LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Department "Institut für Informatik"
Lehr- und Forschungseinheit Medieninformatik
Prof. Dr. Florian Alt

Bachelorarbeit

Context-sensitive Modalities for Interaction with Large Screens

Melanie Gail
melanie.gail@campus.lmu.de

Bearbeitungszeitraum:	1. 6. 2014 bis 06. 11. 2014
Betreuer:	Prof. Dr. Florian Alt, Prof. Dr. Andreas Bulling
Verantw. Hochschullehrer:	Prof. Dr. Florian Alt

Zusammenfassung

In der heutigen Zeit existiert eine große Vielfalt an verschiedenen Eingabegeräten für interaktive Systeme. Im Bereich der Public Displays gibt es deshalb viele verschiedene Möglichkeiten diese Eingabegeräte zu nutzen, vor allem im Hinblick auf den meist großen, nutzbaren Platz vor diesen Displays. In privaten Einsatzmöglichkeiten von interaktiven Displays ist häufig nur ein begrenzter Raum vorhanden.
Um den kompletten Bereich vor einem Display zur Interaktion bereit zu stellen, wird in dieser Arbeit das Konzept von multimodalen Public Displays in Verbindung mit Proxemics vorgestellt. Zu diesem Zweck wurde ein multimodales System enwickelt. Es besteht aus den zwei Sensoren für Gerstenerkennung, dem Leap Motion Controller und der Microsoft Kinect for Windows, einem Touchscreen und einer herkömmlichen Maus. Durch die Kombination der Geräte kann das System aus verschiedenen Distanzen zum Display bedient werden. Das implementierte System wurde anschließend durch zwei Studien untersucht. Eine der beiden Studien stellt ein abgewandeltes Fitts' Law Experiment dar.
Ergebnis der Studien war zum einen, dass die ausgewählten Geräte enorm in Eingabegeschwindigkeit und Beliebtheit variieren. Zum anderen wurde festgestellt, dass die beiden Sensoren für Gerstenerkennung in einem nur sehr begrenzten Rahmen fehlerfrei zu bedienen sind.
Die Entwicklung von multimodalen Systemen in öffentlichen Plätzen kann den möglichen Interaktionsraum enorm vergrößern um den bereitgestellten Platz bestmöglich auszufüllen. Bevor ein solches System entwickelt wird, sollten die einzelnen Geräte anhand von Eingabegeschwindigkeit, Brauchbarkeit und der bestmöglichen Interaktionsdistanz untersucht und verglichen werden. Die Ergebnisse einer solchen Untersuchung können helfen jedes Eingabegerät in seiner besten Interaktionsdistanz und im besten Einsatzbereich und einzusetzen.

Abstract

Nowadays, a large variety of different input devices exists. In the field of public displays, the number of possibilities to use this devices is basically limitless. In purchase of the greater appropriable area in front of the display, the possible purposes become even more. In private places, the area for interactive systems is often limited to a small extend.
To provide interactiveness in the whole area in front of a Public Display, the concept of multimodal public displays in combination with proxemics is introduced in this thesis. For this reason, a multimodal system with two gesture recognizer, a Leap Motion and a Microsoft Kinect for Windows sensor, a touch display and a mouse was implemented. The system works therefore in different distances in front of the display. Furthermore, the system was evaluated trough two studies, containing a adjusted Fitts' Law experiment.
Summed up, the results of the evaluation are, that the selected devices vary enormously in input speed and popularity. Additionally it was proven that both tested gesture recogniser are very limited in their range in which they works the best.
Bring various devices together can enlarge the possible interaction area in front of a display to fit the available place. For this purpose, input speed, usability and best working area of the devices should be tested before implementing a multimodal public display. The received results of such a study can help bringing each input device in its best interaction distance and field of application.

Aufgabenstellung

Topic: Context-Sensitive Modalities for Interaction with Large Screens

Falling hardware prices in the past years led to interactive displays quickly proliferating in public space. Such displays provide access to information (floor planes, time tables, information on exhibits) as well as interactive experiences (games and other playful applications). At the same time, applications are usually tailored toward one particular sensor, for example a touch screen or a depth sensor allowing for gesture-based interaction. As multiple sensors are available at a display, we envision more multimodal applications to emerge that leverage this variety of sensors for novel interaction concepts and a combination of both coarse and fine grained interaction. For example, applications could enable interaction based on body posture from a distance, using depth sensors or high-resolution camera. As the user approaches, more fine-grained interaction may become available using mid-air body and finger gestures as well as touch and gaze.
As of today, very few examples exist, where several interaction modalities are combined and many challenges yet remain unaddressed. These include, for example, how it can be communicated to the user that different techniques are available, how to realize the "hand-over" between different interaction techniques, and how the use of multiple techniques impacts on user performance. The aim of this thesis is to implement a research prototype that allows these challenges to be addressed. At the focus, the student is expected to implement an application that can be controlled using a combination of different interaction modalities.

Tasks:
* Comprehensive literature review
* Development of interaction scenarios for multimodal interaction with public displays
* Identifying suitable interaction techniques and technologies
* Implementation of a multimodal interactive public display application
* Evaluation of different techniques with regard to performance

Requirements:
* Interest in novel interaction techniques
* Independent scientific work and creative problem solving
* Experience with Kinect / LEAP Motion programming is a plus

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, October 15, 2015

. . . .

Contents

1 Introduction

In this first chapter, the underlying motivation of this thesis is explained. Based on the motivation, the fundamental questions of this work are defined. Afterwards, the primary goals of the thesis are formulated and the outline of the resulting work is presented.

1.1 Motivation

Nowadays, public displays or large screens can be found in every major city. One comes across these displays on the street, in shop windows, as information displays in museums, zoos, at sights or other well attended places. Most of them are still not interactive and only used as advertising or information panel with steadily switching content. But also interactive systems can be found few and far between on the streets, even if it is just a digital vending machine. As interaction with digital devices in public places rises steadily, it should be carried on as an important field of research.

In the field of Public Interactive Displays or also Large Screens, new techniques allow an increasing flexibility in designing novel systems for public displays. This does not only mean new display techniques, permanently new input devices for computer systems are invented, varying in their modality. They go far beyond the known modalities as mouse and touch. It is nowadays possible to interact with a computer system by gestures, facial expression, eye tracking and a lot more. These new possibilities should be fully used to facilitate human computer interaction. All the input capabilities have their advantages and disadvantages, working best in different situations or locations for different users. Various devices or modalities should be put together to combine the advantages of more than one device in an application and at the same time compensate the disadvantages of the modalities. This systems could be used by everyone. The displays can be adapted that even persons under a disability can interact with them without any problems or limitations. Despite all advantages of combining different input devices to a multimodal system, multimodal public displays are rarely found on the streets.

Under consideration of Proxemic dimensions as measurement of the devices cooperation, even more facilitated systems of input devices can be developed. But before putting various modalities together to a multimodal system, it requires closer examination how to develop such systems and therefore how to choose the right devices in the right combination. In this work, a concept of bringing various modalities together to one multimodal system is developed and tested in a study. Selected devices are compared to find their advantages and fields of application.

1.2 Central Question

The central question of this thesis is how to combine different input devices, covering various input modalities, to a public multimodal interactive display. In doing so, current fields of research in combination with public displays, like proxemics, should be under consideration.

1.3 Goals

The main goal of this thesis is to find an optimal combination of different input devices for large screens to a multimodal system, for the use in public spaces. But only a section of this field can be analysed in this work. Further ideas for future work are presented in Chapter 6.4. This thesis is reduced to a concept of comparing different eligible devices to bring them to an optimal combination of the selected devices.

To reach the goal, it is subdivided in different parts. Firstly, it should be made out on which characteristics such a comparison can be made and what the results of such a comparison mean for developing multimodal public displays. Therefore, a multimodal application or game with selected devices should be developed. Afterwards, this application should be used for a comparison of the input devices with regard to input speed and accuracy.

1.4 Outline

To work on this theme, several fields of research are combined and need to be introduced in the first chapter. It includes introductions in public displays, proxemics, Fitts' law and multimodal applications. Each field is shortly described and related work to this thesis are summarized. Afterwards, the research questions are defined based on the presented goals. The problems in this research field will be described and a concept to solve this issues by answering the questioning will be presented. In Chapter 3, the software which is only developed for this thesis will be presented. It contains four different devices, combined to a multimodal system. The chapter contains a detailed description of the implemented software by amplify each device partially. After implementing the software, studies will be presented by applying the software. The whole evaluation has been split in pilot and laboratory study, each one is presented with its design, setup, subjects and detailed results. Afterwards, the received results need to be discussed. In a critical reflection it will be taken to account if all research questions has been answered sufficient. The future work in this field and further planned work will be introduced in Chapter 6, based on the limitations of this work. Significant contribution and an overall summary will be presented in Chapter 7. Finally, a conclusion from the author will be drawn in the last subsection.

2 Background and Related Work

2.1 Public Displays

As the main topic of this thesis is based in the field of public displays, this issue needs introduced first.

2.1.1 Definition

Lately, a lot of researches discovered the field of large screens in urban environment, also known as public displays. They appeared therefore in cities around the world. Their purpose is widely spread. While most displays are not interactive and only used for advertisement purpose e.g. in shop windows or as information screen at sights or other tourist attractions, some are even interactive. They should lead passerby to an interaction and enable the possibility to choose between different provided applications or information topics. The interaction allows user and designer a wide range of fields of application. Combined with new technologies being developed all the time, the possibilities appear to be limitless. In consideration of the different input possibilities, there exist several kinds of displays. With one type, users are able to interact with touch input, another kind of displays can be controlled trough mobile phones. Last but not least there are displays that only can be interacted with keeping a special distance.

2.1.2 Related Work

As public displays open a wide range of possible use, a lots of publications in this field exist. Concepts were made how to design those installations the best. While most of them describe only one concept, Müller et al. described general requirements and design space for interactive public displays [1]. They didn't only take one concept under consideration, but described different mental models of public displays, which contain poster, window, mirror and overlay, methods to attract the passerby attention and an overview of different input modalities. Further, they recommend a designer to choose a mental model best matching his goals, one or more input modalities and a type of supported interaction when designing public displays.
Alt et al. [2] created a summary of frequently used evaluation methods in the field of public displays and discussed them in detail. Additionally they provide several guidelines for researches to evaluate Public Displays appropriately.
As people are hesitant when they pass by a public display, the Optionizer system was developed by Brignull et al. [3] to "encourage socializing and interaction" [3] with a public display.

2.2 Proxemics

As mentioned before, the research field of proxemics should be taken under consideration in this thesis.

2.2.1 Definition

The term proxemics was mainly formed by the US cultural anthropologist Edward Twitchell Hall in his book "The hidden dimensions" in 1966. According to Hall, proxemics are a non-verbal communication between humans. The personal space of an individual can be divided in four sections with different meanings. The closest one is the intimate distance located nearer than 46 cm to a person. The personal distance is located between 46 cm and 122 cm. The social distance follows between 122 cm and 370 cm. The last section is the public distance with an distance further away than 370 cm from the person. Every distance is again split in close and far phases. As the names indicate, the first area is primarily used for intimate interaction as touching or whispering.

For interaction with loved ones, the personal distance is used, whereas the third distance is used for interaction with every other person. The public distance is used for public speaking. He further describes that the arrangement of furniture in rooms, consisting of immovable (fixed features) and movable (semi-fixed features) objects, has an influence on the perception of people and their interaction in personal space [18].

