Institute of Visualization and Interactive Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Master Thesis Nr. 3560

Combining Multiple Gesture Sensing Technologies to Interact with Public Displays

Mina Metias

Course of Study:	Infotech
Examiner:	Prof. Dr. Albrecht Schmidt
Supervisor:	Markus Funk, Dipl. -Inf., Tilman Dingler, Dipl. -Medieninf, Florian Alt, Prof. Dr. Andreas Schuller, Dipl.-Inf.
Commenced:	2013-08-01
Completed:	2014-01-31
CR-Classification:	H.5.2

Abstract

Over the past years, the urban landscape has witnessed a revolution with regard to information delivery systems. With the emergence of new hardware sensing technologies, such as multi-touch enabled screens and camera-based technologies, the information delivery business model shifted away from the classic static conventional paradigm to a new dynamic, novel interactive one. The adoption of the new paradigm was proven to have a positive impact with respect to efficiency and effectiveness of information delivery. Public displays are considered to be one of the newest and most popular mediums used when it comes to addressing larger audience. Although a lot of current interactive applications are now deployed on public displays, only a few focus on the different interaction modalities available based on the proximity of users with respect to the screen.

In the study presented in this thesis we declare a system design model combining two different motion sensing technologies to utilize and take full advantage of the interactive spectrum in front of a public display system. Based on the relative proximity from the public display, the model introduced four different interactive zones, each with different granularity of interaction. The study also demonstrates how a seamless transition across the different zones is essential to guarantee an immersive, undisturbed user experience. Moreover, the study proposed a user-defined design guideline for gestural interaction within and across the different interactive zones.

Acknowledgement

I would like to express my deepest gratitude to Professor Albrecht Schmidt, Markus Funk, Tilman Dingler and Professor Florian Alt, my research supervisors, for their patient guidance, enthusiastic encouragement and useful critiques of this research work.

I would like to gratefully acknowledge the supervision by Mathias Peissner and Andreas Schuller, my research supervisors from Fraunhofer. They always supported my ideas from day one and were always there for me when I needed them. I would like to thank them for their openness to new ideas, cooperation and continuous guidance.

I would like to specially thank my closest friends and true partners who started this journey with me, Mirna Ayman, Christine Louka, Michael Girgis, and Youssef Aziz.I wouldn't have been able to do it without you guys.

Special thanks should be dedicated to my one and only thesis buddy, who is soon to be my sister in law Maria Botros; you have always been a true bro!

I am particularly grateful for the assistance given by my friends Miriam Botros,Yomna Ali, Mina Assad, Pierre Ibrahim, Hallawa.

Finally, I wish to thank my parents and my brother for their support and encouragement throughout my study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> *"Yes, getting the technology to work is hard, but the really hard part is getting the human-system interaction right"*
> -Donald Norman

This quote by the expert and evangelist in the field of cognitive sciences and usability engineering summarizes the real challenge of the human computer interaction field. Having the most recent and advanced technology is not enough to have a successful system and deliver the correct user experience. With the emergence of new technologies the task becomes more challenging and even more compelling to investigate. New technologies provide the ability to either replace old fashioned ones or solve the deficiencies previous technologies couldn't solve. Innovations in user interfaces have played a major role in reshaping and restructuring the medium in which information reaches the desired targeted audience.
The marketing field is possibly one of the best examples that witnessed a revolutionary progress alongside to the user interfaces. Any successful marketing strategy utilizes the state of the art in user interfaces to increase its outreach and intensify the focus on the desired target group. The embracement and adoption of technologies can be best observed in the evident transition from classical static flyers and advertisement banners to internet-based social networks and today's state of the art smartphone applications.

As public displays have always been a field of innovation and novelty, and as they represent a medium for information delivery that is increasingly deployed not only by large outdoors but also small retailers, it enticed our research group to put under further comprehensive inspection with respect to new sensing technologies that enhanced and heighten the system's usability factor. Hence, the premises for our research rose: How could new technologies resolve imperfections currently facing information delivery in the field of public displays?

1.1 Motivation

Since the above question is one worthy of thorough exploration, and as it has always been that science has taken the burden to find solutions for every day life problems, a major leap in user interaction evolved specially over the last decade. User interfaces have moved away from traditional WIMP interfaces to the more enjoyable, more intuitive touch interaction, on-screen gestures and mid-air gestures. With the introduction of the Apple Iphone to the consumer market, the touch-screen technology invaded the smartphones market. Not long after, the touch-screen revolution hit the personal computer markets as well, specially with the introduction of Windows 8 where multi-touch gestures such as zoom, pinch, and tap were out-of-the-box ready to meet user's expectations. Parallel to the multi-touch breakthrough, the mid-air gestural interaction has made its way particularly in the gaming and entertainment systems. Naming just a few of the famous motion sensing technologies, Microsoft Kinect and Nintendo Wii [38] would top off the charts.

Establishing such a solid foundation for new novel user interaction techniques encouraged the researchers and practitioners to remodel the concept of user interactions in the area of public displays. Multiple experiments and exciting work have been conducted to explore the possibility of engaging novel motion sensors with public display to enhance the interactive facility and deliver greater user experience . Cost-efficiency and clear-cut precision were the main motivations that fueled such scientific migration from traditional static interfaces to new bilateral dynamic ones.

However, previous work was mainly expanded on the basis of having a limited set of interaction modalities with confined consideration to the variable proximity between the user and the public display regarding gestural interaction. Hence, the main goal for this thesis was determined: To investigate and conceptually understand the effects of human-computer interactivity caused by simultaneous combination of multiple gesture sensing technologies that would allow the public display to be distance-aware. This entails both basic contradictory questions; how can public displays be flexible enough to assign different functionalities or services in accordance to the user's proximity and at the same time maintain the homogeneity of the interaction spectrum across the different interaction zones without introducing extra disturbance for the user? Achieving such a fine-tuned balance would open the gates to a new interaction scheme that penetrates the 3D physical space in front of the public display to its utmost potentials

In order to reach a firm denouement in response to these questions, it was essential to study two types of applications that run on public displays.

A) Continuous real-time feedback applications (e.g. Mouse pointer control [21])

B) Purely gesture-based applications (e.g. Map navigation [8])

One of the many goals of this thesis is to build an interactive prototype system for each application paradigm to examine and help discern the potential problems that might occur in the two dominant application archetypes, as well as propose a potential solution for the major design obstacles. As this thesis was a cooperative project between the University of Stuttgart and Fraunhofer Institut für Arbeitswirtschaft und Organisation (IAO), the interactive wall project at Fraunhofer was found to be the perfect fit as a use case environment for development and evaluation purposes. The interactive wall project is a public exhibit that consists of four adjacent public display segments, each supported with a Kinect motion sensor on top of it to enable interactivity. Figure 1.1 shows the construction of the interactive wall. The project is continuously under extensive interactive design progress and enhancement. Moreover it is also deployed in actual large fairs (e.g. The World Usability Day [1]) and international exhibitions (e.g. Hannover Messe [2]), which is why it was seen as an apt setting for the extensive empirical user studies.

