this exclusion criterion, we applied an improvement to the existing algorithm. Salvador and Chan (2007), facing similar problems, amended the DTW algorithm with the addition of data abstraction and constraints. In the field of DTW constraints refer to limiting the number of cells that are evaluated in the cost matrix, whereas data abstraction points to running the algorithm on a reduced representation of the data. The Fast Dynamic Time Warping (FastDTW) algorithm is based on a multilevel approach, which can be basically described by three operations:

**Coarsening** Create lower resolutions of a data sequence that represents the same curve as accurately as possible. This is done by averaging adjacent pairs of points.

**Projection** Find a minimum-distance warp path at a lower resolution, and use this as a starting point for the next higher resolution. Since the resolution is increased by a factor of two, a single point of a path maps to at least four points in the next higher resolution (or > 4 if |X| ≠ |Y|).

**Refinement** Find the optimal warp path in the neighborhood of the projected path.

Figure 5.7 illustrates a complete run of the FastDTW algorithm. The full DTW algorithm is applied to the lowest resolution (1/8 in this case). The path found at this resolution is projected to the next higher resolution of 1/4, indicated by the heavily shaded cells in Figure 5.7b. A constrained DTW algorithm is run in order to refine the path; the constraint limits the evaluated cells to those in the projected warp path. However, the entire optimal path may not be inside the projected one. To overcome this, a radius parameter extends the search path with additional cells on each side of the projected path, increasing the chance of finding the optimal path. This is illustrated by the lightly shaded cells. The algorithm proceeds until the full resolution (c.f. Figure 5.7d) is reached. The warp path found at a resolution of 1 is the optimal warp path for FastDTW. A high-level description of the algorithm can be found in Algorithm 3.
Implementation

Figure 5.7: Example of four different resolutions evaluated during a complete run of FastDTW \((radius = 1)\) (Salvador and Chan, 2007)

The increase in efficiency, however, is shown to the best advantage for larger data sets, as the overhead of creating multiple resolutions limits its use for smaller ones. The number of cells evaluated by FastDTW scales linearly with the number of acceleration values, while DTW evaluates \(n^2\) cells (if \(N = |X| = |Y|\)). The linearity is due to the constant width of the warp path.

Algorithm 3: FastDTW

begin
\[
\begin{align*}
\text{resolution} & \leftarrow \max(X.\text{minResolution}, Y.\text{minResolution}) \\
\text{sequenceX} & \leftarrow X.\text{getResolution}(\text{resolution}) \\
\text{sequenceY} & \leftarrow Y.\text{getResolution}(\text{resolution}) \\
\text{window} & \leftarrow \text{fillMatrix}(\text{sequenceX}, \text{sequenceY}) \\
\text{path} & \leftarrow \text{findPath}(\text{window})
\end{align*}
\]
while \(\text{resolution} < 1.0\) do
\[
\begin{align*}
\text{resolution} & \leftarrow \text{resolution} \times 2 \\
\text{sequenceX} & \leftarrow X.\text{getResolution}(\text{resolution}) \\
\text{sequenceY} & \leftarrow Y.\text{getResolution}(\text{resolution}) \\
\text{expandedPath} & \leftarrow \text{expandPath}(\text{path}) \\
\text{window} & \leftarrow \text{expandWindow}(\text{sequenceX}, \text{sequenceY}, \text{expandedPath}, \text{radius}) \\
\text{path} & \leftarrow \text{findPath}(\text{window})
\end{align*}
\]
end

In addition to speeding up the algorithm, we switched to an iPhone 3GS with greater processing power (c.f. Table 5.1 for a comparison of iPhone 3G and 3GS). A short test with 20 iterations and a gesture-length of two seconds matching a single gesture produced promising results.
### 5.3 Implementation Details

<table>
<thead>
<tr>
<th></th>
<th>iPhone 3G</th>
<th>iPhone 3GS</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU (MHz)</td>
<td>412</td>
<td>600</td>
</tr>
<tr>
<td>RAM (MB)</td>
<td>128</td>
<td>256</td>
</tr>
</tbody>
</table>

**Table 5.1:** Technical specifications for iPhone 3G and 3GS

**iPhone 3G vs. 3GS** The 3GS with an average calculation-time of 0.287 seconds ($SD = 0.089$) is significantly faster in processing the algorithm than the 3G with an average calculation-time of 1.615 seconds ($SD = 0.353$). The increase in speed is therefore about 560%.

**DTW vs. FastDTW** FastDTW has a clear advantage in speed compared to DTW. Running on the 3GS, FastDTW needs 0.287 seconds ($SD = 0.089$) on average, whereas DTW finishes after 1.491 seconds ($SD = 0.094$). The increase in speed in this case is 520%.

### Decreased Recognition Rate Due to Irregular Training

At the gesture training phase, it can happen that one of the five sample gestures is significantly different from the others. According to the previously shown procedure, this would lead to a higher average cost for this gesture. With regard to gesture recognition, this can mean, in the worst case, that the comparison of two different gestures delivers a positive match.