Proxemics in HCI For humans, using proxemic interaction is typically an implicit communication. When using proxemics in Human-Computer-Interaction, systems have to be enabled to recognize proxemic relations between entities. Therefore, the distance as only parameter is not meaningful enough. A fine grained knowledge about entities in a space and their relationships to each other should rather be taken into account. The proxemic dimensions were therefore adjusted and extended by Ballendat [19] to position, orientation, movement and identity. The position describes the distance of an entity from a certain point in the space. It can be described in absolute or relative terms. For an absolute term, the position of the entity is described by three dimensional coordinates. The relative position can be described by relationships between several entities. Orientation informs about which direction the entity is facing. For this variable the entity has to have a defined front side, like a human has. Movement describes the changes of position and orientation over time. Unique identities describe the entities in the space. Entities can be people, digital devices like large screen or portable devices and non-digital devices like furniture.

Greenberg [23] also developed and defined proxemic dimensions for interaction. His version consists of orientation, distance, motion, identity, and location. orientation, identity and motion (movement) are the same is in Ballendats definition. Ballendats dimension of position is redefined as distance, which is now defined as the absolute distance between any entity. Location describes the setup of fixed and semi-fixed features.

2.2.2 Related Work

Hello.Wall [24] "is a wall-sized ambient display that emits information via light patterns and is considered informative art." [24]. Interaction is possible trough a mobile device. The information shown on the wall and on the mobile device changes based on the users distance from the wall. Therefore, the place in front of the screen is divided in three sections. Hello.Wall can be seen as a pioneer for using proxemics in HCI.

Vogel [22] developed design principles for public ambient displays. His framework consists of four regions in front of a public display, which can be seen in Figure 2.1. The ambient display phase shows general categorized information for passerby. As soon as a person passes by, the screen shifts to the second phase, the implicit interaction phase. In this phase, the display tries to get the user closer to the screen by implicit reaction based on orientation and position of the user. When the user wants to start an interaction, the display gets into the subtle interaction phase in which more detailed information is displayed. The user is now able to interact explicit by simple actions. If a user selects an item, the display switches to the personal interaction phase, where touch input is possible. Users can leave the interaction promptly at every state.

The Proxemic Peddler is a public advertising display that reacts on people in front of it, based on different attentional states of the person. The system uses therefore the proxemic dimension identity, position and orientation [20].
The Proximity Toolkit [21] was developed to simplify the implementation of proximity-aware computer systems. SpiderEyes is a collaborate environment to simplify the analysis of huge multi-faceted datasets. The data is therefore shown on a large display, and can be analysed trough

interaction which is aware of proximity by multiple users. Dostal et al. [26] explored furthermore several modification of the connection of users position and zooming display content.

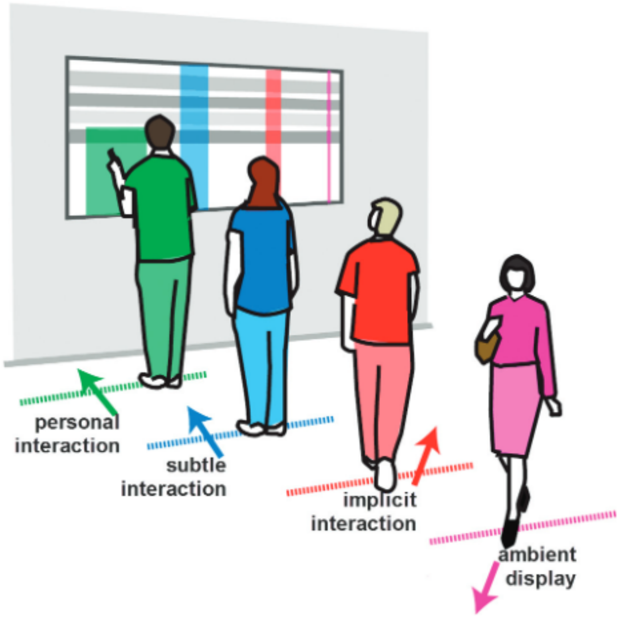

Figure 2.1: Interaction zones of public ambient displays after Vogel [22].

2.3 Multimodal Interaction

When talking about proximity-aware interaction, a system usually can't get along with only one input device. Therefore, the theme of multimodal interaction needs to be introduced.

2.3.1 Definition

Multimodal interaction can be defined as the "interaction with the virtual and physical environment through natural modes of communication such as speech, body gestures, handwriting, graphics or gaze." [25]. This is applicable for all computer systems, requiring any form of input or output for its users. Multimodal interaction should provide a more natural environment for the user trough combining several input or output techniques. Multimodal systems exist in every field of systems controlled trough a user. For example, every smartphone nowadays provides touch and additionally speech as input. Multimodal implementations can either provide the simultaneously use of more than one modality or the exclusively use of one modality with the possibility to choose a favour from a selection.

2.3.2 Related Word

Müller et al. [1] introduced several modalities for the interaction with public displays: presence, body position, body posture, facial expression, gaze, speech, gesture, remote control, keys and touch. All mentioned modalities allow their use for both implicit and explicit interaction and are obviously only a selection of all possibilities.

Bourget [25] defines in her work a simple mechanism for modelling multimodal systems, she proposes to use Finite State Machines (FMS) and introduces her FSM-based toolkit for "designing, implementing and testing multimodal commands" [25]. Raisamo [30] introduced one of the first multimodal interfaces for the use in public spaces. Johnston and Bangalore introduced another multimodal interface for the use in public. Their MATCHKiosk provides the use of an interactive kiosk with "speech, handwriting, touch, or composite multimodal commands combining multiple different modes" [29]. Gaze is often combined with other modalities, as e.g Stellmach et al [27] used the combination of gaze, a keyboard and a mobile multitouch device to eliminate dwell-time and solve the problem of Midas Touch ("unintentionally issuing an action via gaze" [27]). Gaze can also be used in public spaces, as it enables an interaction from a distance or with content out of reach. Turner et al. [28] combined gaze with manual input techniques to enable interaction with objects out of reach.

2.4 Fitts' Law

To compare different devices based on their input time and accuracy, Fitts' law is a commonly used law in Human-Computer-Interaction and is therefore now introduced.

2.4.1 Definition

Fitts' Law was proposed 1954 by Paul Fitts [31]. It is a model for predicting movement time in a pointing task. According to Fitts, the time (MT) a human needs to point rapidly on an object can be determined by the objects size (W) and distance (A) to the current position of the pointing hand. The formula was introduced by Fitts as

$$MT = a + b\log_2(2A/W). \tag{1}$$

Where a and b are constants determined by linear regression, the log term is also known as the index of difficulty (ID).

Several different formulations of this index do exist. The variation proposed by Welford [38] uses the formula

$$MT = a + b\log_2(2A/W + 0.5), \tag{2}$$

whereas MacKenzie proposed his modification [33] as

$$MT = a + b\log_2(2A/W + 1) \tag{3}$$

which is also known as the Shannon formulation.

An important parameter for comparing different Fitts' law results is the throughput (TP), also known as index of performance (IP). Like the ID, many several definitions exist for the throughput. Even Fitts himself defined two versions in the course of the years. The initial version was defined as

$$IP(i) = ID(i)/MT(i) \tag{4}$$

and is dependent of a i*th* point of the regression line [31]. Later he redefined the index of performance as

$$IP = 1/b, \tag{5}$$

where b is the slope received from linear regression [32].

2.4.2 Related Work

Fitts' Law Two-Dimensional In Paul Fitts' experiments, the participants only had to move on a horizontal axis for selection. Amplitude and width of the target were both measured along the same axis and the model is therefore one-dimensional. For applying Fitts' law on modern computing systems, it has to be extended to two-dimensional. For this reason, Scott MacKenzie and William Buxton proposed an extension of Fitts' law in 1992 [38]. For amplitudes disparate 0, the ID value can be negative. To avoid this, Mac Kenzie et al. recommends to use the Shannon formulation as seen in formula 3. Further benefits of this formulation after Mac Kenzie are "the slightly better fit with observation" and it "exactly mimics the information theorem underlying Fitts' law" [38, page 2]. Mac Kenzie introduced several different models of defining the target width in two dimensional tasks. He stated the W' and the smaller-of model as being the best choice. In the W' model, target width is "the extend of the target along an approach vector through the centre" [38, page 13], while in the smaller-of model simply the smaller value of width and height is taken.

Further Work The amount of Fitts' law studies in HCI is enormous. It is therefore one of the most used laws in HCI. It was not only adjusted to 2D as mentioned, but also to 3D [34]. It is commonly used to compare input devices as in [41]. Even an ISO Standard ISO 9241-9 to characterize the performance of different input devices exists. This standard is based on the throughput. As being such a common tool in research, several guidelines exist of how to execute a Fitts' law study.

In 2007, O. Chapuis conducted a Fitts' law study [35] with 24 users over several months. He collected overall two million movements. For this purpose he introduces the length-distance index (LDI) to consider the fact that fewest movements are executed on a straight line.

Weichert et al. [36] conducted a Fitts' law study testing the performance of the Leap Motion sensor. However this study was not conducted with persons as subjects, it was executed by robotics. It therefore can not be compared to studies conducted by humans.

Sambrooks [41] evaluated and compared mouse, touch and Kinect trough a Fitts' law study. He used gestured as selection task for Kinect, a mouse click for the mouse and a touch for touch input. He received poor gesture performance and explains this trough the limitation of the Kinect sensor and peoples precognition in gesture interaction.

There is not only enthusiasm for Fitts' law, but also criticism about its appropriation. Recently, Drewes criticised the overestimation of the law and the argumentations about the right modification. He advices to use the original formulation of movement time by Fitts himself [42].

3 Problem Description

3.1 Scenarios

Scenario 1 Sarah is waiting at the central station for her friend Alexandra to arrive. While walking bored up and down, a large screen a few meters in front of her starts to blink and shows the message "Bored? Want to play a game?". Additionally, there is a little picture of Sarah's silhouette displayed. As Sarah notices her silhouette she is working up curiosity. Sarah is turning toward the display which is now showing a horizontal scroll bar containing several games. A blinking hand appears on the display showing Sarah the possibility to interact with her hands. As moving her hand up, a cursor appears and Sarah is able to choose "Whack a Mole" trough a simple circle gesture. The screen is now containing an illustration demonstrating Sarah to come closer and try to interact through moving her hand over a black bar straight under the display. As Sarah starts to do so, the game begins and she starts to play. After several minutes Sarah quits the game. Multiple buttons are showing up and Sarah gets the chance to save her score and leave a comment about the game. She attempts to select the button through the gestural input as she used to play the game. But when she is moving the cursor across the "Leave a Comment Button", a little picture showing a finger touching a display appears. Sarah is touching the button and the screen reacts, a virtual keyboard shows up and Sarah enters a short comment. During Sarah is walking away, the display changes to its initial situation and waits for the next user.