Figure 1.1: The interactive wall project deployed at a public exhibition [3]

1.2 Example

Having the opportunity to be exposed to an interactive public display that is deployed in real-life application scenarios, made it possible for our research group to inspect and investigate the current system design deficiencies. The interactive wall project was established in 2011 [3], and throughout two years it has been used in multiple public exhibitions and fairs. The two-years public exposure span, made it possible to gather a valuable amount of user's observations regarding the interactive system design. One major and fairly common observation was that at certain positions the system stopped reacting to the interactive gesture, and a user relocation was necessary to return to the interactive spectrum. The

[1] www.worldusabilityday.de
[2] http://www.hannovermesse.de/
[3] http://blog.iao.fraunhofer.de/home/archives/1910.html

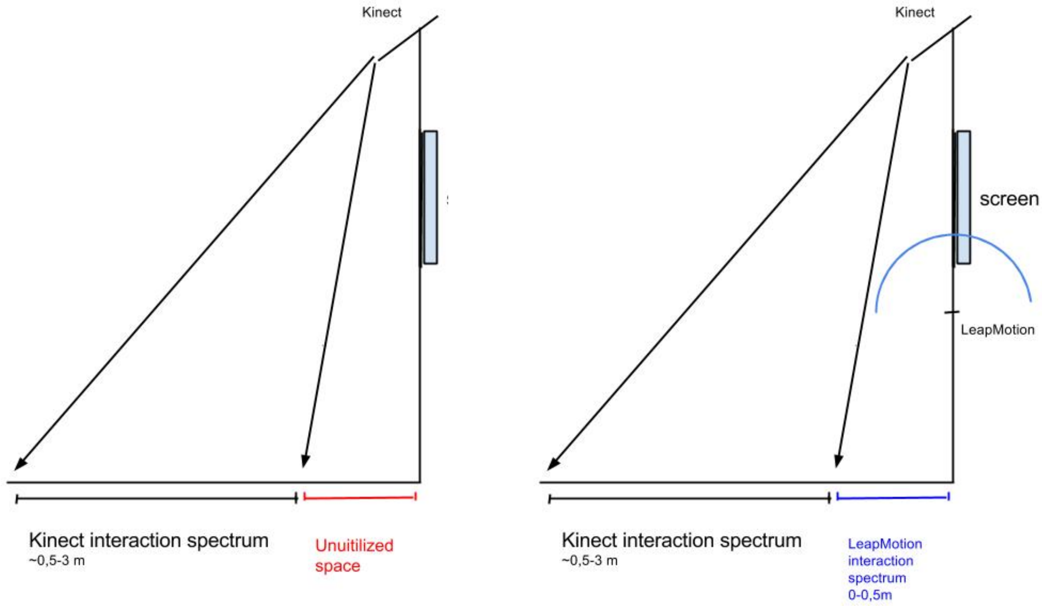
observation was also supported by the inspection of multiple video recordings where the project was deployed. The problem is graphically presented in Figure 1.2. The problem emphasizes a major system design vulnerability. It is obvious

(a) The system successfully responds to user's gestures

(b) System stops responding since user hand is not in the interaction spectrum

Figure 1.2: Real-life example exposing system's vulnerability in close proximities [3]

that the interaction space in front of the public display is not fully utilized. The observation actually exposed a breach in the system design, that is, the interactive wall is "blind" when it comes to interaction from close proximities. In a more general and abstract statement, it Acould be concluded that the public display is not fully distance-aware or the sensing capability is constrained with respect to various interaction proximities. The observation collides perfectly with the core of our research question and represented the perfect motivation to combine multiple input modalities to fully utilize the entire interaction spectrum in front of the display. Figure 1.3 gives an illustrative figure showing how the combination of multiple sensing technologies could present a solution for the exposed vulnerability. Thus, the determination to explore for answers for our research question is further thrust.



(a) Graphical illustration for vulnerable system design

(b) Graphical illustration for the proposed solution

Figure 1.3: Comparison between system design before and after solution deployment

1.3 Application Scenarios

Application scenarios that might benefit from such a setup would be generally described as any workspace that maps different functionalities or interaction modes in consonance to the user's relative distance from the system's screen. Such an application scenario may appear in different life situations for example in a public library or a book signing event where a public display is used for promoting a new book. To comfortably allow the user to scan through the book's content, the public display may allow the user to turn over the pages using left/right hand gestures, while zooming into a certain passage or paragraph could be mapped to a two finger pinch from a closer distance. Text highlighting can also benefit from the fact of having a touchscreen.

A more challenging application scenario may appear in a multi-user environment like in large exhibits for example, where the level of interest between a casual passer-by and an attentive user could be determined through the relative distance to the public display. Since it is obvious that people who are coming closer to the screen are more interested in the content, thus revealing a deeper level of information or even offering a special service would seem legible. On the other hand, just displaying a brief overview about the content might sound more suitable for a casual passer-by. In such a scenario a touchscreen would be most applicable, where attentive users try to enter some textual information like contact details, or even -and thanks to the widespread social networks- share a compelling content on the Internet.

A huge territory for application scenarios would be the gaming field as well. Since gestural-based interaction is already dominating the gaming console area, it would actually be the next smart move to combine multiple sensors as one akin interaction spectrum to deliver players a more enjoyable user experience. For example a first person shooting action game where the player gets to fight against enemies using bigger weapons like sword through bigger hand gestures, while at the mean time allow the user to enjoy accomplishing a more detailed and complex tasks like defusing a bomb or opening a safe just using fine grained finger movements.

Of course these are just sample preliminary conceptual ideas of how the merge between multiple sensing technologies might help in defining new interaction spectrums and how these new spectrums unlock the doors of creative content creation and greater, more amusing, more entertaining and even more realistic user experience. The contribution of this thesis will undoubtedly act as a vital competitive edge for various stakeholders such as content providers or information display owners. It is ambitiously hoped that it will provide the basic knowledge and framework that will guide to an alternation in the way of how interaction spectrums could be optimized.

1.4 Overview

Before inaugurating, a brief recapitulation of the chapters and sections provided in this thesis is offered in this section such that it acts as a reference for the reader. At the commencement of each chapter an introductory paragraph will be also furnished to easily inform the reader about the main contents of the chapter.

The first chapter initially provides the reader with an overview of the topic and the history behind public displays, including a section depicting the motivation behind this study. This part aims at shedding light on the importance of this research and how it can be made use of via future applications that can benefit from it once validating the results of our research.

Chapter Two presents the related work and background knowledge needed by the reader in order to establish a fundamental knowledge base for the chapters to follow. The chapter is divided into three main sections, each represents an area of interest that contributes in finding an answer towards our research question. In the first section of this chapter an overview over the different input modalities that are associated with public displays. The second area of interest is the gestural recognition. Review for the multiple classifications and taxonomies as well as input sensors is included in this section. Proxemic interaction and spacial division of interactive spectrum are the main concerns of the third section where the previous related work is explored.

Chapter Three offers a deeper insight on our first prototype application implemented. The chapter starts by giving a brief introduction of the concept of the study and how the inspiration behind it emerged. The chapter advances by providing a description for the apparatus used in this study. The apparatus is described on various levels, giving a more profound understanding for the software architecture, hardware setup and application's logic. Participants demographics and methodology used in conducting the study is further outlined in the details of the chapter. The results section gives a detailed discussion on how the data is structured and how the approach is used in the data analysis phase. Based on the findings, the chapter introduces an enhanced concept in the integration of multiple sensing technologies. The chapter is concluded by presenting system design implications that were derived from the study results as well.

Chapter Four provides the user with a closer look on the second prototype application implemented. The chapter starts by presenting the conceptual reasoning behind the study and how it serves as an extension of the first prototype. The apparatus used is also described in further detail, exploring the various software, hardware and working logic aspects. Participators' demography as well as the procedures used in experiment execution and data collection is presented. In the results section, the users' performance and system's evaluation is highlighted, describing how these findings imply a general gesture design framework and guidelines.

Last but not least, Chapter Five wraps up the an adequate summary of all the previous chapters, adding focused attention on shortcomings and limitations that were faced during the course of the study. The chapter also includes a section presenting the public exposure and various events where the system was already deployed and feedback was received by the audience. Finally, the chapter explores the potential forthcoming improvements and enhancement ideas as well as interesting application scenarios where the study outcomes could present a positive impact when deployed.

Chapter 2

Related work & Background

As the topic presented in this thesis is multidisciplinary taking into considerations various technical as well as general human computer interaction concerns, a diverse research in the related work was essentially needed. In this chapter the related previous work upon which our research was built is being presented. The dive into the related work could be divided into three major subcategories:

Public Displays:
: Focusing on the different input modalities associated with public displays

Gestural Recognition:
: Concentrating on its taxonomy, various input sensors and challenges.

Proxemic Interaction:
: Reviewing previous work related to proximity-based interaction and proxemic spatial division of interaction space.