Since the detection of so-called ‘outliers’ is crucial for statistical applications, there are a number of methods to detect these. We applied *Grubb’s test for outliers* (Grubbs, 1969) in order to limit the possibility of false positive recognitions.

\[ Z = \left| \frac{\text{mean} - \text{value}}{\text{SD}} \right| \]  

(5.5)

If the calculated value of $Z$ is greater than a predefined critical value (see Table 5.2), the result is significant at a level of $p \leq 0.05$. After all five samples have been recorded, the $Z$-value is calculated for each sample. If a sample is identified as an outlier ($Z > \text{critical value}$), it is excluded from all further calculations.
Implementation

<table>
<thead>
<tr>
<th>Number of samples</th>
<th>Critical Z value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.15</td>
</tr>
<tr>
<td>4</td>
<td>1.48</td>
</tr>
<tr>
<td>5</td>
<td>1.71</td>
</tr>
</tbody>
</table>

Table 5.2: Critical values for Z (see GraphPad Software)

5.4 Concurrent Validation

Finally, a verification test is run to show that the implemented gesture recognition algorithm meets the requirements. Identified goals were the best possible recognition rate within an acceptable execution time.

Seven gestures (see Figure 5.9), which are later used in the usability study, were trained and then performed each 50 times. Recognition results are presented in Figure 5.8. With a recognition rate of 96.87% (2.28% not recognized, 0.85% incorrectly recognized) and an average execution time of 0.447 seconds ($SD = 0.05$), the implementation meets the previously stated targets.
Figure 5.8: Recognition rates for seven gestures, 50 tries per gesture, single user
Figure 5.9: Used gestures, represented by long exposure photos
6 Evaluation

This chapter deals with a usability study to evaluate the previous work. The remainder of this chapter defines the objectives, states hypotheses and describes the test design to evaluate these. Subsequently, the findings of this study are presented and subjected to analysis.

6.1 Goal

This study evaluates the approach presented in this thesis. As already demonstrated in Chapter 5.4, the application performs reliable gesture recognition operations. Away from the technical side, it is of prime importance whether potential users of such a system would ever use this modality and how they would cope with it. For this reason, it is a further object of this study to make a comparison between alternative methods for controlling home appliances.

6.2 Hypothesis

Given that the input modality is a novelty for all participating users, it can be assumed that some time is needed in order to get accustomed to this method. As the experience with motion-controlled devices is likely to differ from participant to participant, the following hypothesis can be established:

Hypothesis 1 Experienced participants achieve higher recognition rates.

Since interpersonal communication is mainly composed of voice and gestures, it seems likely that this combination is also reasonable for [HCI].

Hypothesis 2 Participants prefer multi-modal interaction.

It is assumed that the disappointment of poor recognition rates may be reflected in the evaluation of each modality - this leads to the following hypotheses:
Evaluation

Figure 6.1: Sketch of the experimental room, including (i) TV and hi-fi equipment, (ii) the task-displaying monitor, (iii) wall light, (iv) ceiling light, (v) blinds, (vi) a camcorder and (vii) the ‘wizard’s’ room

Hypothesis 3  The higher the recognition rate, the better the rating of the system

Participants who have experience in dealing with motion-controlled devices, are assumed to be more open-minded towards the gestural interface, while potential scrupulosity on behalf of inexperienced users might constitute an obstacle, which leads to another hypothesis:

Hypothesis 4  Experienced participants rate the gestural modality better than inexperienced participants.

6.3 Test Design

The study was conducted in a fully furnished living room (see Figure 6.1), where the electronic devices are connected to a network. Among all available devices, only the TV, Hi-Fi, lighting and blinds were used for the study. The participants were seated on a sofa, from where they had all devices in the field of view. For later analysis, the user’s interaction was recorded on video.
At the beginning, participants were given a questionnaire, asking for general information (age, gender, and level of experience with iPhone or Wii). Afterwards, a brief introduction to the experimental setup was given.

The study itself was divided into three parts, each examining the usability of one input modality: speech, gestures, multi-modal. The order of these three parts was randomized. Because of the immaturity of the speech recognition system, we used a ‘Wizard of Oz’-setting for the corresponding tasks. Generally, a ‘Wizard of Oz’-experiment is conducted by an experimenter (the ‘wizard’), simulating the behavior of an intelligent computer system. In this case, a researcher replaced the speech recognition system, which led to a 100% recognition rate.

Each test run began with a short introductory video, illustrating the basic functionality of the current modality. This was followed by a training phase. To ensure a consistent procedure, even the simulated speech recognition system was trained by reading a given text aloud. Participants were given a clip-on wireless microphone, which should allow for communication to the system (in this case, the ‘wizard’). Gestures were each shown as a video and then, as described in Chapter 5, recorded. After completion of the training, there was a guided test task. Upon fulfillment of the task, the briefing person left the room for the following run-through.

6.3.1 Task Design

Each test run consisted of several tasks (nine to eleven), which could in turn consist of sub-tasks.\textsuperscript{1} A complete task block is shown below:

- Lower the blinds and stop them. (guided task)
- Turn on the ceiling light and try to dim it.
- Turn on the TV. Navigate to the television-stations Sat1 and RTL. Turn down the volume and then turn off the TV.
- Play the biathlon video. Mute the sound.
- Delete two tracks from your ‘favorites’ playlist. Add two new titles.
- Find out which movies are running tonight and record one of them.
- Zap through the radio stations. Turn down the volume. Switch to the next station. Mute the sound.