Scenario 2 John wants to buy tickets for his favourite rock band. A new 24-hour digital ticket kiosk was recently opened in his city. He would like to try this new service but does not know how to handle the new machine. As he is walking towards the vending machine, which looks like a large tv combined with a ticket machine, the next concerts in his city are appearing on the screen. John is detecting his favourite band among them. As he is getting closer and is looking on the concert poster and ticket price, a pop up is appearing with further information, like free seats in the concert hall. John gets closer as he wants to interact with the machine, till he stands in a distance of 1 m in front of the display. A bar on the right side of the screen is showing him trough pictures that he can interact with usual touch or gestures without touching the screen. Not only one but two different interactions with gestures is possible. The screen indicates the possibility to interact closely to the screen with finger gestures and the possibility to interact with hand gestures from-a-distance. Since John has incredible fear of germs and doesn't want to go backwards, he is deciding to try the finger gestures. As soon as he is lifting up his hand, a little animation appears on the screen, showing the best way to interact in detail. John has now the possibility to try the interaction by a little game moving a ball. Once he is secure enough he stops the training and moves back to the previous screen where he started to interact. He is now able to choose the four seats he needs and is buying them. The bar on the right side steadily shows that a change on the modality is possible at any time. After the tickets are print out, John is leaving the display without further input. The display recognizes that the costumer has left and gets back to is initial situation. As John walks across the display a hour later, he recognizes a woman interacting with the touch function.

3.2 Research Questions

Based on the current state of research and the defined goals of this thesis, five research questions are defined:

- How can multimodality be combined with proxemics in public displays?
- How can the area in front of the display be used to create a gapless interaction area? Is proximity useful to achieve this goal?

- Based on the selected devices: Which device can be used the best at which distance to the screen and how affects granularity of the input device on the quality of interaction in different distances?
- Is there a difference in peoples' favourite input device?
- Can new input techniques be brought in public spaces or are they to unfamiliar for people to use?

3.3 Combining Multimodality with Proxemics

When thinking about combining both concepts multimodality and proxemics, it came to mind that the proxemic dimensions orientation, distance, motion, identity and location may be useful as parameter for changing the modality. For this reason, each dimension is now presented with ideas how to match them with multimodality approaches. The definition of Greenberg [23] is used for further consideration.

Orientation Orientation describes in which direction an entity is directed. It can be used to determine the persons centre of attention and to attract attention if the passerby looks towards the screen. There is not a large variety of feasible states of orientation, merely it can be divided into the user being orientated towards the screen and being not. When the orientation state changes from orientated towards the screen to not orientated towards the screen, the system only can try to make the display attractive again, e.g. by changing the output modality to sound, as the user won't see a change on the display content. Changing the input modality based on changes in users orientation is hard to imagine but surely possible.

Distance The proxemic dimension distance is very versatile in use. The distance describes the absolute space between two entities in the scene. It can be used to determine how interested a passerby is trough his distance to the screen, whereby orientation should be regarded additively. Additionally, the distance a passerby is away from the screen can be used to adjust the displayed content. The content should be detailed when a user is near the screen and more general if a user is distantly. The space between users in the scene can indicate relations between them, e.g. if the people are a related group or single user. The distance appears to be a suitable parameter on which the active modality could be chosen. By a change of the current distance, the current modality could be adjusted to match every devices perfect interaction distance to one installation.

Motion The dimension motion implies that distance, location or orientation of an entity has changed. Thus the motion is the actual dimension on which the modality switch is based on in case of distance, location and orientation. A motion can hence be used for changing the modality when a passerby occurs or leaves the scene, if a user changes the distance or orientation towards the screen. Another case for a motion based switch of the modality would be a explicit motion of the user which causes an explicit choice of an input device.

Identity Identity is an unique identifier for entities in the area in front of the screen. It normally does not change during an interaction. Identity as a parameter for changing the modality is therefore hard to imagine but not impossible. It could be made possible trough, e.g. a person is assigned to a unique identifier trough log-in or face recognition. Every identified person can choose and save one favourite modality, whenever a person is recognized by the system, it switches to the saved modality. In this case, the system would not be multimodal within the interaction of one user.

Location The location is defined as the setup of fixed and semi-fixed entities in the area in front of the display. As fixed objects like furniture don't change in general, their locations use as parameter appears to be difficult. Whereas the relationships between fixed and semi-fixed entities can change, like person A sits down on chair C and the screen switches to a modality providing seated interaction. Another use of location is possible when a group of people wants to interact at the same time. The system should now enable only modalities which allow the simultaneous input of various users. In this case, groups with three people or more should interact with modalities enabling interaction from-a-distance, since all users need to have a look on the display.

The proposed ideas are obviously not all fields of application as combining multimodality with proxemic has a wide range of opportunities. The suggestions should only demonstrate this wide range received from the combination of proxemic and multimodality.

As considering all dimensions would exceed the scope of this thesis, only the distance between a user and the display is regarded in the following course of the thesis. Furthermore, the distance seems to be the most multifaceted proxemic dimension for the use as parameter for changing the modality.

3.4 Interaction Distance

The distance a person is away from the screen is a significant parameter especially in public spaces as interaction locations and the available place in front of the screen are widely differing. The interaction distance to a computer system is mostly not changed in interactions at private places or even variable at all. Interactions at private places are rather the same and repeated again and again by the same few people. The diversity of users possibilities of interaction in public spaces changes this issue. Besides, for most public displays, a larger area in front of it is available than for any device at private places.
As mentioned before, the definition of proxemic dimension from Paul Greenberg [23] is used in this thesis. In his definition, location and distance were in two dimensions. The distance is therefore the space between two entities in the regarded area. For this thesis, the distance is defined as the distance between the screen with its input devices and the primary user, further on referred to as interaction distance. In combination with proxemic dimensions, interaction distance is a commonly used parameter for interaction. It can be used to determine the presence of a person which can in turn be used for implicit interaction as the screen reacts on the presence or be an factor for determining how interested a user is in an interaction. Explicit interaction is another way to use it, e.g. when the user is told to come closer if he wants to interact.
As the interaction distance can change rapidly during an interaction, it is further used as parameter for the current input device. Due to the fact that the possible interaction distance varies for every device or modality, the possible distances for each device should be considered by developing multimodal public displays. Additionally, it would be desirable to combine the devices so that the area in front of the screen enables interaction at every distance and create a gapless interaction area.

3.5 Devices for multimodal Interaction

In this thesis, a system will be implemented, resulting in a multimodal system in which the single devices are only provided for its single and explicit use. As from now on only the distance between user and display is considered as parameter for the use of different modalities, a closer look should be made which modalities work in which distance or if even a specification on the possible interaction distance can be made. The different modalities are chosen from the work of Müller et al. [1]. It is assumed that every device of the modalities is placed directly at the screen.

Presence The modality presence states if a user is in the tracked area of the system. It is typically used for implicit interaction in combination with public displays to attract passerby attention. Possible distance for tracking the presence is the possible tracking area the used sensor provides. Presence is the furthermost possible interaction, as it is enabled as soon as a user is in visual range.

Body Position Body position does not only identify presence but also the exact position of a person in front of the display. It is used like the presence as implicit interaction, by adjusting the displays content based on the persons coordinates. The position of a user can be determined in the range the used sensor works.

Body Posture Body posture wants to determine the proxemic dimension orientation. The passerby orientation is mostly used as implicit interaction in the possible tracking area.

Facial Expression Special software and hardware is able to determine the facial expression of a person to adjust the screens content based on the user's mood. As presence, body position and body posture, facial expression is mostly used as an implicit input and only possible in the range of the used hardware and software.

Gaze Gaze can be used in different ways. The direction of the user's gaze can be used the same way as body posture. Eye-tracking allows to follow the users path of gaze. Therefore, eye-tracking can even be used as explicit input device. Eye-tracker exist as mobile devices like glasses or helmets, but also as remote eye tracker which doesn't need any device on the users body. Remote eye tracker usually need a calibration. The possible interaction distance must be separated by the different techniques.

Speech Speech can be used for input and output. For implicit input, conversations can be sensed for keywords, to change the display's content based on the conversations taking place around the display. Based on different voices, speech recognition can additionally estimate how many people are located near the display. Furthermore, speech can of course be used as an explicit input, which is provided nowadays even by smartphones. A microphone with the right software is used as input device in this case. The possible range of this modality is the range of the used technology.

Gesture Gestures are integrated in many technologies like touch sensors, eye-tracking or the mouse. In combination with public displays, mostly hand gestures are used. Hand gestures are additionally applicable in game consoles. They are particularly suitable for public display since they work without any remote. Care should be taken to avoid systems in which a calibration is needed for every user. There exist several different techniques for gesture recognition. Some of them work only from-a-distance, while others only work directly near the screen.

Remote Control Remote control describes the input with an external wireless connected device. In public spaces, usually the mobile phone represents the remote control, as the remote should be ready for everyone. Remote controls are additionally used in game consoles, like the Nintendo Wii. The possible interaction range is once more the range of the type of connection but is usually primarily used for from-a-distance interactions.

Keys The term keys is understood to represent mouse and keyboard. The two devices are the usual way to interact with a computer. Both devices exist with a cable connection or wireless. Their interaction distance is usually limited to the desk on which they are placed on. But the distance is theoretical only limited by the cables length or the range of the wireless connection.

There exist wireless keyboards and mice with a range greater than 10 m. As keys are common for every computer user, they are optimal for a comparison with other input devices.

Touch Touch interaction is beside mouse and keyboard the mostly used input modality, ever since smartphones, tablets and other mobile touch devices occurred on the market. Touch input is further available for laptops, PC displays, interactive tables and a lot more. There exist several techniques for touch input, the mostly known are resistive and capacitive touch screens.
Devices with touchscreens are also popular in public spaces. Public displays and ticket machines for public transportation are only examples. Touch input requires direct contact by the user and is handled trough touching the screen.
Caused by the needed contact, the interaction distance is limited to the length of the user's arm. The interaciton with touch is therefore the nearest possible interaction with a large screen.

3.6 Realization

In this chapter, it is introduced how the modalities for the later implemented software were selected and which methods are used for further resolution of the presented research questions.

3.6.1 Selection of Modalities and Devices

The system developed in this thesis should contain different input devices to cover several modalities. As the modalities should be switched based on the distance between user and screen, the devices are selected in such a way that the greatest possible range in front of the screen is covered. Furthermore, only explicit input possibilities are considered, as in the first step, the study should take place in a laboratory and not in public spaces.
Presence, position, posture and facial expression aren't easy to use as explicit input modalities, they are hence mostly used as implicit ones and seem to be not suitable for the application in this thesis.

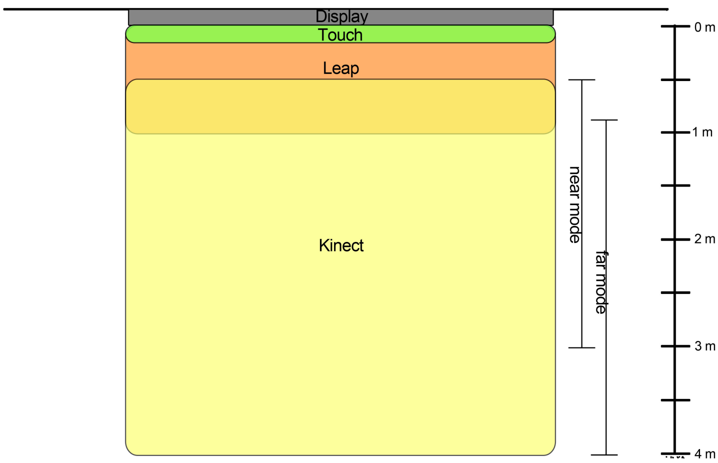

Figure 3.1: Selected modalities with their ranges.