Having covered these three subcategories should enable our research a solid foundation and a good starting point in choosing and combining the most suitable methodologies for our approach. By the end of each subcategory, it is briefly described how the research benefited from the previous effort made in the corresponding area.

The chapter also gives a brief introduction to the sensing technologies used, emphasizing the main features of the software development kits that helped in establishing the technical backbone of the prototypes.

2.1 Related work

2.1.1 Public displays

Deployment of public displays in urban life has witnessed an increase over the past few years. This expansion could be noticed in different everyday life aspects including public libraries [1], museums [36], shopping centers [29] or even as central information booth in city centers [32]. As public displays represented the first facility in the system under investigation, a deeper understanding for the different interaction modalities was seen as a necessity. The work of Möller et al. [28] declared the ten basic interaction techniques that take place in a public display environment.

Presence:

The presence of the user can be used as a trigger for an application that runs on a public display. User's absence may also be used for display turn off for energy saving purposes. The switching from off/on states of the public display may serve for awakening the user attention and attracting awareness. With the adoption of the proper sensor set this can be applied. Sensors may vary depending on the nature of the application and the ambient in which the public display is set. The Hello.Wall [33] project resembles this idea using RFID-based ViewPorts carried by the users, and upon approaching the wall the corresponding information is being displayed.

Body position:

In a very similar, but slightly different manner the exact position of the system's user may be used as an input modality. In order to extract the exact position with respect to the public display, pressure sensors or depth aware cameras maybe used. Knowing user's proximity allows the application to have a more complex output modalities that depends on the user's position. This is best illustrated in the work of Bayer et al. [6] who used a cylindrical public display and a camera to detect user's position. Content displayed on the wall followed the user around the cylindrical display which results in increasing user's attention as well as interactivity.

Body Posture:

Orientation and proximity of the body could also be used as an input modality for a public display system. Using the suitable setting of hardware sensors like 3D cameras, motion sensors or even low-frequency waves allows the public display to recognize body posture features like whether the user is facing the display or just passing by. The work by Vogel et al. [42] demonstrates this while focusing how the body postures could influence and helps to decide between implicit or explicit interaction.

Facial Expression:

The field of facial recognition is now booming in the industrial sphere. One of the best frameworks now available in the market was developed by the cooperative partner in supervising this master thesis, i.e. Fraunhofer. The SHORE [22] framework detects facial expression and hence allows the system to recognize certain feelings and emotions like happy, sad, or angry.

Gaze:

New complex eye-tracking algorithms allow the system to extract the full gaze path and not only the direction of gaze. Where this could be fruitful in measuring the exposure to digital signage, the technology could be more beneficial in crucial situation regarding security aspects. The work done previously by our research group addressed the problem of authentication of ATM systems in public using eye-tracking algorithms [11].

Speech:

Microphone arrays installed in public displays could also be used as input modality. The number and relative users' position could be estimated using the information supplied by microphones. Speech recognition could also be used as a technique that displays targeted advertisement based on specific keywords [34]. The Liberated Learning [4] illustrates the idea by having a system that displays lecture notes on public classroom displays once they are spoken by the lecturer.

Gesture:

While gestural interaction has witnessed a revolution in the gaming market, the same technique is also adopted for public display, specially regarding marketing and advertisement fields. The typical set of sensors used for this input modality is a camera-based sensor for detecting hand and bodily gestures. The *bubble game* system [2] presents a study on how interactive systems could influence the recall and recognition factors of items displayed on a public display.

Remote Control:

In case of distant public displays or too large ones, the need to remotely control the display rises. The usual interaction media in this case are mobile phones that nowadays embrace huge technical capabilities. Connections are typically established via Bluetooth or HTTP protocols. The work of Boring et al. [7] gives a practical example for the concept through allowing control of live video image streamed on the phone.

Touch:

Touchscreen technology has been in increasing use by the smart phone market segment. Consequently a large user base has established good understanding and a solid mental model regarding touch-based interaction. The "*City Wall* project [32] introduced the idea of having multiple users interacting with large public display in parallel.

Keys: All the above mentioned interaction modalities are new innovative techniques that resulted in extensive research. However the deliberate Keyboard and mouse combination is still valid and often desirably employable in some application scenarios. This combination is specially needed when the minimum effort in understanding the communication medium is wished. Brignull et al. [9] presented this technique in the system *"Opinionizer"* to assess the socialization behavior of users around public displays.

After the thorough inspection of the various interaction modalities, our research group came to the conclusion that our system consists of an overlapping between different input modalities. In the study presented in this thesis an overlap between presence, body position, gesture and touch input modalities is used to communicate interactivity to the system's user.

2.1.2 Gestural Recognition

Having explored the area of input modalities and settling upon the perfect combination that matched our research requirement, left us with the next area to explore. That is, the area of gestural recognition. Before proceeding with the research a deeper insight into the definition of gestural recognition, its hardware taxonomy, different algorithms, as well as the main challenges was the main motivation behind conducting further research exploring this area.

Definition

A fundamental goal for HCI systems that raised incremental attention specially over the past few years was to migrate "natural" means of interaction that are used among humans to the computer-human interfaces. Alongside with verbal communication, gestural interaction belongs to humans daily and most expressive interaction technique. Hence the need to interpret human gestures by computer systems arose over the past few years.
According to McNeill [26] gestural behavior is described as "movements of arms and hands which are closely synchronized with the flow of speech". The definition clearly emphasized the correlation between speech and gesture in human communication. While Luciani et al. [24] differentiated between different terms gesture, motion and action. In tier study they defined motion as the resultant movement of

physical objects (device attached or held by the human; including the movement of the human body itself). Whereas a gesture was defined to be more related to the mental model or concept of the performer. Actions were defined to be resultant from a set of gestures in a more symbolic or abstract level (for example drinking a glass of water).

Gestural Classification

Feldman and Rime [13] described a taxonomy for gestural classification in their book *Fundamentals of nonverbal behavior* that classifies gestural interaction into : symbolic, dietic, iconic, and pantomimic. Symbolic gestures are the ones that hold only one meaning, like the "OK" gesture. While dietic ones are commonly used to point or direct the receiver's attention to a specific target. On the other hand iconic gestures used to describe shape, size or orientation of an object, like the path of a moving car. Pantomimic gestures are most used when trying to describe how an invisible tool or object should be used.

In order to represent these classification to the the computer model it is highly dependable on the application scenario that is to be implemented. Vladimir et al. [31] has introduced a a taxonomy for the spatial classification of gestural models into two main domains: a 3D model-based and appearance-based. In appearance-based model the perceptual system tries to recognize the intended gesture through interpreting a sequence of visual images. Whereas in 3D model it is a two step approach. First a representation of the intermediate tool used for gesture production (human hand or arm) is modeled first. Then through the represented motion or posture the intended gesture could be inferred. For example some applications require simple interaction schemes such as just pointing, where the appearance based is more suitable. While other application domains require more complex attributes regarding hand posture and finger joints details. Figure 2.1 depicts the classification mentioned by Vladimir et al. Cadoz et al. [12] presented another gestural classification into three classes. The first class is called epistemic and is concerned with tactile or haptic interaction with the environment. Ergotic interactions rep-

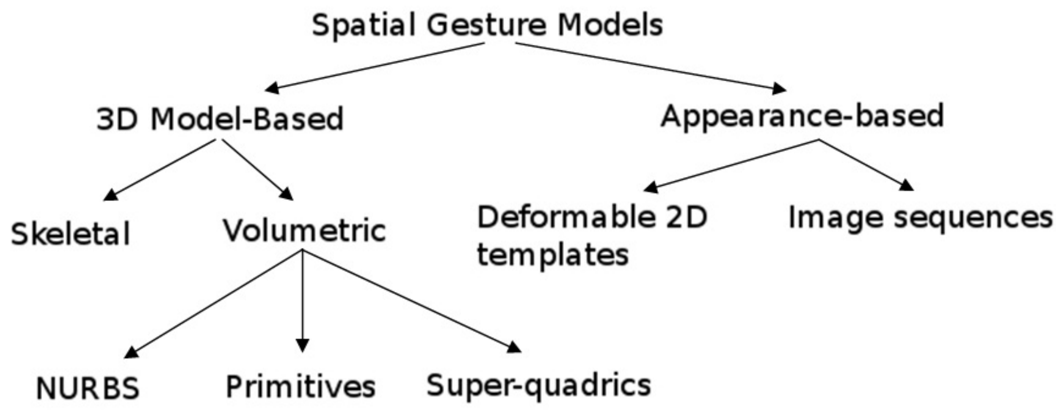


Figure 2.1: Spatial classification for gestural models [31]

resent the second class and are mainly deal with gestures that are responsible for manipulating virtual objects on the display. The third class groups the semiotic interactions where the communication of meaningful information takes place.