\textsuperscript{1}The task blocks can be found in the appendix.
Evaluation

- Show the list of MP3s. Navigate to the albums of an interpreter.
- Now show the titles of an album. Play a track and then stop it.

These tasks were displayed on a screen in front of the participant (one task at a time). Depending on the modality, the participant had different options for fulfilling the task:

**Smartphone** Participants were asked to complete the tasks using the GUI and gestures. The tasks that could be solved by means of gestures were turning on/off the TV and radio, switching the channel (TV, radio), showing the EPG and the list of available radio stations, and controlling the blinds (lower, raise, stop) and lighting (on/off, brighten/dim).

**Voice** The system should be controlled using natural language.

**Multi-modal** For each task, users could choose the most appropriate modality for them. Participants were free to switch back and forth between the modalities.

Upon fulfillment of a task, the next task was displayed. Each task block completed with a questionnaire.

### 6.3.2 Questionnaire

Participants were asked to complete a questionnaire\(^2\) for the respective modality after each run. The questionnaire consisted of five pages with a total of 60 questions. The questionnaire itself was composed of different questionnaires:

- A general question regarding the quality of the system
- AttrakDiff (condensed form): asks for experienced attractiveness, in terms of usability and appearance ([Hassenzahl et al.] 2003)
- Quesi: assesses intuitiveness of the system ([Humane-Machine Systems Department of Psychology and Ergonomics, TU Berlin])
- SUS: evaluates usability of a system ([Brooke 1996])
- USE (condensed form): gathers information on usefulness and ease of use ([Lund 2001])
- A questionnaire regarding system output quality

\(^2\)The questionnaires can be found in the appendix.
6.3.3 Participants

14 women and 13 men took part in this study, aged between 21 and 35 (\(\text{Mean} = 26.1, \text{SD} = 3.98\)). Five of them were in possession of an iPhone, four rarely used one; three participants said that they often play with a Wii, two rarely. Participants with prior experience with one of the two systems are referred to as ‘experienced’. In total, the group sizes of ‘experienced’ and ‘inexperienced’ participants were balanced.

6.4 Results

This section presents the results of the usability study, subdivided into two parts. In the first part the gesture recognition is evaluated, the second part deals with the analysis of questionnaires.

6.4.1 Evaluation of the Gesture Recognition Rate

In total, participants used 1100 gestures in order to control the available devices, which means an average of 42.31 gestures per participant (\(\text{SD} = 11.64\)). However, the number of recognizable gestures is reduced to 893 after analyzing the recorded videos. The term ‘not recognizable’ refers to gestures, which apparently can not be assigned to any of the previously trained gestures. The percentage of unrecognizable gestures is thus 18.82% with an average of 7.96 unrecognizable gestures per participant (\(\text{SD} = 6.04\)). The resulting recognition rates are shown in Figure 6.2. The figure is divided into two pie charts: Figure 6.2a contains all gestures, while Figure 6.2b is limited to recognizable gestures. This distinction has been made to prevent a dilution of the recognition rate by unrecognizable gestures.

The remainder of this section refers to recognizable gestures and omits unrecognizable gestures for the time being. Recognition rates varied widely among participants, for example, the maximum was 28 out of 34 gestures (82.34%) for a participant, the minimum only six out of 21 (28.57%). The time it took the program to check a gesture is 0.57 seconds on average (\(\text{SD} = 0.497\)).

6.4.2 Analysis of Survey Data

An analysis of variance regarding the ratings of the overall quality of the three modalities yielded a difference that was just under being significant (\(F(2, 71) = 2.84, p = 0.065\)). Friedman’s analysis of variance by ranks, however, shows that using gestures and the
**Evaluation**

![Recognition rates for gestures and gestures limited to recognizable gestures](image)

(a) Recognition rate for gestures that are (i) correctly recognized; (ii) unrecognizable; (iii) incorrectly recognized; (iv) not recognized

(b) Recognition rate limited to gestures that are (i) correctly recognized; (ii) incorrectly recognized; (iii) not recognized

**Figure 6.2:** Recognition rates for (a) all performed gestures; (b) recognizable gestures

GUI as input modality is graded lower than multi-modal or voice input ($\chi^2 = 17.65, p < .001$).

A factor analysis of the questionnaires provided three factors (KMO: 86.2, explained variance: 62.7%) that can be interpreted as follows:

1. Pragmatic quality (27.6% explained variance, Cronbach’s alpha: 0.97)
2. Information presentation (20.8% explained variance, Cronbach’s alpha: 0.976)
3. Hedonic quality (14.3% explained variance, Cronbach’s alpha: 0.914)

Subsequently, an analysis of variance should show whether the modality has an impact based on the three identified factors. For the factor that was interpreted as pragmatic quality, a significant impact for the modalities was found ($F(2, 50) = 14.58, p < .0001$). A post-hoc analysis showed that the modality using the mobile phone was rated worse than the multi-modal approach, which was in turn rated worse than voice input. Concerning this factor, ratings by participants who were experienced with the **Wii** were significantly higher for all modalities ($F(2, 15) = 4.74, p < .05$). The factor, which aimed at information presentation, also showed a significant impact of modalities ($F(2, 50) = 5.92, p < .005$). Again, the modality using gestures and the **GUI** was rated worse than the other two modalities. In contrast to the first two factors, the one regarded as hedonic quality showed no significant difference ($F(2, 50) = 2.58, p = .086$).
6.5 Discussion