Speech and remote control were removed, because they can be used in every distance and therefore don't need to be replaced trough distance changes. It is possible to use them in

multimodal systems which have supplementary devices the whole input time.
For gesture recognition, a Kinect for Windows and a Leap Motion sensor were provided for this thesis. In the following comparison of the modalities, gesture recognition is replaced by the Kinect and Leap specification. Further, the modalities touch, gaze and keys made it to the final round of selection.
Finally, gaze was removed. Although it would cover the furthest range from the screen, at this distance an interaction would only be possible with eye-tracker glasses, which contradict the concept of public displays for everyone without any required remote. Additionally, Fitts' law does not apply to eye movements [42]. This fact would make a comparison of the selected devices trough Fitts' law impossible.

Keys weren't fully removed from the system. A mouse will be used for purposes of comparison with the other devices. As no text input is necessary, no keyboard will be integrated in the implementation. For further study purposes after the laboratory studies, the mouse should be excluded.

3.6.2 Methods

After selecting the modalities and devices for the implementation, the devices should be evaluated on their possible input distance and if there exists a position in front of the display where the interaction works the best for the respective device. Moreover, the devices should be tested on needed input time, granularity and user experience. The results should be used to compare the devices based on this factors.
The whole evaluation is split in two studies, firstly the possible distances are determined. Afterwards, the implementation should be adjusted if necessary. As second study, a classical Fitts' law study seems to be plausible to compare the devices on the named factors. The original definition of the formula of movement time (see Formula 1) by Paul Fitts is used in this study. Furthermore, the second definition for throughput is used (see Formula 5). User Experience will be captured trough questionnaires and interviews. For determining usability of the input devices, a Mini-AttracDiff[1] questionnaire is used. As some modalities like gesture recognition can be physically demanding and others may be complicated to handle and therefore mentally demanding, a NASA-TLX[2] questionnaire is used additionally to determine physical and mental demand.

[1] http://attrakdiff.de/
[2] http://humansystems.arc.nasa.gov/groups/tlx/

4 Implementation

This chapter contains a description of the implemented system and its single parts. Initially, the concept is described. Afterwards, the implementation for each device and the software on the whole are explained in a detailed way.

4.1 Fitts' Law as a Game

While brainstorming about games for a multimodal public display, the game Whac-A-Mole came into mind. After choosing Fitts' Law as evaluation method, the idea of combining Whac-a-Mole with Fitts' Law came into mind. It seems to be an interesting alternative for Fitts' law studies.

Whac-A-Mole Whac-A-Mole is a simple game available as mechanic implementation or computer game. The original version contains 9 objects (molehills) arranged on a 3 x 3 grid. When the game starts, a mole starts to pop out of his molehill. The user has to hit (whack) it with a gavel to bring it back in its hill. As soon as the mole is hit, the next one pops up. In some variations, several moles can be outside their holes at the same time. When the user can't whack the mole fast enough, the mole disappears. To complicate it, the game gets continuously faster. The game ends after a constant time. Whac-A-Mole exists in various variations as it can be applied to almost everything, e.g. Whac-A-Banker or Whac-A-Monster.

Combining Fitts' Law and Whac-A-Mole Hitting an object as fast as possible is basically the same aim in both, Fitts' law and Whack-A-Mole. Combining them may lead to a more interesting and less boring and monotonous study, through a playfully interpretation of Fitts' law. That's why they are now combined to one implementation.

For matching the idea of Fitts' law, the Whac-A-Mole game has to be slightly adjusted. First of all, the partition in a 3x3 grid has to be removed due to the fact that the distance between to moles should be variable as it represents the amplitude parameter needed for an implementation of Fitts' law. Furthermore the moles can't be similar size, as the size of a molehill represents now the width parameter. To execute a valid Fitts' law study, only one object is allowed to be on the screen simultaneously. On account of this, moles and molehills can't appear separate, a mole and its hill appear and disappear simultaneously in this implementation. To get comparable results for every four devices it is abdicated to use a explicit selection of a mole. Simply moving the cursor above the current mole fulfils the move. The cursor is replaced by an image of a sledge. Finally, a move has to be executed no matter how long it takes. The mole does not disappear after a time limit, like in the original version.

4.2 Devices

The software was implemented in Visual Studio 2013, using C# as programming language. WPF was used to generate the graphical user interface. To show the current active input device, four rectangles on the right side of the screen were built-in. Once a device is active, the appropriate rectangle turns the colour from white to pink. This can be seen in Figure 4.1, where the mouse is active. The different input devices require different implementation and certain libraries. For each device, a function is required to get the corresponding coordinates of the cursor depending on the active device. The particular methods are now presented.

Figure 4.1: Screenshot while playing.

4.2.1 Mouse

As mentioned before, a mouse was used additively for comparison purposes. Therefore, standard Logitech optical USB-Mouse was used for implementation, testing and evaluation.

For implementing the mouse part, the predefined WPF-Event MouseMove was used. When the mouse mode is active, the event gets fired every time the user moves the mouse. The current cursor position can be extracted from the event. The cursor position is continually updated through a thread as long as the mouse is on the move.

4.2.2 Touch

A 23 inch Acer T232HL Full HD LED Touch monitor was used as touch device. The display is able for ten-point multitouch and has a 5 ms reaction time.

Like for mouse input, WPF contains a library for touch input. Therefore, the predefined TouchDown-Event was used. Unlike the other three devices, the cursor can't be updated when the user targets on an object. The updating of the cursor is therefore left out for touch mode. As soon as a TouchDown-Event is fired, the move is performed as long as the next target was hit.

4.2.3 Leap

The Leap Motion is a 3 cm by 8 cm small device released in December 2013. It is equipped with two CCD cameras and three infrared LEDs. The LEDs generates a flashing pattern of IR light dots. The sensor is very sensitive to bright light or dirt, which should be avoided while using it. The cameras stream maximal 290 frames per second trough USB 3.0 with a precision about 0.01 mm. Bringing the light-dots and the camera frames together is enabled by complex maths, implemented in the Leap Motion software. Getting further official specification is not feasible because of patent and property rights. The possible tracked area around the Leap is indicated as approximately 1 x 1 x 1 m above the device. The leap airspace, a software automatic installed by the first leap setup, contains a lots of different software for interaction with the Leap Controller. Mainly it is used for 2D and 3D modelling and design but also games are available. A big computer manufacturer offers actually notebooks with an integrated Leap Controller [39].

The implementation for the Leap Motion is not as easy as for mouse and touch, the standard C# libraries are not sufficient. To get the required coordinates of a pointing finger, a few steps are necessary. Firstly, the official Leap-Developer-Kit is needed, it provides the SDK with a lots of predefined functions as e.g predefined gesture recognition. Additionally, several sample implementations are provided for developers. The Leap sensor captures data as single frames. The SDK makes a class Controller available. By adding a Listener class to the Controller, the access to the frames of the tracked data is possible as the Listener handles all events from the Frame class. 60 previous frames are stored, which can be accessed. A instance of Frame contains ID, Timestamps, Hands, Fingers, Tools and Gestures. The access on the required X, Y, Z is enabled trough the detected fingers, which are available trough the Listener. In this implementation, a custom listener is implemented to enable the access on the event OnFingersRegistered. This event is fired as soon as a finger is detected by the sensor. Furthermore, in this custom listener, it is implemented that the whole calculation is not done every single frame, as this would be a lot of redundant computation. The received coordinates trough the captured fingers need to be normalized to match the display coordinates. The cursor coordinates are updated as soon as a new position of the pointing finger arrives, as long as at least one finger is visible for the sensor. If the sensor can't detect any finger, the cursor disappears.

4.2.4 Kinect

Less resolution but therefore equipped with a much bigger observation area than the Leap Motion is the Microsoft Kinect sensor. Primarily, the sensor was developed for the Microsoft game console XBOX 360 by Microsoft. Later, an advanced sensor called Kinect for Windows was developed and a Kinect-SDK was published to enable developers to implement own solutions with the Kinect. Kinect for Windows enables the simultaneously use up to four Kinect devices and owns a separate near mode, enabling an interaction from 500 mm to 3000 mm afar the sensor. The default mode enables detection of from 800 mm to 4000 mm away the sensor [40].

A Kinect sensor is equipped with an IR emitter, a colour sensor, an IR depth sensor and four microphones for tracking. The IR emitter projects a pattern across the room using a laser diode. This pattern is captured by the IR depth sensor and a chip inside the Kinect analyses that data to depth data. The colour sensor is a normal camera which is recording the scene.

Microsoft provides a Kinect for developers toolkit, which includes a lot of samples for speech and gesture recognition applications of all kinds. When starting programming with the Kinect, the sensor has to be initialized first. Afterwards, the different streams can be started. The sensor provides three data streams: A SkeletonStream containing the tracked users as skeletondata, a DepthImageStream containing the depth-images and a ColorImageStream containing the pictures taken by the colour sensor. The SDK provides further the class InteractionStream which enables to map the hand position of a user to a cursor on the screen. Access to the required coordinates is therefore possible trough the InteractionStream. Trough this implementation, users are able to use their favourite arm and can even change the arm during interaction.

4.3 Game Logic and Debug Mode

A variation Model-View-ViewModel MVVM was used as design pattern, which is mostly used in combination with WPF. After the integration of all four devices was shown, the game logic and the debug mode for study purposes are now presented.

The game logic is as simple as the game is to play. At every time, only one mole object exists. Only X and Y coordinates and the size of the mole needs to be adjusted after every move. A constant array with 16 combinations for amplitude and size exists. The array is randomly ordered after every 16 executed moves. The X and Y coordinates for the next move are determined by a pair of amplitude and width and a random angle (from 45°,90°,135°,180°,225°,270°,315°,360°). If the determined coordinates are outside the screen, the next angle is taken. As soon as a new mole appears on the screen, the logging time starts. Time ends as soon as the cursors centre hits the mole. Logging files contain width, amplitude, angle and required time for every move. A new file is created for every input device. The four files of one user are stored in a separate folder.

A debug mode is integrated to relieve study procedures. It enables switching between all single devices mouse, touch, Kinect, Leap. Additionally a training-mode which enables all devices simultaneously, resulting in a multimodal-mode, is implemented. Furthermore, basic functions like pause, start new and end are built-in. The debug mode enables the examiner to have better control of the study procedure. If the debug mode is opened, tracking and logging is paused. If a device is chosen, the other devices stop tracking data to exclude the influence on the results.

5 Evaluation

The evaluation was split in two parts. Firstly, a pilot study with only a few subjects was conducted to exclude the influence of several factors on later studies. Secondly, a larger laboratory study with fifteen people was arranged to determine the performance and usability of the selected devices. The design for each study, which is containing details about subjects, setup and procedure, and the results are presented.

5.1 Pilot Study

First of all, a Pilot study was conducted. Its aim was to determine if the manufacturer specifications in terms of the distance correspond to reality. Furthermore, it should be made out if there is a distance in which the two gesture recognizer, Kinect and Leap, work the best. With this information it could be excluded that the distance of a user to the device has an effect on the interaction. Additionally, it was tested if the size of the pointer has had an effect on the interaction and if the two devices could be used without previous training or explications.

5.1.1 Design

In this subsection, participants, setup and procedure of the pilot study are defined.