Input Sensors

To be able to use recognized gestures as an input for the application, a quantification process needs to take place. The process of quantification needs to be done through a hardware sensor, that is usually selected upon the application needs. And it was decided a study to inspect the different available sensors was seen to benefit our research purpose.

Bodily sensors:

Bodily sens0rs usually consist of external devices that are held in the user's hand or directly attached to user's body. Information is straightforward delivered to the system when the movement occur.

Wired Gloves:

Data gloves are used to detect and capture hand as well as finger movements. The main features delivered to the system are hand movement, position and finger bending. The first commercial product of wired gloves was DataGlove [45]. The biggest advantages of wired gloves are the high data rates delivery as well as the sharp precision. However the system requires a external devices for the interface and includes large number of cables that connects the glove and the computer, that may obstacle user movements [14].

Controller based:

With the usage of the matching software, the external controllers perform as extension for the body parts. Schloemer et al [38] have presented a sample project where the Nintendo Wii remote control was used to recognize gestures performed using the accelerometer technology. Another example is illustrated in the work of Jun Rikomoto [35] where accelerometer was also used in a wristwatch-type peripheral to detect gestures made by forearm and hand.

Visual sensors

Visual sensors are typically used to detect shapes and properties like color and texture. It can be seen in facial expression recognition systems, that analyse and detect human expressions through facial curvatures [30]. Another popular domain where visual recognition is widely used is the domain of hand and motion sensors , where the 3D position and posture of user's hands is detected and interpreted by the system [3].

Depth-aware Cameras:

Another interesting technique how depth maps could be generated is the structured light technique. Freedman et al. [15] presented in their patent the process. Abstractly described, the first step is projecting a predefined pattern of pixels-usually in a form of a grid- on the scene in sight. Followed by a second step to calculate the deformations of this grid after hitting objects present in the scene. Hence, the relative depth and surface information can be estimated by the vision system. The very interesting work of Izadi et al. [19] illustrated the construction of accurate 3D model for the perceived scene using a Microsoft Kinect sensor.They presented a novel tracking and reconstruction techniques.

Stereo Cameras:

On the other hand, stereo cameras-based vision system follow a different logic in estimating the 3D scene. The basic idea behind this paradigm is the presence of two different cameras separated by two a known distance called the baseline. The two cameras take a picture for the same scene. The generated two pictures are then post-processed computationally in order to extract depth information. Jenkin et al. [20] describe the different techniques alongside with its limitations in this area.

This comprehensive study for the gestural recognition techniques, its taxonomy and the various input sensors has enabled our research groups to categorize and further define the scope of the research. For our study we selected the Microsoft Kinect and the LeapMotion sensors to comprehend our gestural interaction requirements. Both sensors could be categorized to deliver a 3D model-based for the users engaged, and they both fall under the segment of visual sensing devices. While Microsoft Kinect is a depth aware camera-based system, the LeapMotion uses stereo disparity algorithms to define the depth information. Ergotic and semiotic gestural sets also raised our attention in defining the appropriate gesture set to allow virtual object manipulation and command execution on a public display respectively.

Challenges

One of the major challenges that face the deployment of natural user interfaces and gestural computing interfaces is the ergonomic factor. The question of comfort and muscular effort typically rises in the field of gestural interaction. As Boring et al. reports in his study in using mobile phones to control pointers on large public displays, that the long term usage of hand gesture as pointer for vertically mounted display may lead to an increase of a fatigue feeling in the user's arm, usually referred to as the *"gorilla arm effect"*. The same effect was also reported by Young et al. [44] in their comparison of various movements in specific workspace configurations while evaluating the performance and speed factors.

Fysh et al. [16] and Sparks et al. [40] also studied the health related problems that are associated with gameplay of the Nintendo Wii. With 44% the studies showed that hand lacerations were the most common injuries of the total incidents reported. Although studies mainly investigated the Nintendo Wii, they are not constrained to it and could be generalized, according to Sparks et al.

These results were considered while designing the system, where the alternating switch between interaction modalities would provide a constant switching between arm/hand usage and finger usage, that should result in temporary muscular relaxation when the switching frequency is not too high.

2.1.3 Proxemic Interaction

Since our system was concerned with assigning different input modalities depending on the user proximity, the third area of interest that required reviewing for the previous related work, was the proximity-based and multimodal interactions.
Edward Hall introduced in his book *The hidden dimension* [18] the concept of personal space with respect to different proximities where he differentiated between four basic interaction spaces.

Intimate:

The physical space ranges around 46 cm from the individual and is dedicated for very familiar interactions. Family members, loved ones, close friends and pets belong to the group of people that are comfortably welcomed into this interaction zone. Affectionate feelings are usually expressed within this ranges unless communicating with strangers the individual might feel threat or danger.

Personal:

The space extends from 0.46 cm- 1.2 m where the existence of colleagues and friends is desired. The space is usually reserved for greetings among two individuals , however and depending on the culture another form of greeting (kiss on the cheek) might penetrate the boundaries of the interaction space.

Social:

Extends from 1.2-3.7 m and is usually associated with formal and business meetings or non-personal interactions like paying at supermarket cashier.

Public:

From 3.7m and beyond the space for public communication and is not reserved for any particular communications.

Streiz et al.extended the findings of Edward Hall to the public display interaction medium through their work in *The Gossip Wall* [41] and *Hello.Wall* [33]. In their work they used an unusual medium to display and convey information to the user interacting with the display. The display consisted of multiple light emitting cells, that worked cooperatively to display a certain light pattern. Depending on the user's identity and user's proximity the light pattern changed accordingly. User's identity could be recognized through an external handheld device called the ViewPort that included an RFID tags. Tow RFID readers are embedded into the lower part of the public display allowing the detection of RFID tags of the passer-by. Through the RFID technology used, the public display could identify 3 interaction zones.

Ambient Zone:

Where general information is displayed, regardless of the identification of a particular individual.

Notification Zone: If a passer-by shows more interest and approaches the public display, the person enters the notification zone and being identified using the embedded RFID tag. In this zone, light patterns change to convey some information that could only be understood and realized by the engaged user.

Interaction Zone: This zone becomes active, once the individual approaches very close distance to the public display. In such case, Streiz et al.present the concept of "borrowing an artifact". The public display starts to communicate more personal information to the user on the dedicated ViewPort. Depending on the user authentication level, decoding, downloading or only browsing the information is made available. Figure 2.2 gives graphical representation for the different interaction zones.