<table>
<thead>
<tr>
<th>Modality</th>
<th>Frequency</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartphone</td>
<td></td>
<td>72.56</td>
<td>14.86</td>
</tr>
<tr>
<td>Voice</td>
<td></td>
<td>43.59</td>
<td>6.63</td>
</tr>
<tr>
<td>Multi-modal</td>
<td>Smartphone</td>
<td>38.96</td>
<td>19.41</td>
</tr>
<tr>
<td></td>
<td>Voice</td>
<td>14.35</td>
<td>10.12</td>
</tr>
</tbody>
</table>

Table 6.1: Differences in the frequency of use of modalities

For the questions that have been grouped under this factor, however, *iPhone* users rated significantly better \( F(2, 15) = 5.20, p < .05 \).

A final interview should provide participants with space for comments and indicate the reason for their answers in the questionnaire. For example, 17 participants said that voice input worked well, the gestures, however, were difficult to remember or did not work as intended. In this manner, 12 participants tended to use voice control during the multi-modal test phase, whereas five were inclined to control the devices via gestures or GUI. However, when analyzing the frequency of the input types used in the three phases, an opposite trend was found. Table 6.1 presents the average frequency of use of the modalities.

6.4.3 Correlation between Recognition Rate and Ratings

A final examination should show, whether there is a correlation between recognition rate and ratings (overall quality and grades). Since there was no true speech recognition by the system, only the gesture recognition was considered.

Using *Spearman’s rank correlation coefficient*, a significant correlation of recognition rate and overall quality rating of the system was found \( r_s = .423, p < .05 \). However, a correlation between recognition rate and school grades could not be detected \( r_s = −.12, p = .544 \). A visualization of the correlations is given in Figure 6.3.

6.5 Discussion

As realized in hindsight, there was a bug in the gesture recognition algorithm. When filling the cost matrix (as depicted in Algorithm 1), two values were embezzled by mistake. Algorithm 4 shows an excerpt from the complete algorithm, the two missing assignments are indicated by strikeout. That is to say, the construction of the warp path may not find the optimal alignment and therefore delivers suboptimal results. Consid-
Evaluation

Figure 6.3: Correlation between recognition rate and (a) overall quality ratings; (b) grades

As indicated by the participants’ comments, the reason for the high number of unrecognizable gestures may be that the training period was too short to remember all the gestures. While the participants went through a relatively short training period, the series of tests described in Chapter 5.4 was conducted by an expert who was able to rehearse the gestures over a longer period of time. Even though there was no significant correlation between recognition rate and grades, this might be one reason for the difference in the modalities regarding the factor that was identified as pragmatic quality.
As the experience of often false or not recognized gestures arose out of the participant’s interaction, it would be understandable if this modality would be rated worse than the modality with a perfect recognition rate. For this reason, the second hypothesis cannot be confirmed, as voice input was the preferred modality. Not only in this case would it be interesting to know how an error rate of about the same size between the modalities would have affected the results.

Despite the experience with motion-controlled devices on the side of participants who were in possession of a *Wii* or *iPhone*, there was no significant influence of experience on recognition rates. Thus, this finding does not confirm the first hypothesis. This could be explained by the fact that the nature of the gestures was different, as well as the interaction scenario from the one the participants were familiar with. These two points may diminish the presumed advantage of the experienced users. Hypothesis 4, based on the assumption that experienced users are likely to give better ratings concerning the gestural modality can be partially confirmed. As found for pragmatic and hedonic quality, experienced users gave better ratings in general, not only for this specific one. The initially assumed more open attitude of the experienced users could provide the reason for this. That this does not apply to the information presentation and that the interaction method using the *iPhone* was rated worse than the other modalities, may be due to a generally poor feedback, which was needed even more when using gestures and the GUI. For example, there was a positive feedback (vibration) in the event of an incorrectly recognized gesture. This led to confusion among the participants, since the response of the smart-home system was different than expected. A brief note on which gesture (or the corresponding referent) was recognized would certainly have been helpful.

As described in section 6.4.3 a significant correlation between recognition rate and overall quality ratings was found. This partially confirms hypothesis 3, as ratings were divided into overall quality ratings and grades. After each test run, participants were asked for an overall quality rating, while grades were given at the final questionnaire after the three modalities were tested. Thus one can assume that when comparing the modalities, the detection rate plays a minor role in the direct evaluation of each modality. This minor role may also be true for the hedonic quality, in which the interaction through the *iPhone* is possibly considered fancier, compensating for the low recognition rate in comparison to voice input. As already mentioned in the previous section, 12 participants said that they tended to use voice input in the multi-modal phase. According to the data analysis, this is only true in four cases. In the remaining cases, the smartphone has been used more frequently (see Table 6.1). It should be noted that in the case of controlling
Evaluation

the room lighting with the smartphone, an additional selection of the lamp is necessary, which increases the number of interactions.
7 Conclusions

The research in this thesis addresses the question of how controls for home appliances can be made more user-friendly and easier to use. Today’s home environments are equipped with increasingly more technology. However, interaction techniques remain the same, leading to a clutter on the coffee table. An existing approach to this problem is the combination of multiple remote controls to a single ‘universal’ remote control. Although eliminating the issue of a multitude of remote controls, these devices confront the user with a plethora of buttons and modes.