Subjects Four computer science students took part in the first study. There were three females and one male, aged between 20 and 22. Participants were recruited from the authors university. All subjects were regular computer users resulting from their studies. None of the participants have had any previous knowledge about the Leap Controller. One had good precognition about the Kinect, the other three none. For taking part at the study, the subjects got paid.

Setup A Lenovo U410 Ultrabook with a NEC WT615 beamer for output was used as a replacement for a large screen. The beamer was configured to 52 inch. For input, only the Kinect sensor and Leap Motion sensor were used. User positions were marked on the floor. The implemented program was adjusted to match the purpose for this study.

Procedure When a participant entered the room, he firstly was told to try to interact with the Kinect sensor without any explication or further information about positioning in front of the device. If he didn't know how to handle it, the instructor explained how to execute an interaction. Afterwards, subjects had two minutes for trying and getting used to the interaction at a distance of 2 m. Devices were tested one after another. The currently not used sensor was covered to exclude its influence on the other one. This is necessary since both devices produce infrared light pattern for tracking. To test the possible interaction range of the Kinect, the area in front of the sensor was divided in 0.5 m long sections. Subjects tried to interact in every section between 0.5 and 4 m at least 30 seconds. Afterwards, they had to rate the possibility to interact as: not possible, possible but only bad interaction, satisfying interaction, and good interaction. The same procedure was executed with the Leap Motion sensor, but this time with 10 cm long sections between 0 and 50 cm in directions: left, right, forwards, backwards and upwards. Finally, questions about the two devices were asked.

5.1.2 Results

The received results of the conducted pilot study are now presented.

Kinect Microsoft specifies the range in which an interaction is possible from 80 cm to 400 cm away from the sensor, since this is the range in which the depth sensor can track users. [40]. Although the sensor may detect users in this range, the results show that an interaction is not possible at every distance between 80 and 400 cm like . All subjects were able to interact optimal between 1.5 m and 2.5 m. The first interaction was possible at 3.5 m but participants called it troublesome. At any distance further away than 3.5 m no interaction was detected by the sensor. When being closer than 1.5 m, interaction becomes laborious again. Three out of four subjects called the input with the Kinect physically demanding especially when taking a longer input. Interaction was perspicuous for all subjects, mostly even without any precognitions or explanation.

Leap A number of rumours are circulating that the Leap Motion works in an 1 x 1 x 1 m large hemisphere above the device. In this study, an optimal interaction with the Leap Motion sensor was only practicable when the subjects placed their hands directly above the device or maximal 10 cm away at one side. Everything farer than 30 cm wasn't even noticed. Upwards, the sensor works best at a distance of 20-30 cm. Just one participant mentioned that the input was physically demanding, his hand even began to tremble after a few minutes of interaction. Without having precognitions with the Leap, none of the participants was able to interact without explanation. An optimal interaction was only possible after a short training, even with explanation.

Interviews After the practical part of the study, the subjects were asked several questions. First of all, they had to state which device they would prefer for an interaction. Two subjects preferred Kinect, the remaining two preferred the Leap. As an advantage of the Leap was mentioned the discreet interaction, while interacting with the Kinect requires sometimes overhanging gestures. In contrast, two subjects stated they would prefer an interaction in a greater distance than the Kinect, as this enables to keep a better track of the whole screen. Additionally one subject felt it confusing to have the input device below his hand but in his understanding, an interaction forwards was required. The participants stated that objects and pointer were conspicuous at every distance. Finally, subjects were asked if they could imagine to change the input modality in public spaces for any reason. Three subjects answered positive to this question, whereas the fourth participant negated the question caused by his laziness.

5.1.3 Significance for Laboratory Study

This pilot study was conducted to exclude the influence, the user's distance and other factors have on the interaction. For this aim, the two gesture recognition sensors Leap Motion and Kinect for Windows were tested. The main findings where the optimal distances to interact with this devices. As a consequence of the presented results, participants in the laboratory study and further studies should all be in the same distance from the devices, as the quality of interaction differed enormously at the tested distances. In case of the Kinect, optimal interaction is able from 1.5 m to 2.5 m. Two meters as a mean of the obtained values will be taken as distance for further studies. In case of the Leap, users will be told to hold their hand directly above the device with an upwards distance from about 20 cm. As subjects stated that all objects and the cursor on the screen were visible without difficulty at every distance, the proportions and sizes remain the same in further studies. Further, the whole interaction should not last too long, since subjects felt the interaction physically demanding, especially interacting with the Kinect sensor.

Since the interaction with both devices does not seem to be intuitive, the needed interaction should be explained to the subjects. Furthermore, a time for practising should be scheduled. This is expected to bring subjects skills in gestural interaction up to nearly the same level.

Frequency	Computer	Touch	Gesture
never	0	0	9
monthly	0	0	4
weekly	2	4	2
daily	13	11	0

Figure 5.1: Participants and the frequency with which they use computer, touch and gesture input.

Two out of four subjects stated the Leap sensor as favourite device, the other two the Kinect sensor. This may indicate the usefulness of multimodal user input, as every user could interact with its favourite technique. Another hint may be the fact that three out of four subjects reacted very interested in the possibility of changing the interaction modality within one application. All this motivates further investigation like the following laboratory study.

5.2 Laboratory Study

The second conducted study was a classical Fitts' Law experiment on a larger scale than the last study. Its aim was to compare the different devices by input time and user experience. Firstly, the study design is defined. Afterwards the detailed results are presented.

Goals Having excluded several factors with an influence on the Fitts' Law experiment, the aim of this study is now to compare the selected devices. The dependent variables in this study are the time, needed for an input, and user experience.

5.2.1 Design

The main study was conducted in the same environment as the pilot study. Hardware and software were adjusted to match the goal.

Subjects Fifteen subjects within the age from 18 to 51 took part in the study. Five of them were male, ten female. 14 out of 15 indicated they would use a computer more than weekly and having therefore good skills in handling a computer. The mouse is the mostly used input device for interacting with a computer, second is the touchpad of a laptop. Touch input is used by all subjects at least weekly. 14 out of 15 are using a smartphone and about the half of them additionally a tablet computer. For this reason, skills in handling Touchscreens were consistently estimated as very good. Only six participants are using gesture recognition at least monthly, with only one of them using the Kinect and the other five the Nintendo Wii as input device. Skills in gestural input were estimated rather bad. As in the previous study, none had any precognitions in handling with the Leap Motion sensor. All subjects and their precognitions are summed up in Table 5.1.

Setup A 23 inch Acer T232HL LED Touch monitor was used for output with a Lenovo U410 Ultrabook. The two devices Kinect and Leap Motion tested in the pilot study, were extended by the touch input of the monitor and an optical USB mouse. A mark at the floor indicated the place were users should stay during an input with the Kinect. The optimal placement for this mark was extracted from the pilot study results. For each participant, the height of the display was adjusted, so that the subjects outstretched arm could reach on the top edge of the display.

Procedure Participants used the different devices in counter balanced order to exclude the order of devices influencing the results. The procedure was the same for every user and every device. Before handling one device, participants were shown a picture with a short explanation about interacting with the actual device. After 30 seconds, the picture disappeared and 30 seconds of training began. This explanation and practising results from the recognition of the pilot study, which yielded that at least the Kinect and the Leap sensor can't be handled without previous knowledge or explanation. The shown explanation pictures can be seen in Figure 5.2-5.5. After this 30 seconds, the actual measurement and logging began. Data from the training was not collected and therefore not used in the analysis. Subjects had to perform 10 sets of 16 movements, ending up to 160 movements for each subject. Each set consists of all possible combination of amplitude and width. Timing began when the last object was hit and timing stopped when the cursor hit the object. Subsequently, subjects had to fill in a questionnaire. This procedure was iterated for every device. Participants were interviewed after having handled all four devices.

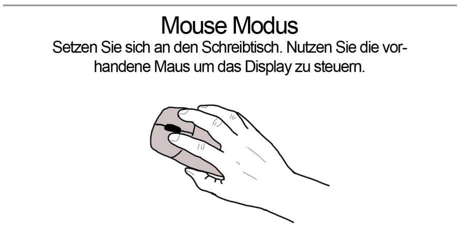

Figure 5.2: Mouse Mode explanation: Sit down at the desk. Use the available mouse to control the display.

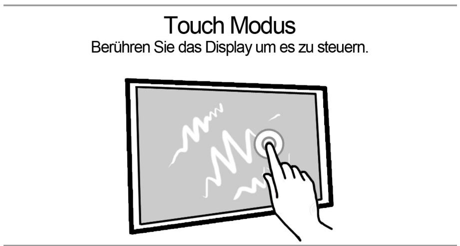

Figure 5.3: Touch Mode explanation: Touch the display to interact with it.

Figure 5.4: Leap Mode explanation: Hold your hand above the device. Try to control the display by moving your hand.

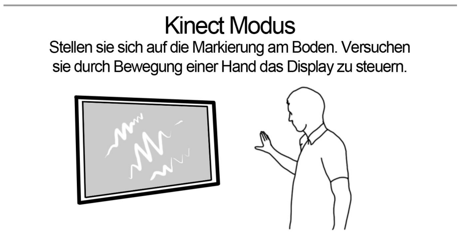

Figure 5.5: Kinect Mode explanation: Stand on the mark on the floor. Try to control the display by moving your hand.

Fitts' Law According to Fitts' Law, the independent variables for this study are amplitude A and target width W. Values for these variables were A=1,2,4,8 and W=1,2,4,8. Height could be excluded as variable, because of having quadratic objects. Crossing the width and amplitude values ended up in a set of 16 combinations including 10 different pairs of A and W. All 16 combinations appeared in a random order as moles on the screen. An unit has the size screenwidth/25 to fit on every output device. The only dependent variable of the Fitts' law study is the movement time.

Capture User Experience Subjects had to fill in five single sheets of a questionnaire. This happened between the interaction with the devices, afterwards they have been interviewed. The first page of the questionnaire consists of asking about precognition of the different input techniques. The following four pages are the same and asking about the subject' experience of the use of the four single devices. Each of this four pages consisting of a shortened Mini-AttracDiff, the Nasa-RTLX and two further questions about participants subjective assessment of getting better or faster during the input. The interview consists of two questions. First, subjects were asked to rank the devices after their subjective sensation and why they ranked them in this order to find individual advantages and drawbacks. Second, the Scenario 1 from Chapter 3.1 was presented, so that participants could imagine how a real world multimodal public display could work. Afterwards participants had to rate the introduced multimodal system and to state if they would go trough the single modalities or cancel the interaction for any reason.

5.2.2 Results

The collected logging data was adjusted and put into a linear regression test. After analysing the single devices, a comparison of the devices was executed. The detailed questionnaire and interview results are presented afterwards.
The grand mean input times of the four devices were:

- Mouse: 460 ms
- Touch: 507 ms
- Leap: 568 ms
- Kinect: 772 ms

A One-Way-ANOVA test, considering all data points, showed that the differences in movement-times between all four devices are significant with having all p-values smaller than 0.001.

Adjustment of Data Before attempting any statistical test, data needed to be adjusted. Therefore, all trials with a movement time more than three standard deviations from the mean of the corresponding device were removed. Furthermore, all trials with a movement time less than 100 ms were deleted, caused by the fact that the minimal human reaction time is defined as 100 ms [37]. The few trials with such low reaction time seem to be hit more incidental than caused by a normal reaction.