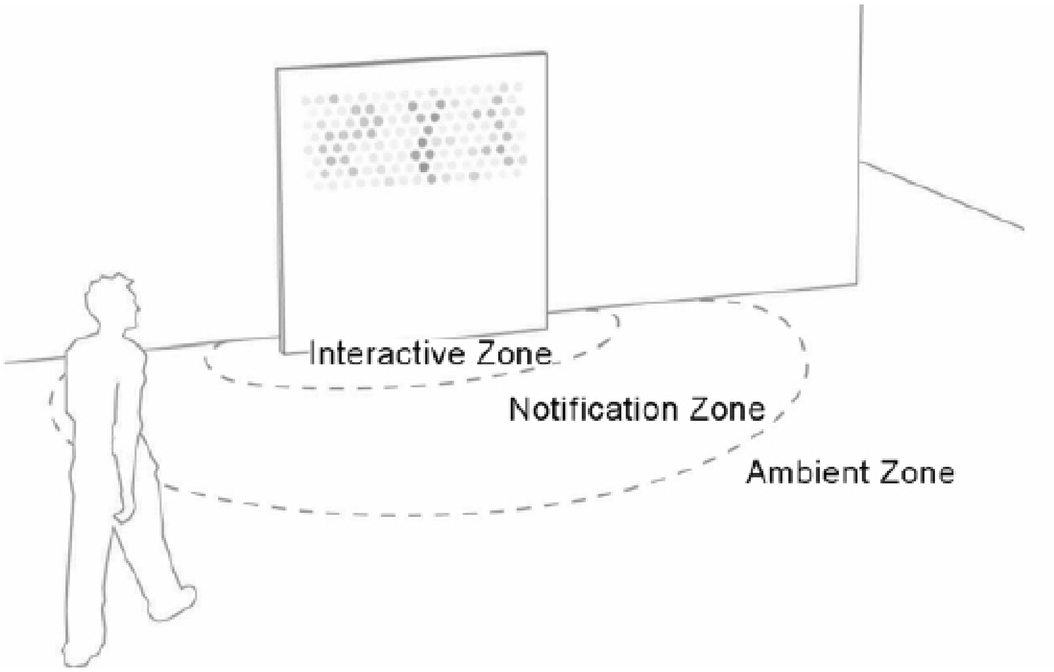


Figure 2.2: Interaction zones defined by Streitz et al. [41]

Brignull and Rogers [9] presented another spacial arrangement in their work regarding the *Opinionizer* system. The system consisted of a public display supplied by a keyboard and is to be deployed in public events and exhibits. Using the keyboard, users get to write their opinion concerning the ongoing event and this was shown on the public display. As a result of observing user's behavior around the public display system, Brignull et al. could separate the physical space around the public display into three main categories of activity-related spaces as described below. Figure 2.3 depicts the space division according to Brignull et al.

Peripheral awareness spaces
: In these spaces event participants usually are aware of the existence of the display, but don't know much about it or how to operate it.

Focal awareness spaces
: Engagement with the public display usually starts in these spaces. Participants usually try to know more about the public display including topics like how to operate it, and what is it used for.

Direct interaction spaces
: A total engagement is present at this stage, where an individual -or a group- are directly acting with the public display.

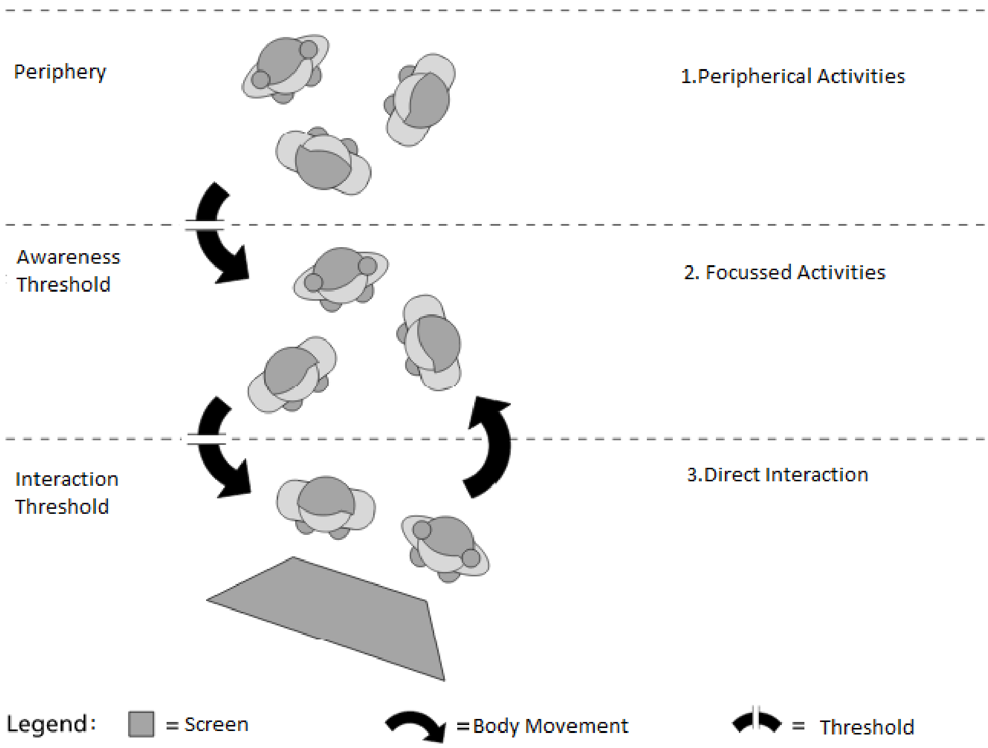


Figure 2.3: Interaction zones defined by Brignull et al. [9]

Vogel et al. [42] extended the work of Streiz et al.to further explore proximity-based interaction. Vogel et al. differentiated their work from previously mentioned framework by extending Streitz et al.'s *"interaction zone"* into two further zones, namely a *"subtle"* and *"personal"* Interaction areas. They also presented a more generalizable notions for the *"notification zone"* to be *"implicit interaction zone"*. The four interaction zones are graphically depicted in 2.4 and are described as follows:

Ambient Display Phase:

Considered to be the default phase in which the public display operates. General information are displayed randomly giving an overall view for the regular passer-by.

Implicit Interaction Phase:

Once the user approaches a closer distance to the public display, it is considered as an the implicit interaction phase. The system recognized the body posture and orientation in order to reach a solid conclusion about the user's openness to receive information. The content on the display should implicitly adapt the content to be present a general information or a general notification.

Subtle Interaction Phase:

Further approaching the public display and providing an implicit indicator like pausing in front of the screen for several seconds, should signal the system to enter the subtle interaction phase. In this phase a more detailed information regarding the previous general information/notification should be presented to the user. Personal information that are dedicated to the identified user could also be presented if no confidential data is to be exposed.

Personal Interaction Phase:

When the participant is at the closest proximity possible to the public display, the personal interaction phase is activated. In this phase the user should be able to interact with the public display through a touch gestures. Due to the extreme close proximity to the display, the user's body most probably occludes the information displayed, and hence information presented could be rather personal.

Having carefully investigated the various frameworks proposed by the previous related work, made it clear for our research group that our study scope is to spatially segment the interaction space in front of a public display but with a greater focus on seamless gestural interaction. In order to come up with the appropriate gesture set that best suited out study needs, we followed a similar approach to the work presented by Wobbrock et al. [43] in investigating gestures that allow the connection between mobile devices and interactive surfaces.

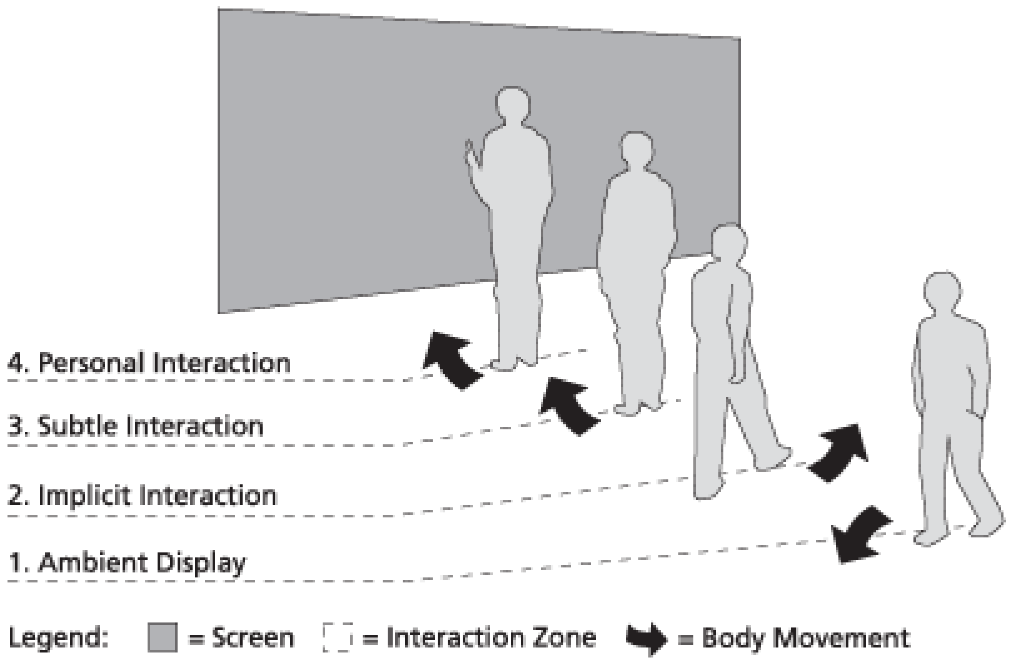


Figure 2.4: Interaction zones defined by Vogel et al. [27]