There are also approaches to the design of smart environments with alternative input techniques. Due to their requirements, only a few are eligible for the use in a smart-home environment. Voice-based interfaces seem to be well suited, but are still subject to high error rates, which complicates the practical use.

This chapter will now give a final consideration of set goals and their implementation and provide an outlook on future work.

7.1 Summary

In this thesis, the problem of an inadequate usability for controlling home appliances is addressed. Previous approaches were promising, but may not measure up to users’ expectations of a user-friendly control modality. This led to the approach of designing an alternative input modality using gestures, which should facilitate users to control their home appliances. This approach provided for the use of a commercially available smartphone with built-in acceleration sensors. Using the smartphone SDK, it was possible to create a custom application being able to record gestures and additionally providing remote control-like functions. In a preliminary study, users were involved in the process of designing a set of gestures for smart-home control. This study included three steps: at first, users were asked to conceive gestures for predetermined tasks, resulting in a first gesture vocabulary; in the second step, users were asked to reassign commands to given gestures from the gesture vocabulary; the final test should examine the memorability of the gesture-command mappings. Additionally, a gesture recognition system was imple-
Conclusions

mented on the mobile device. This system allows controlling devices via gestures or by using the GUI. Finally, a study with 27 participants evaluated the usability of the implemented approach. Participants compared speech-based input with gestural control and a combination of both modalities.

7.2 Results

The solution presented in this thesis meets the requirements that were initially established. This section casts light on crucial points that have contributed to the outcome of this work.

An initially conducted user survey showed that participants were generally interested in a gesture control for their home appliances. As motion-controlled devices gain in popularity, this trend may continue. By means of a further series of studies we have identified a set of gestures, which showed themselves as adequate for use in a smart-home environment.

The implemented gesture recognition yields good results (97%) in an acceptable time for a mobile device. However, gestures need to be rehearsed over an extended period of time to make the system applicable for use outside of laboratory conditions. The final usability study showed that participants were sometimes very uncertain in dealing with the new input modality. This places additional requirements on the design of the application. This is especially true for error handling and support that have not served in part to inexperienced users. However, it is worth noting that a program error that was discovered too late certainly contributed its share to the confusion.

7.3 Future Work

The solution presented in this thesis still leaves points of contact for further research. The most obvious ones will be presented in this section.

As already implied by participants of the usability study, it is advisable to conduct a field experiment over a longer period of time. This would provide the participants the opportunity to rehearse the gestures and become familiar with the new modality. Furthermore, they could, if they wished to, customize gestures or add their own to the vocabulary.

By switching to a newer version of the used smartphone, we could not only notice an increase in processing speed, but the new mobile device came with a magnetometer indicating the direction it is facing. With this improvement, it is possible to change the
7.3 Future Work

design of the application so that device buttons become obsolete, and selecting a device is done by merely pointing at it. Because the direction of devices from the user’s point of view change with the user’s location, the user has to stay at a predetermined location, for example a sofa. In turn, this would mean a disadvantage, to which the determination of the user’s position could provide a remedial option.

We focused on a specific algorithm for recognizing gestures. Although this algorithm delivers good results for an experienced user, it seems reasonable to investigate further techniques (e.g. applying Hidden Markov Models (HMM)) that possibly fit even better in this field of application. Besides that, the threshold could be increased to minimize the number of unrecognized gestures. However, caution needs to be exerted, because the rate of false recognitions may increase at the same time.

The smart-home system that we used in the usability study was still in development. Occasionally, system crashes occurred which led to confusion on the part of participants. A more detailed feedback on the mobile phone could take remedial action.

As development of smart-home environments is still striving forth, the prospects of the integration of a more sophisticated gesture-based interaction modality seem promising.
Bibliography


Bibliography


Bibliography


65
Bibliography


Bibliography

of the 22nd annual ACM symposium on User interface software and technology, New York, NY, USA, 2009. ACM. [9][10]


The Nielsen Company. iPhone Users Watch More Video... and are older than You Think, 2009a. URL http://blog.nielsen.com/nielsenwire/online_mobile/iphone-users-watch-more-video-and-are-older-than-you-think/ Retrieved November 24, 2009. [33]


Appendices
List of Abbreviations

DTW Dynamic Time Warping
EMG Electromyography
EPG Electronic Program Guide
FastDTW Fast Dynamic Time Warping
GUI Graphical User Interface
HCI Human Computer Interaction
HMM Hidden Markov Model
IDE Integrated Development Environment
INSPIRE INfotainment management with SPeech interaction via REmote-microphones and telephone interfaces
IR Infrared
LED Light-Emitting Diode
KMO Kaiser-Meyer-Olkin measure of sampling adequacy
MEMS Micro Electro-Mechanical Systems
NFC Near Field Communication
OS Operating System
PDA Personal Digital Assistant
SD Standard Deviation
SDK Software Development Kit
UI  User Interface
WLAN  Wireless Local Area Network
Task Blocks