After plotting the particular mean movement times for each subject and device, one enormous outlier attracted the attention as seen in Figure 5.6. This subject has had a really outlying movement time, but only in combination with the mouse as input device. After matching logging data with questionnaires results, it became clear that the subject had never handled a computer before and therefore never had had contact with a mouse before. As this subject has an crucial influence on the general results, all data of it were removed from further analysis. Further significance of this outlier is discussed in Chapter 7. Significance tests on the remaining data

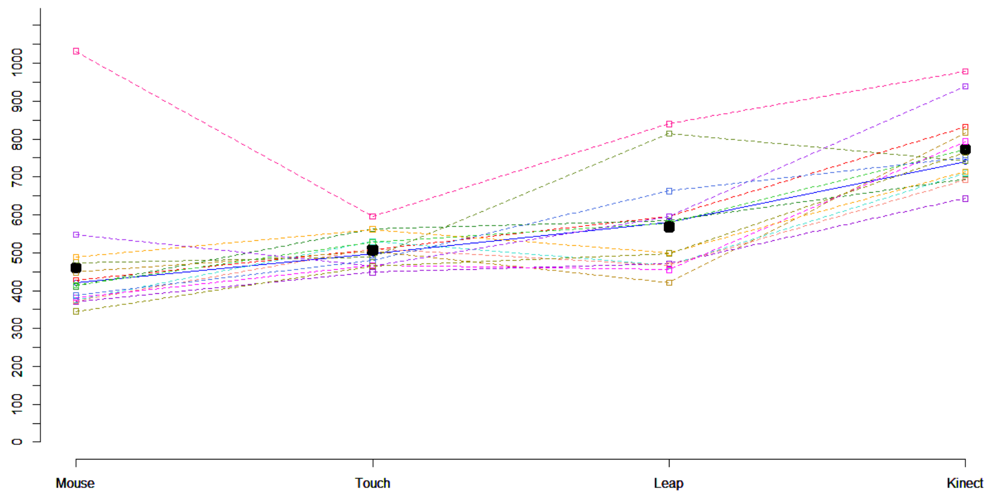

Figure 5.6: Mean movement times of all 15 participants. The black dots represent the grand means.

didn't show other significant differences between users and movement times.
A linear Regression showed no differences between movement time and duration of the interaction. If there had been any difference, like subjects improved their achievement after the first sets of moves, the first sets would have been removed as further trial time.

Fitting Fitts' Law To analyse the trials after Fitts' law, the data of all four devices was entered separately in a test of linear regression.

The results of the linear regression tests are shown in Figures 5.7 - 5.10 and Table 5.1. All p-values were smaller than 0.001, all R^2 were bigger than 0.5800. Resulting from the p-value being constantly smaller than 0.001 for all devices, the linear regressions can be accepted as significant.

Device	Intercept, a	Slope, b	R^2	TP, 1/b	Error
Mouse	340	60	0.7736	16.67	8.28
Touch	474	44	0.6564	22.73	7.20
Leap	438	95	0.5807	10.53	19.94
Kinect	596	141	0.7908	7.09	16.52

Table 5.1: Results of the linear regression of the four devices.

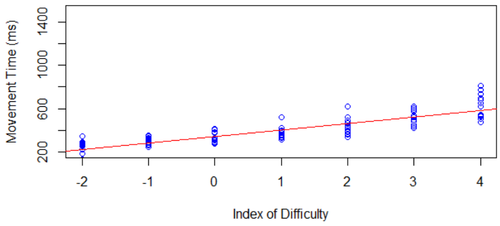

Figure 5.7: Linear Regression Mouse

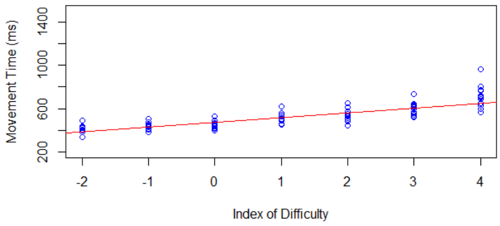

Figure 5.8: Linear Regression Touch

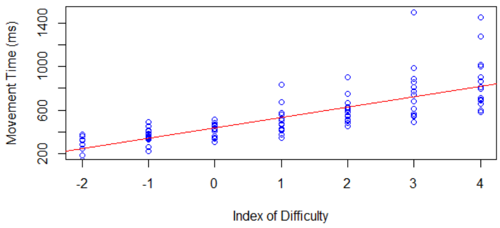

Figure 5.9: Linear Regression Leap

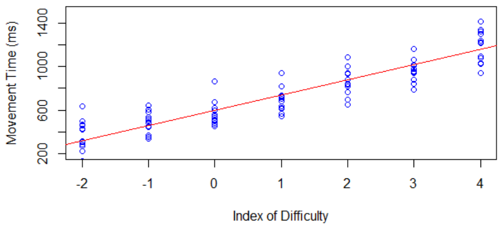

Figure 5.10: Linear Regression Kinect

Questionnaires The already short Mini-AttracDiff was shortened again, as a few questions do not fit to the belonging tasks. It consisted finally of five opposing word pairs, which can be seen in Figure 5.11, where the most frequent answer of each question for each device is shown.
To determine physical and mental demand of the task, the NASA Task Load Index was used in an adjusted form, called NASA-RTLX. The single results of each question contained in the NASA-RTLX are presented in Figure 5.12, showing the most frequent answer.
In Table 5.2, the determined values of AttracDiff and NASA-RTLX are shown. In case of the AttracDiff, the closer the value is to zero, the better was the belonging task evaluated by the subjects. The same applies to the NASA-RTLX, whereby a smaller value means a less mental and physical demand of the subject. The maximum of the AttracDiff would be seven, whereby the maximum of the NASA-RTLX would be 20.

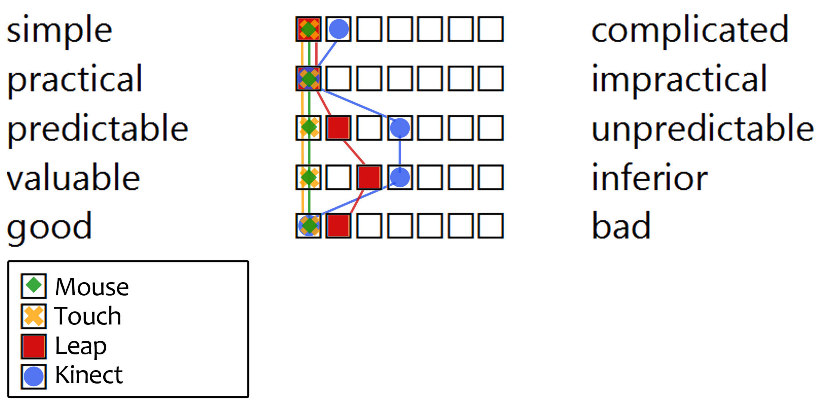

Figure 5.11: Most frequent answers on the five questions of the shortened Mini-AttracDiff. mouse and touch achieved same results.

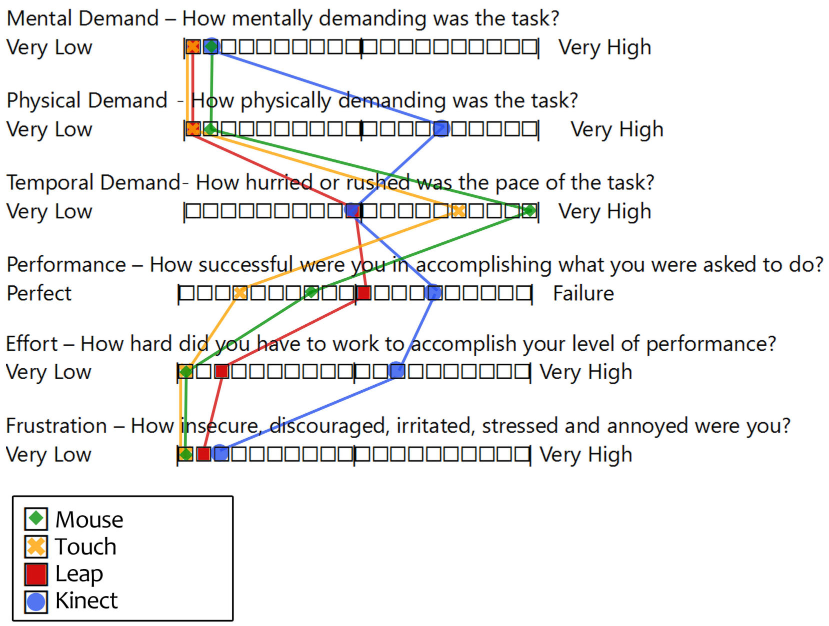

Figure 5.12: Most frequent answers on the questions of the Nasa-TLX questionnaire.

Method	Mouse	Touch	Leap	Kinect
AttracDiff	1.68	1.59	2.55	3.19
NASA-RTLX	5.80	6.62	7.12	8.89

Table 5.2: Results of the questionnaires for AttracDiff and NASA-RTLX.

Significance tests trough using t-tests, showed that in both cases, AttracDiff and NASA-RTLX, the interaction with the Kinect was rated significantly worse than the other three devices. In the AttracDiff, Leap was also rated significantly worse than mouse and touch. No significant difference can be figured out between mouse and touch in both cases. In the case of RTLX, even touch, mouse and Leap do not show any significant distinction in their interactions rating.

Device	faster	same	slower
Mouse	5	9	1
Touch	4	9	2
Leap	7	5	3
Kinect	4	9	1

Table 5.3: Results to the question if subject thought they got faster or slower during one interaction.

Device	preciser	same	more inexact
Mouse	1	11	3
Touch	5	9	1
Leap	8	5	2
Kinect	3	10	1

Table 5.4: Results to the question if subject toughed they got preciser or more inexact during one interaction.

Most subjects experienced their interaction as being constantly in speed and accuracy with mouse, touch and Kinect. In the case of Leap, most subjects experienced that they went preciser and faster during interaction with the sensor.
Several linear regression did not show any significant differences in input speed over the time of the interaction with one device. This contradicts the subjective experiences of the subjects, in the case of Leap.
Inaccuracy can't be determined in the scope of this study, as there are no selection coordinates storable for the devices Kinect, Leap and mouse. Caused by the fact that no explicit selection was needed, as moving the cursor over the object is regarded as selection.

Interviews As the mouse was only added to have reference values, it was omitted in the interviews. That is why there aren't any interview results of the mouse.

The results of the interviews are now presented. Like in the pilot study, users were asked which device they would like to use in public spaces. This time they were allowed to name more than one device. The answers were again various. While eleven out of 15 subjects would like to interact with a touch display, seven named the Leap and six participants named the Kinect sensor. One subject even named all three devices. Four subjects would only like to interact with touch, three subjects preferred the single use of the Leap, while none of them preferred the Kinect sensor over the other two.

Afterwards, the subjects were asked to reason their answers. Subjects who favoured the gesture interaction over touch interaction mentioned mostly the lack of hygiene on public touch displays, with displays often found dirtied and smeared with fingerprints. Whereas the other participants stated that a touch display is the most precise method and simultaneously the most intuitive one for input on large screens, without requiring to learn anything new. Two subjects stated that touch input is the safest input method as passerby cannot easily look on the screen while one interacts with it.