2.2 Technology used

2.2.1 Microsoft Kinect

Microsoft Kinect, that was first developed under the name of *"Project Natal"*, got first introduced to the world on the 1st of June 2009. The hardware peripheral was mainly developed as an extension for Microsoft's game console Xbox 360. Microsoft Kinect allowed its user to enjoy the interaction experience without the need for an extra controller, but through natural user interfaces that consisted basically of fully or partly bodily gestures [1]. Microsoft officially launched the first commercial generation of Kinect on November 2010 specially designed for the Xbox. On February 2012, the Kinect for Windows was launched to the market to be compatible with the windows system, and targeted to broaden the user base in the field of Windows application developers and companies. Kinect for windows is bundled with a very powerful Software Development Kit (SDK) that allows

[1]http://blogs.msdn.com/b/mssmallbiz/archive/2010/06/16/microsoft-kinect-project-natal-announced-and-new-xbox-360-released.aspx

developers to communicate with the Kinect via .NET Framework [2].

Working logic & Architecture

The Kinect sensor consists of four main components. An infrared emitter (IR emitter), an infrared sensor, a color sensor and a microphone array. As it can be seen in Figure 2.5, the devices architecture could be described as follows:

IR Emitter:
: The emitter's job is to emit infrared light beams in order to "explore" the area in front of the Kinect sensor.

IR Depth sensor:
: The reflected IR beams bouncing back from the objects are received and interpreted by the IR sensor. The reflected beams could be converted into depth information and the distance between the object and the Kinect sensor can be calculated. In such a technique extracting depth image information is made possible.

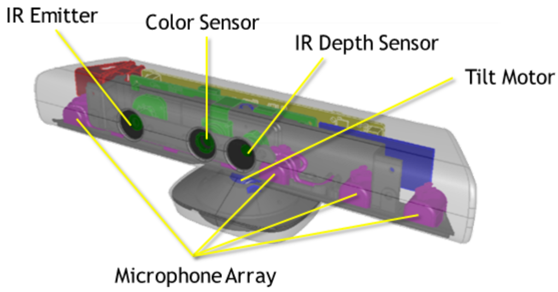
RGB color camera:
: The RGB camera has got two different working frequencies and correspondingly two different resolutions. It can work at a pace of 15 frames per second at resolution 1280x1024, or 30 frame per second at resolution 640x480.

Microphone Array :
: The multi-array microphone is composed of four microphones, that are dedicated to capturing sound information. Through capturing audial information the location of the sound source and the direction of sound wave can be estimated. This comes an additional for simple sound recording feature.

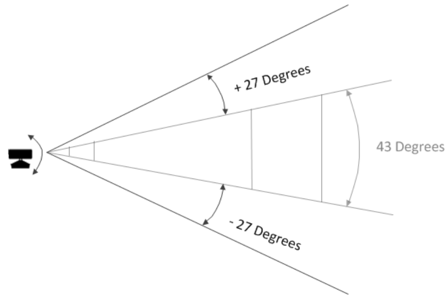
Important to mention that the sensor enjoys a 43 degree vertical x 57 degree horizontal wide range of sight, in which the performance peaks. The tilt motor attached to the device gains the sensor a ± 27 degrees of freedom to enable better

[2] http://blogs.msdn.com/b/kinectforwindows/archive/2012/01/09/kinect-for-windows-commercial-program-announced.aspx

adjustment and user tracking as well as larger interaction space. The utility was applied to ensure the correct positioning combination with the LeapMotion sensor to cover the entire area in front of the public display. An illustrative description is given in Figure 2.5.



(a) Microsoft Kinect's sensor components



(b) Microsoft Kinect's tilt motor and the available degrees of freedom

Figure 2.5: Microsoft Kinect hardware components [3]

[3]http://msdn.microsoft.com/en-us/library/jj131033.aspx

SDK features

The Kinect API allows the developer to receive multiple data streams from the connected Kinect sensor. Four different types of Kinect data streams can be generated: a color data stream, a depth data stream, a skeleton data stream and an audio stream. For each of the data streams available a dedicated class is provided by the API. The class allows the user to determine the data format, the frame rate and the resolution of the pixel data. The default working settings for the Kinect sensor is 640x480 resolution at a frame rate of 30 frames per second. Once the skeletal tracking is activated the sensor can detected up to six users, of which two simultaneously users can be active. The default settings is a random approach to select any two users, however this could be overridden to select the nearest two users. In order for the user to be correctly recognized by the Kinect sensor, 20 skeletal joints are by default needed. However Kinect offers two different operational tracking modes:

- **Default or standing mode:** needs 20 skeletal joints to track the user.
- **Seated mode:** needs only 10 skeletal joints to track the user.

This feature was fully utilized while developing our prototypes systems. Examples where the seated mode was valuable included scenarios where younger audience participated in the hamster game, a full skeletal track was not available, but through focusing on the upper 10 skeletal joints the user was correctly recognized. Another interesting feature that was heavily exploited during the implementation of the prototype, and it was even vital for the merging operation with the LeapMotion sensor was the near mode operational mode provided only in the Kinect for Windows devices. Important to mention that this mode is *NOT* available in the Kinect for Xbox sensors, since it only operate in the default mode. A comparison between both modes is given in Figure 2.6

[4] http://blogs.msdn.com/b/kinectforwindows/archive/2012/01/20/near-mode-what-it-is-and-isn-t.aspx

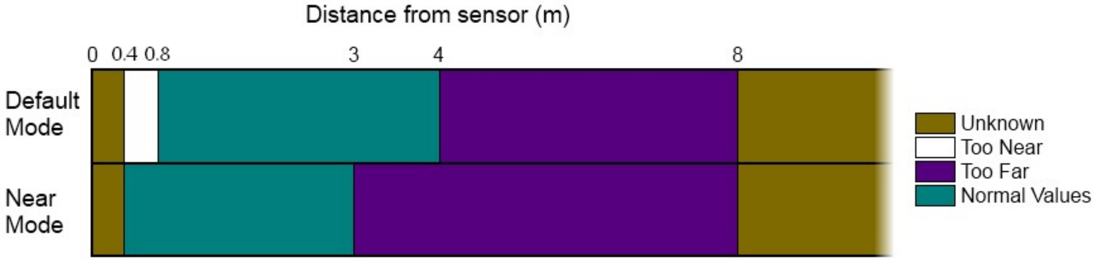


Figure 2.6: comparison between Default and Near tracking modes [4]

2.2.2 LeapMotion

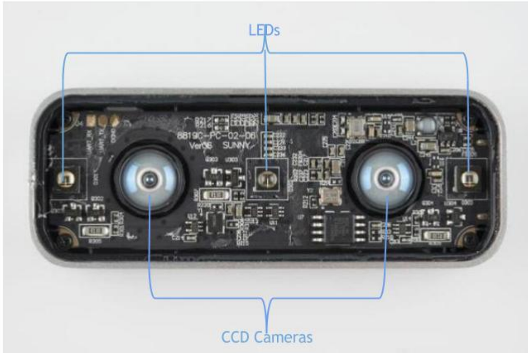
The LeapMotion is a 80x30x10 mm^3 rectangular USB peripheral that creates a 3D interaction space of 8 cubic feet. The device is produced by Leap Motion Inc. and was recently introduced to the market on the 22nd of July, 2013 [5]. Leap motion allows touch-less interaction to occur with hands, fingers or pointing tools (like a pen or a chopstick). It comes as a stand-alone peripheral that can be plugged into any computer that supports a minimum requirement of USB 2.0 or USB 3.0. The sensor can be used for various types of application that range from simple mouse control to complex gaming console on a regular PC. Buyers of LeapMotion get an access to the Airspace market [6]. It is a software market where LeapMotion applications are hosted and could be bought or downloaded for free . In December 2013 LeapMotion reported a partnership with Hewlett-Packard (HP) [7]. indicating the embed of the sensor device as an all-in-one feature into eleven of HP models