Task Block One

- Training: Fahren Sie das Rollo herunter und stoppen Sie es.
- Schalten Sie erst die Wandlampe ein und verändern sie die Helligkeit, bis es Ihnen angenehm erscheint, schalten Sie nun die Deckenlampe ein.
- Welche Sportsendungen laufen heute Abend im Fernsehen? Nehmen Sie eine davon auf. Schließen Sie die Ansicht.
- Welche Radiosender können Sie empfangen? Wählen Sie den Sender, den Sie jetzt hören möchten.
- Schließen Sie die Ansicht. Reduzieren Sie die Lautstärke. Schalten Sie nun das Radio aus.
- Wechseln Sie in die Ansicht Ihrer mp3s. Lassen Sie sich die Alben eines Interpreten anzeigen.
- Navigieren Sie jetzt zu den Titeln eines Albums. Spielen Sie ein Lied ab und beenden es direkt.
- Öffnen Sie ein Album eines anderen Interpreten und fügen Sie einen Titel zu Ihrer Playlist ‘Favoriten’ hinzu. Schließen Sie die Ansicht.
- Schalten Sie den Fernseher ein. Wechseln Sie zunächst zum Sender Pro7 und dann zum Sender RTL. Schalten Sie den Fernseher wieder aus.
- Lassen Sie sich die Liste Ihrer Videoaufnahmen anzeigen und spielen Sie „Die Simpsons“ ab. Unterbrechen Sie den Film, indem Sie den Fernseher ausschalten.
- Öffnen Sie Ihre Playlist ‘Favoriten’. Spielen Sie einen der ersten 20 Titel ab. Unterbrechen Sie das Lied (Pause).
Task Block Two

- Training: Fahren Sie das Rollo herauf und stoppen Sie es.

- Passen Sie die Helligkeit einer Lampe an. Schalten Sie alle Lampen aus.

- Schalten Sie das Radio ein. Lassen Sie sich die Übersicht über die verfügbaren Radiosender anzeigen.

- Wählen Sie einen anderen Sender aus und wechseln Sie zu diesem. Schließen Sie die Ansicht. Schalten Sie das Radio wieder aus.

- Schalten Sie durch die ersten drei der verfügbaren Fernsehsender. Reduzieren Sie die Lautstärke soweit wie möglich. Schalten Sie den Fernseher aus.

- Lassen Sie sich das Fernsehprogramm anzeigen. Wählen Sie einen Film aus und nehmen Sie diesen auf. Schließen Sie die Ansicht.

- Lassen Sie sich die Liste Ihrer Videoaufnahmen anzeigen und löschen Sie ‘Die Simpsons’. Schließen Sie die Ansicht.

- Lassen Sie sich die Liste Ihrer mp3s anzeigen. Lassen Sie sich die Alben eines Interpreten anzeigen.

- Navigieren Sie jetzt zu den Titeln eines Albums. Spielen Sie ein Lied ab und beenden es direkt.

- Erweitern Sie Ihre Playlist ‘Favoriten’ um einige Titel aus dem Album eines anderen Interpreten.


Task Block Three

- Training: Fahren Sie das Rollo herunter und stoppen Sie es.

- Schalten Sie die Deckenlampe ein und versuchen Sie diese dunkler zu machen.

- Schalten Sie den Fernseher ein. Navigieren Sie zu den Sendern Sat1 und dann RTL. Reduzieren Sie die Lautstärke und schalten Sie dann den Fernseher wieder aus.
• Spielen Sie die Aufnahme vom Biathlon ab. Schalten Sie den Ton aus.

• Löschen Sie zwei Titel von Ihrer Playlist ‘Favoriten’. Fügen Sie zwei neue Titel hinzu. Schließen Sie die Ansicht.

• Finden Sie heraus, welche Spielfilme heute Abend laufen und nehmen Sie einen davon auf. Schließen Sie die Ansicht.

• Zappen Sie durch die Radiosender. Reduzieren Sie die Lautstärke. Schalten Sie einen Sender weiter. Schalten Sie nun den Ton aus.

• Lassen Sie sich die Liste Ihrer mp3s anzeigen. Navigieren Sie zu den Alben eines Interpreten.

• Lassen Sie sich jetzt die Titel eines Albums anzeigen. Spielen Sie ein Lied ab und beenden es dann.
Questionnaires
Sehr geehrte Teilnehmerin, sehr geehrter Teilnehmer,

Vielen Dank, dass Sie sich die Zeit für diesen Versuch nehmen.

Sie befinden sich in einem Wohnzimmer, das mit einem Multimedia- und Hauskontrollsystem ausgestattet ist. Für dieses System möchten wir mit Ihrer Hilfe drei neuartige Eingabevarianten (Spracheingabe, Steuerung über ein iPhone und die Kombination aus beidem) mit der traditionellen Bedienung (Lichtschalter, Fernbedienung) vergleichen.

Im Folgenden werden Sie jeweils ein kurzes Video zur Erläuterung der einzelnen Eingabeformate sehen und bekommen dann die Möglichkeit, eine Aufgabe unter Anleitung zu lösen, um mit der entsprechenden Eingabevariante vertraut zu werden.

Anschließend bekommen Sie Gelegenheit, die jeweilige Eingabevariante anhand von einigen Beispielen ausführlich zu testen. Die dabei als Anleitung dienenden Aufgaben werden auf einem Bildschirm rechts vor Ihnen erscheinen. Bitte bearbeiten Sie jeweils die aktuelle Aufgabe und warten dann ab, bis die nächste Aufgabe erscheint.