The second favourite device in the interviews was the Leap Motion sensor. Several advantages of the controller were mentioned by the subjects, the distance in which the interactions takes place is for four subjects the most pleasant one for longer interaction. When using the Leap, the user is far enough to have an overview of the whole display but simultaneously near enough that other people cannot walk trough in front of the user. Another stated advantage of the Leap Motion is the preciser input possibility than with the Kinect. One user mentioned he would prefer the Leap because it is the device which affords least physical demand. In opposition to that stands the answer of another subject who stated the Leap would be to physical demanding for longer interaction. Further mentioned disadvantages were the unfamiliar and rather more complex handling which is in need of for practice when users have no previous knowledge with the Leap sensor.

The Kinect sensor was favoured the least in the interviews. No relevant benefits for the Kinect were stated by the participants. The six subjects who indicated in the previous question that they would interact with the Kinect explained their answer as the interaction with the Kinect would be rather acceptable than desirable. Several drawbacks were mentioned therefore. Most subjects are of the opinion they are located to far away from the display while interacting. Caused by the huge distance, passerby could walk trough user and display and disturb the ongoing interaction.

Various subjects stated furthermore that the interaction with the Kinect sensor looks strange and may be not understood as interaction by passerby. This is advanced by the needed sweeping movements and gestures for interaction. Further drawbacks are the high physical demand when interacting longer and the less granular gesture recognition.

After that, the subjects were told Scenario 1 and were firstly asked, if they can imagine to go trough all modalities, or if they would break off the interaction prematurely. 12 out of all 15 subjects affirmed this question, whereas the other three negated it. They had to justify their answers afterwards.
Seven of the subjects who answered positive explained their answer by their interest in such new interactions. They would try the system and if everything is explained in a detailed and comprehensive way, an interaction trough all modalities is imaginable for them. Three out of the 12 subjects had contact with unimodal public displays and had made positive experience. They went a little bit enthusiastic after telling the scenario. All three stated they would immediately try the display when coming around it. Another subject understood the concept behind. He stated that the multimodal display sounds practical for longer input, as roughly input is made with the not so granular Kinect and preciser interaction takes place with the more preciser touch device. The three subjects who negated the question named different reasons. One stated he would feel observed by the display if it reacts specifically on him. Another one said that the reactions of the display would appear untrustworthy. The third subject wouldn't interact with the display alone as she would feel strange doing this interactions in public space. She can imagine to interact when friends are with her and they interact as a group.

Comparison of the Devices The mean movement times of all devices indicated a difference between the general input speed of all devices. Therefore, the differences are now considered more closely. In an One-Way ANOVA it was proven that the differences in movement-times between the devices are significant. The grand mean movement-times were for the mouse 460 ms, in case of touch 507 ms, for the Leap 568 ms and with the Kinect 772 ms. Further t-tests showed that movement-times of the mouse were in general significant smaller than movement-times of touch, Leap and Kinect. Movement-times on the touchscreen were significant smaller than the ones of Leap and Kinect. Last but not least, input times on the Leap Motion were significant smaller than times for an input on the Kinect.
Linear regression emphasises this rating, as intercept rises steadily in the order as mean movement time. Whereas in the case of slope and throughput, mouse and touch are interchanged. All regression coefficient emphasise the Kinect sensor being the slowest device, where throughput is even only half as large as the throughput of mouse and touch.

6 Discussion

After evaluating the developed system and presenting the results, a discussion about the received results is necessary. Next, a critical reflection will be made to clarify if all research questions were answered sufficiently.

6.1 Discussing the Results

This subsection will summarize and then interpret the results of the evaluation.

Summary Looking on the presented results of the laboratory study, it is clear that significant differences in the input speed and usability of the tested devices do exist. It seems that the subjective evaluation of the participants corresponds to the measured data.
In both, Fitts' law and user experience, the Kinect sensor scored the worst result in comparison with the other devices. The Leap sensor scored worse than touch and mouse in movement time and the Mini-AttracDiff. Whereas mouse and touch scored nearly the same in the questionnaires, mouse scored better in the Fitts' law test.
The same pattern yields the interviews, Kinect resulted at the least favoured one with a lot of mentioned disadvantages. On the second place is again the Leap sensor, with several mentioned advantages and disadvantages. When it comes to interaction with public displays, the most subjects clearly named the touch input as favourite input device.

Interpretation When beginning to interpret the results, it comes to mind that the proposition, which was made at the beginning was reasonable. It was proposed to use from-a-distance interaction for rough input, and using interaction near the screen for preciser input.
The conclusion for the Kinect is, that it is perfectly suitable for short and rough input from-a-distance. This is motivated by the facts that an optimal interaction is only possible between 1.5 and 2.5 m. Further, it can be physical demanding after a certain amount of time and a long time is needed for granular input. Based on the several disadvantages and less advantages, it could also be taken in consideration to remove the Kinect from further applications or replace it with another from-a-distance input modality, like gaze.
When using the Kinect, it should be possible that a user can change as soon as possible the input device as the difficult interaction with the Kinect could discourage the user to continue his interaction.

The conclusion for the Leap sensor is, that it is a device which a user needs to get used to when interacting for the first time. Once a user is used to handle the Leap, it is a conceivable replacement for touch input. Arguments that state, the interaction with the Leap would be more physical demanding than interaction with a touch screen, are difficult to comprehend as both devices require nearly the same movements of the users hand.

Touch seems to be the over all favourite input device resulting from captured user experience. Additionally it got the best results in the Fitts' law study as it was the secondly fastest device after the mouse. It seems to be intuitive and therefore easy to handle. Furthermore, most people nowadays are used to touch interaction and no explanation would be required.
But when talking about hygiene, the touch screen reaches its limits. Additionally, when considering large screens, the touch screen may be not reachable for all possible users and needs to have an alternative.

In the interviews it was argued that elderly people will not be able to interact with such innovative and uncommon input devices as the Leap and Kinect sensors are. The outlier-user who was removed from data in Chapter 5.2.2 may indicate the contrary. This subject without any precognitions in using a computer received better movement-times with touch, Leap and Kinect than with the mouse. This fact could be interpreted as touch and gesture input being more intuitive than the usage of a mouse.

6.2 Critical Reflection

At this point, it needs to be clarified if all research questions were answered in this thesis.

- *How can multimodality be combined with proxemics in public displays?* The concept of combining multimodality with proxemics on a large screen was proposed and implemented in a prototype. Caused by the scope of this thesis it wasn't possible to conduct a further field study. This should be fetched later to emphasize this work.

- *How can the area in front of the display be used to create a gapless interaction area? Is proximity useful to achieve this goal?* In this combination of devices, it is not possible to create a gapless interaction area. The Leap does not work further than 10-20 cm away from the device and the Kinect can only be used afar 1 m. A gap of at least 80 cm exists therefore. For several users, the Kinect interaction is not even possible from 1 m distance. Proximity can be useful to create a gapless interaction area as the proxemic dimensions can be used as parameter on which the modality or input device is chosen. This is not only useful for creating a gapless interaction space, but also for multimodal installations itself.

- *Based on the selected devices: Which device can be used the best at which distance to the screen? How affects granularity of the input device on the quality of interaction in different distances?* As the Kinect seems to be the less granular input device and its best working interaction distance is located between 1.5 and 2.5 m, a user should be lead to this distance when aspiring to interact with the Kinect. The best interaction distance of the Leap and touch is limited to their technical specification. As both devices are developed for input from a short distance, the are preciser than input with the Kinect.

- *Is there a difference in peoples' favourite input device?* Yes, it is. There are people who would prefer gesture interaction in public spaces to touch interaction and the other way round.

- *Can new input techniques be brought in public spaces or are they to unfamiliar for people to use?* This question is difficult to answer. In any case, this thesis made clear that a precise explanation is required for new input devices to prevent to scaring away possible users who don't understand the required moves to interact with the screen. Furthermore, the installation should respond to the users need by considering further proxemic dimensions for changing the modality. This could also be used to simplify the input with novel devices and modalities.

Summed up, the main goals of this thesis have been achieved. Nevertheless, a lot of future directions can be pursued. This can be seen in the next chapter.

6.3 Limitations

As a bachelor thesis is limited to a specific workload, not all firstly planed steps of procedure could be executed in this scope. In a first draft of the steps of procedure, a field study was planned additionally to the pilot and laboratory study, to test the multimodal proximity-aware system in the wild. After beginning to work on this thesis, it quickly became clear that the third study would exceed the scope of this thesis.
Furthermore, there weren't all devices available which would have been useful for further

comparison. The planned display size from 52 inch, which was also used in the pilot study, couldn't be used in the laboratory study as no touch screen in this size was available at the authors university.

6.4 Future Directions

Future directions are primarily based on the limitations of this thesis. Firstly, the study can be reproduced on a 52 inch touch screen to examine the influence of display size on quality of user input.

In a second step, a field study should be conducted. To do so, the created interaction area trough touch, Leap and Kinect needs to be adjusted, as it was found out that a gap where interaction is not possible exists between Leap and Kinect. Therefore, either a replacement for the Kinect or an additional device to fill the gap in the created interaction area should be added.

In the scope of this field study, it would be interesting to integrate explicit selection, as gestures for Leap and Kinect. Furthermore, smooth transitions should be made between the modalities to enable fluent interaction despite a modality change. To allow smooth transitions, it is necessary to provide a gapless interaction area.

Furthermore, the removed eye tracker should be taken to account for previous work in this field, as gaze could increase the possible interaction area enormously.

Another possible future work would be the use of two devices simultaneously. This could be implemented as explicit use of the different devices, but also as a real simultaneously use of more than one device by a user.

7 Conclusion

In this last chapter, an overall summary of this thesis will be made. Additionally the contribution will be summarized and a conclusion by the author will be drawn.

7.1 Summary

The topic of this thesis was primarily to combine more than one modality to a public interactive display. Therefore, the methods of proxemic and multimodality were put together. Proxemic dimensions were used as the parameter on which the modality changes. In Chapter 2, background and related work to the topic were introduced, including public displays, proxemic and multimodal interaction. For each field, a general definition and related work was stated. Chapter 3 is about the problem description opened up by the goals of this thesis. The chapter includes two scenarios, describing the usage of proximity aware multimodal public displays. Afterwards the research questions were defined. The following subsections dealt with answering the research questions. In Chapter 4, the implementation of the Fitts' law game was declared, containing the general concept and detailed information about the implementation for each device. Both conducted studies with their design and results were described in Chapter 5. The results were discussed in Chapter 6 which includes additionally a critical reflection and future directions.

7.2 Summary of contribution

Firstly, it was tested in which distance the devices Leap and Kinect work the best. When trying to interact outside of this ranges, problems like inaccuracy can occur. Therefore, the Leap should be placed 10 cm ahead of the Display, to cover the first 20 cm in front of it. The Kinect works best between 1.5 m and 2.5 m for explicit interaction.
This thesis further conducted a comparison of four different input devices consisting of three different input modalities. The devices include two standard input devices for computer systems, whereas the other two are rather new devices. While there exist lots of publications for performance of the input with mouse and touch, performance tests of the Kinect and Leap sensor are rather rare. Furthermore, the existing studies can't be used for comparison purposes of these four devices, as the same data set should be used for a Fitts' law study with performance comparison.
Additionally, the concept of combining proxemics with multimodality was introduced. This opens up a new field trough combination of two existing fields of research in Human-Computer-Interaction. As it was not possible in the scope of this thesis to examine more possibilities, it should be taken to account for future work. This combination of fields of research allows a new way to use the whole area in front of a public display and enable interaction in this area without any gaps. Unfortunately, it was not possible to create a gapless interaction area with the selected devices, as Leap and Kinect can't track all users interaction as stated in both manufacturer specifications.