Working logic & Architecture

The small in size but great in capability motion sensor consists mainly of three main components : 2 CCD cameras, 3 Infrared Light LEDs and on board processing unit. Due to the patent rights that LeapMotion claimed, not much about the algorithm on how it works is given. Nevertheless, the basic simple idea behind the working logic is stated. The system could be easily mapped to an analogy to

[5] http://blog.leapmotion.com/post/56106835762/shipping-and-order-processing-is-ongoing
[6] https://www.leapmotion.com/product
[7] http://www8.hp.com/us/en/ads/envy-leap-motion/overview.html
[8] https://learn.sparkfun.com/tutorials/leap-motion-teardown/all

(a) LeapMotion sensor from outside



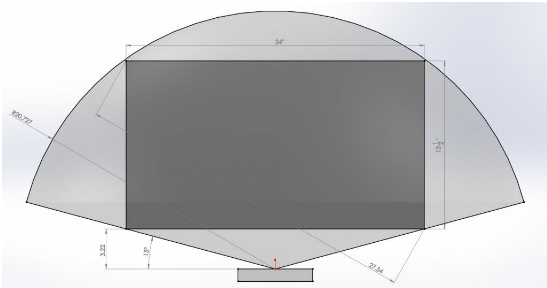
(b) LeapMotion components from inside

Figure 2.7: LeapMotion sensor components [8]

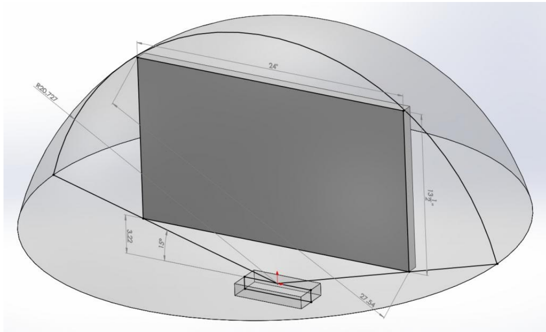
the human vision system. The two CCD cameras in this scenarios work as the eye, each capturing the same image but from different perspective or angle. While the LEDs could be mapped to the fill flash light in this scenario, where they just illuminate the interaction spectrum for the CCD cameras to take a clear picture. The on-board chip could be mapped to the human brain, where the processing of the images happen and the 3D model is constructed through the two 2D captured frames. However, not all the processing is done on the chip, since very complex mathematical model is being processed, part of it is done through the software supplied by the LeapMotion SDK on the host device. The device architecture presented in figure 2.7 to provide better understanding.Having said this, it is now important to explain how big the interaction spectrum is and how it looks like. The Leap creates a 3D interaction space of about 8 cubic feet. It could be best described as a cubic volume of space shaped like the top forty percent of a one-meter beach ball. Maximum range is roughly about 1 meter [9]. Figure 2.8 gives an illustrative clarification for the interactive spectrum.

[9]https://forums.leapmotion.com/forum/general-discussion/general-discussion-forum/434-the-unofficial-leap-faq?420-The-unofficial-Leap-FAQ

[10]https://forums.leapmotion.com/forum/general-discussion/general-discussion-forum/1058-technical-specifications



(a) Frontal-view for the coverage area



(b) Side-view for the coverage area

Figure 2.8: 3D representation for LeapMotion's interaction spectrum [10]

SDK features

LeapMotion works on a frame basis. Each frame holds a data that represent potential objects in the range of sight. The frame rate can be adjusted to one of the following modes:

- **Balanced mode:** with a maximum data rate of 120 frames per second.
- **Speed:** with a maximum data rate of 230 frames per second.
- **Precision mode:** with a maximum data rate of 60 frames per second.

This property was used to set the leap to the precision mode in our prototype in an attempt to minimize the frame rate between the used sensor as well as to gain the highest precision possible.
LeapMotion comes with a very compelling software development kit (SDK), that allows tracking the following objects: hand, finger, tools and predefined set of gestures. In addition to the potential pointing objects, the the API also provides the possibility to compare successive frames in order to detect whether a general motion occurred or not. General motions detected include: rotation, scaling and general translation. The API represents each of the pointing objects through a model that posses a set of attributes assigned to it. Through the values of these attributes across the different frames, the object's motion can be detected and a gesture can be recognized. The LeapMotion API supplies developers with a hand

the hand model allowing them to manipulate and extract data about the following attributes:**hand palm position, hand palm velocity, hand palm normal, sphere center and sphere radius**. The greatest limitation the the hand model representation that the LeapMotion can not detect whether the hand is right or left. Each of the recognized hand objects contains an array of fingers assigned to it. If no fingers are visible or the hand is clinched, then the list is empty.
Finger and tool objects are modelled using the same set of attributes as they exhibit the same behavior with respect to the sensor. Both objects are modelled using the following attributes:**Length of object, width of object, direction, tip position, and tip velocity**.

The API also obtains a predefined set of gestures that can be detected for all the previously mentioned pointable objects. The recognized movement patterns are circle, swipe, Key tap (vertical tap) and screen tap (horizontal tap,perpendicular to the screen surface).

One of the most extensively used features in our project, is the *"interaction box"* property provided by the LeapMotion SDK. The feature allows the user to define a virtual interaction space in a form of a rectangular prism. For each of the previously mentioned pointing objects that are present within the defined interaction box, the feature allows the developer to extract the object's coordinates. Moreover, the API provides various functions that help to transform the raw extracted LeapMotion's coordinates into 2D or 3D application's coordinates. Visual presentation of the rectangular prism-shaped interaction space is given in Figure 2.9.

[11] https://developer.leapmotion.com/documentation/

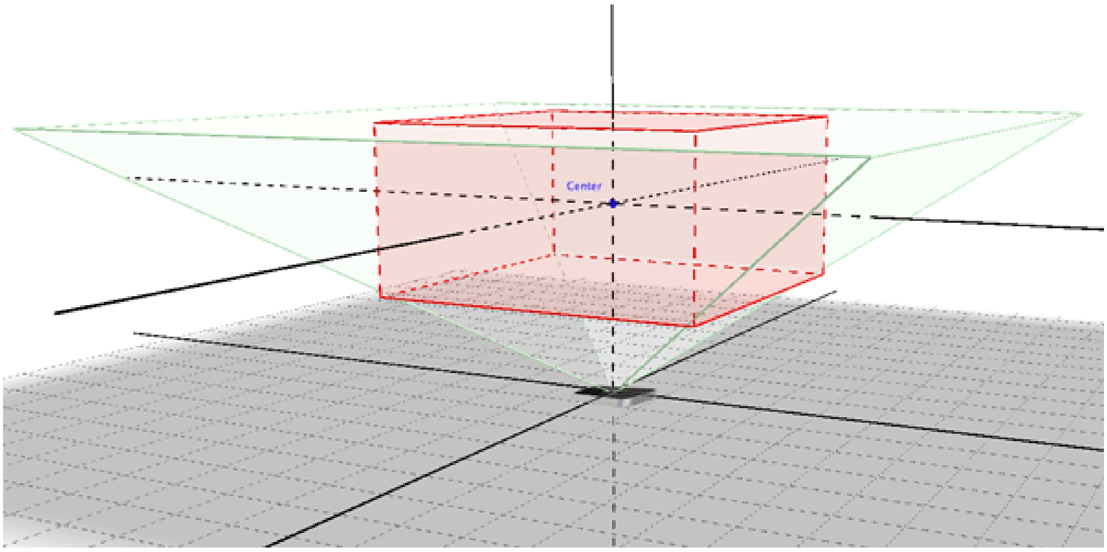


Figure 2.9: Visual representation of the rectangular prism-shaped interaction space supplied by LeapMotion's API [11]