Nach jedem Durchgang bitten wir Sie, einen Fragebogen zu der jeweiligen Eingabevariante zu beantworten. Im Folgenden sehen Sie zwei Beispiele möglicher Fragen:

<table>
<thead>
<tr>
<th>Wie ist Ihr Gesamteindruck von dem System?</th>
</tr>
</thead>
<tbody>
<tr>
<td>○ Sehr gut ○ Gut ○ Unentschieden ○ Schlecht ○ sehr schlecht</td>
</tr>
</tbody>
</table>

*In diesem Beispiel ist der Gesamteindruck des Teilnehmers sehr gut.*

Bitte geben Sie mit Hilfe der folgenden Wortpaare Ihren Eindruck des Systems wieder.

| einfach ○ ○ ○ ○ ○ ○ ○ kompliziert |

*In diesem Beispiel fand der Teilnehmer das System sehr kompliziert.*
Ablauf des Versuchs:

1. Kurzer Fragebogen
2. Erste Eingabevariante
   a. Erläuterungsvideo
   b. Kurzes Training
   c. 7 Aufgaben
   d. Fragebogen
3. Zweite Eingabevariante (7 Aufgaben)
   a. - d.
4. Dritte Eingabevariante (7 Aufgaben)
   a. - d.
5. Vierte Eingabevariante (7 Aufgaben)
   a. - d.
6. Abschließender Fragebogen

Beurteilen Sie selbstbewusst und bedenken Sie während des gesamten Versuchs: Nicht Sie werden getestet, sondern Sie vergleichen die Eingabevarianten und bewerten diese!
Uns ist Ihre offene Meinung über das System wichtig.

Für die Auswertung des Tests benötigen wir einige Informationen über Sie, die selbstverständlich anonym bleiben. Deshalb beantworten Sie bitte zunächst die einleitenden Fragen.
Bitte unterzeichnen Sie bei Zustimmung noch die Geheimhaltungs- und Zustimmungserklärung auf der folgenden Seite.

Wir können Sie hören. Bei Fragen oder Problemen einfach kurz über das Mikrophon Bescheid sagen.
Sie können den Test jederzeit ohne negative Folgen für Sie abbrechen!

Und nun: Viel Spaß beim Versuch!
Zustimmungs- und Geheimhaltungserklärung des Teilnehmers

Hiermit erkläre ich, dass ich darüber informiert bin, dass die von mir heute zur Verfügung gestellten Informationen und die Videoaufzeichnungen zum Zweck der Analyse und nur zu diesem Zweck gespeichert werden und ich gebe mein Einverständnis dazu. Mir ist zudem bekannt, dass alles, was ich während der heutigen Sitzung gesehen und gehört habe, vertraulich ist und erkläre, dass ich diese Informationen nicht weitergeben werde.

Name: ___________________________________________________
Datum: ___________________________________________________
Unterschrift: ______________________________________________
Allgemeine Angaben

Geschlecht: weiblich männlich

Alter: 

Wie oft benutzen Sie die Nintendo Wii Remote?
nie selten oft

Wie oft benutzen Sie ein iPhone?
nie selten oft
Gesamteindruck

Bitte geben Sie mit Hilfe der folgenden Frage Ihren Eindruck von dem System wieder.

Wie ist Ihr Gesamteindruck von dem System?
○ Sehr schlecht ○ Schlecht ○ Unentschieden ○ Gut ○ Sehr gut

Bitte geben Sie mit Hilfe der folgenden Gegensatzpaare Ihren Eindruck von dem System wieder.

einfach ○ ○ ○ ○ ○ ○ ○ kompliziert
hässlich ○ ○ ○ ○ ○ ○ ○ schön
praktisch ○ ○ ○ ○ ○ ○ ○ unpraktisch
stilvoll ○ ○ ○ ○ ○ ○ ○ stillos
voraussagbar ○ ○ ○ ○ ○ ○ ○ unübersichtlich
minderwertig ○ ○ ○ ○ ○ ○ ○ wertvoll
phantasielos ○ ○ ○ ○ ○ ○ ○ kreativ
gut ○ ○ ○ ○ ○ ○ ○ schlecht
verwirrend ○ ○ ○ ○ ○ ○ ○ übersichtlich
lahm ○ ○ ○ ○ ○ ○ ○ fesselnd
Bitte beantworten Sie jetzt folgende Fragen:

<table>
<thead>
<tr>
<th></th>
<th>Ich denke, ich würde das System gerne häufiger benutzen.</th>
<th>trifft gar nicht zu</th>
<th>trifft wenig zu</th>
<th>trifft teils zu</th>
<th>trifft ziemlich zu</th>
<th>trifft völlig zu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>2</td>
<td>Ich finde das System unnötig komplex.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>3</td>
<td>Ich finde das System war einfach zu benutzen.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>4</td>
<td>Ich denke, ich würde die Unterstützung einer technisch erfahrenen Person brauchen, um in der Lage zu sein, das System zu benutzen.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>5</td>
<td>Ich finde, die verschiedenen Funktionen des Systems sind gut integriert.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>6</td>
<td>Ich finde, es gibt zu viele Widersprüchlichkeiten in dem System.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>7</td>
<td>Ich kann mir vorstellen, dass die meisten Leute das Bedienen des Systems sehr schnell lernen würden.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>8</td>
<td>Ich fand, dass das System sehr umständlich zu bedienen ist.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>9</td>
<td>Ich fühlte mich sehr sicher bei der Benutzung des Systems.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>10</td>
<td>Ich musste zuerst viel lernen, bevor ich mit dem System zu Recht kam.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td></td>
<td>Statement</td>
<td>trifft gar nicht zu</td>
<td>trifft wenig zu</td>
<td>trifft teils zu</td>
<td>trifft ziemlich zu</td>
<td>trifft völlig zu</td>
</tr>
<tr>
<td>---</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>--------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>1</td>
<td>Das System hilft mir dabei, effektiver zu sein.</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
</tr>
<tr>
<td>2</td>
<td>Das System ist unnütz.</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
</tr>
<tr>
<td>3</td>
<td>Das System hilft mir dabei, produktiver zu sein.</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
</tr>
<tr>
<td>4</td>
<td>Das System macht es mir nicht leichter zu erreichen was ich will.</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
</tr>
<tr>
<td>5</td>
<td>Das System gibt mir mehr Kontrolle über Aktivitäten in meinem täglichen Leben.</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
</tr>
<tr>
<td>6</td>
<td>Das System hilft mir nicht dabei, Zeit zu sparen.</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
</tr>
<tr>
<td>7</td>
<td>Das System ist leicht zu benutzen.</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
</tr>
<tr>
<td>8</td>
<td>Das System macht nicht immer, was ich erwarte.</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
</tr>
<tr>
<td>9</td>
<td>Das System ist einfach zu benutzen.</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
</tr>
<tr>
<td>10</td>
<td>Das System ist nicht benutzerfreundlich.</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
</tr>
<tr>
<td>11</td>
<td>Das System benötigt die wenigsten Schritte, um zu erreichen, was ich mit dem System machen möchte.</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
<td>⊗</td>
</tr>
<tr>
<td>Nr.</td>
<td>Aussage</td>
<td>trifft gar nicht zu</td>
<td>trifft wenig zu</td>
<td>trifft teils zu</td>
<td>trifft ziemlich zu</td>
<td>trifft völlig zu</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------------------------------------------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>1</td>
<td>Es gelang mir, das System ohne Nachdenken zu benutzen.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Ich habe erreicht, was ich mit dem System erreichen wollte.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Mir war sofort klar, wie das System funktioniert.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Der Umgang mit dem System erschien mir vertraut.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Bei der Benutzung des Systems sind keine Probleme aufgetreten.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Die Systembenutzung war unkompliziert.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Es gelang mir, meine Ziele so zu erreichen, wie ich es mir vorgestellt habe.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Es fiel mir von Anfang an leicht, das System zu benutzen.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Mir war immer klar, was ich tun musste, um das System zu benutzen.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Die Benutzung des Systems verlief reibungslos.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Ich musste mich kaum auf die Benutzung des Systems konzentrieren.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Das System hat mich dabei unterstützt, meine Ziele vollständig zu erreichen.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Die Benutzung des Systems war mir auf Anhieb klar.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Ich tat immer automatisch das Richtige, um mein Ziel zu erreichen.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hilfreich</th>
<th>○</th>
<th>○</th>
<th>○</th>
<th>○</th>
<th>○</th>
<th>○</th>
<th>○</th>
<th>nicht hilfreich</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinnlos</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>sinnvoll</td>
</tr>
<tr>
<td>Förderlich</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>hinderlich</td>
</tr>
<tr>
<td>Nutzlos</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>nützlich</td>
</tr>
<tr>
<td>Destruktiv</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>konstruktiv</td>
</tr>
<tr>
<td>Unkompetent</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>kompetent</td>
</tr>
<tr>
<td>Verständlich</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>verständlich</td>
</tr>
<tr>
<td>Sinnvoll</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>unsinnig</td>
</tr>
<tr>
<td>Wertlos</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>wertvoll</td>
</tr>
<tr>
<td>Vernünftig</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>unvernünftig</td>
</tr>
<tr>
<td>Inkonsistent</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>konsistent</td>
</tr>
<tr>
<td>Unbrauchbar</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>brauchbar</td>
</tr>
<tr>
<td>Hemmend</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>unterstützend</td>
</tr>
<tr>
<td>Unbefriedigend</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>befriedigend</td>
</tr>
</tbody>
</table>
Bitte vergleichen Sie jetzt die vier Systemvarianten indem Sie den Varianten Schulnoten (1 - sehr gut bis 6 - sehr schlecht) zuordnen.

<table>
<thead>
<tr>
<th>Systemvariante</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditionelle Eingabe über Lichtschalter und Fernbedienung</td>
<td></td>
</tr>
<tr>
<td>Spracheingabe</td>
<td></td>
</tr>
<tr>
<td>Eingabe mittels Gesten und GUI über das iPhone</td>
<td></td>
</tr>
<tr>
<td>Multimodale Eingabe</td>
<td></td>
</tr>
</tbody>
</table>
Erklärung der Urheberschaft

Ich versichere, dass ich die vorstehende Arbeit selbstständig und ohne fremde Hilfe angefertigt und mich anderer als der im beigefügten Verzeichnis angegebenen Hilfsmittel nicht bedient habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht.

Ort, Datum

Unterschrift