7.3 Authors Conclusion

The issue of this thesis seems to be a promising approach for the field of public displays, since a lot of space is usually present around a display installation. This space should be fully used, for both implicit and explicit interaction. Interaction afar a display should be quick and uncomplicated as interaction with the Kinect allows to, whereas interaction near the display can require more granular input actions.
Further, it was shown in this thesis that new input devices such as the Leap can be brought in public spaces when an explanation is present. This indicates the feasibility of creating multimodal displays with even more novel combinations of modalities. The whole topic is even more supported by the received result that users notion of the favourite input device differs significantly.

Furthermore, it was proven that each of the selected devices has his specific interaction distance, in which it works the best. This interaction zones should be respected.
The Kinect had received bad results in both studies. Furthermore, a gapless interaction is not possible with the selected devices, as the received interaction areas for Leap and Kinect from the pilot study had shown. For this reasons, the Kinect should be replaced in possible future works.

Content of the enclosed CD

I. Bachelor Thesis

   a. PDF: Contains the thesis.pdf
   b. Latex: Contains the thesis.tex

II. Source Code

   - WhacAMole : The software implemented for this thesis.
   - README.txt : Contains explanations for execution.

III. Bibliography

   - References: All sources which were available in electronic format.
   - Web-References: All internet sources which were available in electronic format.

IV. Questionnaire

   - Questionnaire.pdf : The Questionnaire used in the Laboratory Study

References

[1] Jörg Müller, Florian Alt, Daniel Michelis, and Albrecht Schmidt. Requirements and Design Space for Interactive Public Displays. In Proceedings of the international conference on multimedia, MM'10, pages 1285-1294, New York, NY, USA, 2010. ACM.

[2] Florian Alt, S. Schneegaß, A. Schmidt, J. Müller, N. Memarovic. How to evaluate public displays. In: Proceedings of the 2012 International Symposium on Pervasive Displays. ACM, 2012, p. 17.

[3] Harry Brignull, Yvonne Rogers. Enticing people to interact with large public displays in public spaces. In: Proceedings of INTERACT. Vol. 3, 2003, pp. 17-24.

[4] Stefan Schneegass and Florian Alt. SenScreen - A Toolkit for Supporting Sensor-enabled Multi-Display Networks. In Proceedings of the 3rd ACM International Symposium on Pervasive Displays, PerDis '14, New York, NY, USA, 2014. ACM.

[5] Fabius Steinberger, Marcus Foth, and Florian Alt. Vote With Your Feet: Local Community Polling on Urban Screens. In Proceedings of the 3rd ACM International Symposium on Pervasive Displays, PerDis '14, New York, NY, USA, 2014. ACM. [PDF] Nora Broy, Stefan Schneegass, Florian Alt, and Albrecht Schmidt. FrameBox and MirrorBox: Tools for Prototyping User Interfaces for 3D Displays. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, New York, NY, USA, 2014. ACM.

[6] Jonna Häkkilä, Maaret Post, Stefan Schneegass, Florian Alt, Kunter Gültekin, and Albrecht Schmidt. Let me catch this! Experiencing Interactive 3D Cinema through Collecting Content with a Mobile Phone. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, New York, NY, USA, 2014. ACM.

[7] Florian Alt, Alireza Sahami Shirazi, Thomas Kubitza, and Albrecht Schmidt. Interaction techniques for creating and exchanging content with public displays. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13, pages 1709-1718, New York, NY, USA, 2013. ACM.

[8] Rui Jose, Jorge Cardoso, Florian Alt, Sarah Clinch, and Nigel Davies. Mobile applications for open display networks: common design considerations. In Proceedings of the 2nd acm international symposium on pervasive displays, PerDis '13, pages 97-102, New York, NY, USA, 2013. ACM.

[9] Nemanja Memarovic, Keith Cheverst, Marc Langheinrich, Ivan Elhart, and Florian Alt. Tethered or free to roam: the design space of limiting content access on community displays. In Proceedings of the 2nd acm international symposium on pervasive displays, PerDis'13, pages 127-132, New York, NY, USA, 2013. ACM.

[10] Nemanja Memarovic, Marc Langheinrich, Florian Alt, Ivan Elhart, Simo Hosio, and Elisa Rubegni. Using public displays to stimulate passive engagement, active engagement, and discovery in public spaces. In In proceedings of media architecture biennale 2012, MAB'12. ACM, ACM, 2012. [

[11] Nemanja Memarovic, Marc Langheinrich, and Florian Alt. The Interacting Places Framework: Conceptualizing Public Display Applications that Promote Community Interaction and Place Awareness. In Proceedings of the 2012 international symposium on pervasive displays, PerDis'12, pages 71-76, New York, NY, USA, jun 2012. ACM.

[12] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. Looking Glass: A Field Study on Noticing Interactivity of a Shop Window. In Proceedings of the 2012 acm conference on human factors in computing systems, CHI'12, pages 297-306, New York, NY, USA, apr 2012. ACM.

[13] Florian Alt, Moritz Balz, Stefanie Kristes, Alireza Sahami Shirazi, Julian Mennenöh, Albrecht Schmidt, Hendrik Schröder, and Michael Gödicke. Adaptive User Profiles in Pervasive Advertising Environments. In Proceedings of the european conference on ambient intelligence, AmI'09, pages 276-286, Berlin, Heidelberg, nov 2009. Springer-Verlag.

[14] Florian Alt, Nemanja Memarovic, Miriam Greis, and Niels Henze. UniDisplay - A Research Prototype to Investigate Expectations Towards Public Display Applications. In Proceedings of the 1st Workshop on Developing Applications for Pervasive Display Networks, PD-Apps '14. IEEE, 2014.

[15] Tongyan Ning, Jörg Müller, Robert Walter, Gilles Bailly, Chachatvan Wacharamanotham, Jan Borchers, and Florian Alt. No Need To Stop: Menu Techniques for Passing by Public Displays. In Proceedings of the chi workshop on large displays in urban life, Vancouver, BC, Canada, apr 2011.

[16] Nemanja Memarovic, Marc Langheinrich, Keith Cheverst, Nick Taylor, and Florian Alt. P-layers - a layered framework addressing the multi-faceted issues facing community-supporting public display deployments. ACM Transactions on Computer-Human Interaction (ToCHI), 2013.

[17] Florian Alt, Jörg Müller, and Albrecht Schmidt. Advertising on Public Display Networks. Ieee computer, 45(5):50-56, may 2012.

[18] Edward Twitchell Hall. The hidden dimension. Doubleday, 1966.

[19] Till Ballendat, Nicolai Marquardt, Saul Greenberg. Proxemic interaction: designing for a proximity and orientation-aware environment. In: ACM International Conference on Interactive Tabletops and Surfaces. ACM, 2010, pp. 121-130.

[20] Miaosen Wang, Sebastian Boring, Saul Greenberg. Proxemic peddler: a public advertising display that captures and preserves the attention of a passerby. In: Proceedings of the 2012 International Symposium on Pervasive Displays. ACM, 2012, p. 3.

[21] Nicolai Marquardt, et al. The proximity toolkit: prototyping proxemic interactions in ubiquitous computing ecologies. Proceedings of the 24th annual ACM symposium on User interface software and technology. ACM, 2011, pp. 315-326.

[22] Daniel Vogel, Ravin Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In: Proceedings of the 17th annual ACM symposium on User interface software and technology. ACM, 2004, pp. 137-146.

[23] Saul Greenberg, et al. Proxemic interactions: the new ubicomp?. In: interactions. 18.1, 2011, pp. 42-50.

[24] Thorsten Prante, et al. Hello. wall?beyond ambient displays. In: Adjunct Proceedings of Ubicomp. 2003, pp. 277-278.

[25] Marie-Luce Bourguet. Designing and Prototyping Multimodal Commands. In: INTERACT'03 Human-Computer Interaction, IFIP, 2003, pp. 717-720.

[26] Jakub Dostal, et al. SpiderEyes: designing attention-and proximity-aware collaborative interfaces for wall-sized displays. In: Proceedings of the 19th international conference on Intelligent User Interfaces. ACM, 2014, pp. 143-152.

[27] Sophie Stellmach, et al. Designing gaze-supported multimodal interactions for the exploration of large image collections. In: Proceedings of the 1st Conference on Novel Gaze-Controlled Applications. ACM, 2011, p. 1.

[28] Jayson Turner, Andreas Bulling, and Hans Gellersen. Combining gaze with manual interaction to extend physical reach. In: Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction. ACM, 2011, pp. 33-36.

[29] Michael Johnston,Srinivas Bangalore. MATCHKiosk: a multimodal interactive city guide. In: Proceedings of the ACL 2004 on Interactive poster and demonstration sessions. Association for Computational Linguistics, 2004, p. 3.

[30] Roope Raisamo. A multimodal user interface for public information kiosks. In: Proceedings of PUI Workshop, San Francisco. 1998.

[31] Paul M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. In: Journal of Experimental Psychology. 47.6, 1954, p. 381.

[32] Paul M. Fitts, B. K. Radford. Information capacity of discrete motor responses under different cognitive sets. In: Journal of Experimental Psychology. 71, 1966, pp. 475-482.

[33] I. Scott MacKenzie. Fitts' law as a research and design tool in human-computer interaction. In: Human-computer interaction 7.1. 1992, pp. 91-139.

[34] Atsuo Murata, and Hirokazu Iwase. Extending Fitts' law to a three-dimensional pointing task. In: Human movement science. 20.6, 2001, pp. 791-805.

[35] Olivier Chapuis, Renaud Blanch, Michel Beaudouin-Lafon. Fitts' law in the wild: A field study of aimed movements. 2007.

[36] Frank Weichert, et al. Analysis of the accuracy and robustness of the leap motion controller. Sensors 13.5, 2013, pp. 6380-6393.

[37] Prof. Dr. Gerhart Groos. Künstliche neuronale Netze. In: Vorlesungen über Informatik. Springer Verlag, Berlin, Germany, 1998, pp.105-164.

Web-References

[38] I. Scott MacKenzie, W. Buxton. Extending Fitts' Law to Two-Dimensional Task. In: CHI'92 Proc. ACM CHI Int. Conf. on Human Factors in Computing Systems. New York, NY, USA, 1992, pp. 219-226. 30. Oct. 2014. http://www.billbuxton.com/fitts92.html

[39] Leap Motion, Inc. 4. Sept. 2014 https://www.leapmotion.com/product

[40] Microsoft Corporation, Mircrosoft Developer Network: Kinect Sensor. 4. Sept. 2014. http://msdn.microsoft.com/en-us/library/hh438998.aspx

[41] Lawrence James Sambrooks, School of Computer Science, Engineering, and Mathematics, Faculty of Science and Engineering. Evaluation of Touch and Gestural Interaction Techniques for Post-WIMP User Interfaces. 14. Oct. 2014 https://wiki.csem.flinders.edu.au/pub/CSEMThesisProjects/ProjectSamb0014/samb0014_thesis_final.pdf

[42] H. Drewes, Ludwig Maximilian Universität München. A Lecture on Fitts' Law. 2. Oct. 2014. http://www.cip.ifi.lmu.de/~drewes/science/fitts/A%20Lecture%20on%20Fitts%20Law.pdf