The *Pointable* class present at the standard SDK provided by the LeapMotion, offered a very practical and interesting feature. The feature is called *"Touch Emulation"*, and offers the user the illusion to have a virtual touch surface. This virtual touch surface is located roughly parallel to LeapMotion's x-y plane, however the touch emulation accommodates the user's finger or hand position along the Z axis as well. Simply explained the API divides the physical perpendicular space in front of the X-Y plane of the LeapMotion into three different interactive zones, a *"Hovering zone"*, a *"Touch zone"* and a *"None"* zone. When ever a pointable object (Finger, hand or tool) is present in front of the x-y plane, it could be easily extracted through the Z coordinate in which interactive zone this pointable object is present and thereupon the application's logic and UI could be correspondingly adapted. The API uses a normalized values scheme called the *"touch distance"*, in other words it assigns masked values for the real raw data in order to simplify the decision. The *"touch distance"* ranges from +1 to -1. Whenever a pointable object enters the Hovering zone, it is assigned the value of (+1) and the value starts to decrease along side the decreasing direction of the Z axis, i.e. nearing the touch surface. Once the the pointable penetrates the *"Hovering"* zone and enters the *"None"* zone, its value becomes 0. If the pointable proceeds further, the value continues to decrease but never exceeds -1, expressing it is now present in the *"Touch"* zone[11]. Figure 2.10 gives a more illustrative graphical description to how the touch emulation feature works.

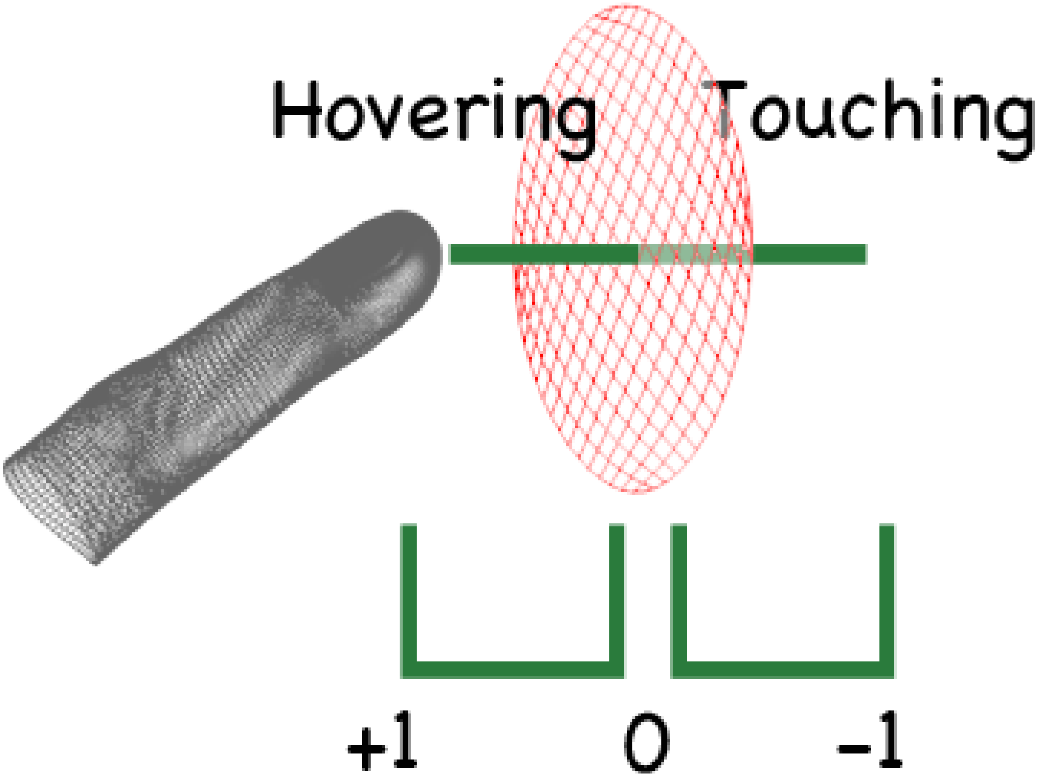


Figure 2.10: Touch Emulation feature supplied by LeapMotion's SDK [11]

[11]https://developer.leapmotion.com/

Chapter 3

Preliminary Study

This chapter focuses mainly on our first prototype implemented. It gives a detailed description to the apparatus used,how the data is logged and a characterization for the user base participated in this study. The chapter discusses the results and implications in two separate sections.

3.1 Concept

The basic focal point of this preliminary study is to explore and give a first judgement on the idea of fusing multiple, physical interaction spaces together in order to come up with one consistent, smooth and seamless broader interaction spectrum. In order to test the effects of such a combination, the appropriate environmental ambiance had to be entrenched first.

Accordingly the pertinent and most applicable hardware sensor technologies that are sensitive to different proximities were needed to be selected. After deliberate inspection for the different currently available sensing technologies in the market, and after thorough research in the previous related work, the decision was taken to limit the number of hardware sensors to two. The famously widely used Microsoft Kinect topped the list of the available resources. Microsoft Kinect already possessed all the desired attributes that fitted our research requirements.

It has a clear-cut efficiency in tracking the body skeleton as well as the ability to determine the current depth of different objects in the available range of sight, which was in addition to its suitable spatial coverage area that ranged from approximately 0.5-3 m (when working in the near mode)the two main crucial criteria in our selection process. Kinect was also extensively used in previous work and it went through a colossal amount of inspections specially in the area of public displays, which easily flourished the backbone for our research and resembled a solid starting point. Not to forget, Microsoft Kinect comes with a very powerful Software Development Kit, that enabled us to manipulate and extract the input data correspondingly.

The second hardware sensing technology that was seen as a perfect fit upon our selection criteria, was the new, off-beat, freshly-introduced to market LeapMotion. Although the introduction of LeapMotion to the technology market raised up the question whether this new sensing technology would be a killer to the already dominating one, namely the Microsoft Kinect, it was seen by our research group to be the exact counterpart of that. LeapMotion was mainly selected for its novelty in motion tracking and its high capability and extremely fast gesture recognition algorithms used. By the time this research started LeapMotion was still it the beta version phase, nonetheless we were lucky enough to be one of the few first research institutes to receive one of the rare innovative devices for our research needs. Thus, its availability and cost-efficiency attributes contributed also to our main selection benchmarks. However the essential, decisive and compelling property upon which the selection decision was taken, was its spatial coverage area, that ranged from 0m-0, 5m , which meant that only from the device technical specification the entire range in front of a public display could be covered through the simultaneous combination of LeapMotion together with Microsoft Kinect.

After such a comprehensive study and selection for the most suitable hardware sensors, the environment was almost ready to inspect the first kind of applications archetypes, specifically the real-time feedback type. The basic application scenario and probably the most obvious one is a mouse cursor control application that goes through both interaction spectrums. Users should be able to manipulate and maneuver the cursor's position through hand or finger gestures with respect to their position to the public display, i.e. whether they are located in Kinect or LeapMotion coverage areas respectively. In an extended version of the prototype application users should also be able to employ other cursor functionalities like single click, double click and right click through-out the different interaction zones.

3.2 Apparatus

In order to test such an application scenario, a more entertaining, enriching and enjoyable content needed to be created in order to spare users the boredom and to disguise the real intention behind the prototype. In this fashion the idea of having a mouse-controlled game would satisfy this condition. Just like that the first conceptual design behind our first prototype emerged, through having a labyrinth-type of maze, where the user needs to guide a hamster through hand or finger gestures across the different levels in order to reach the final level where the hamster finds a desired nut. Having our first prototype masked in a game application was also extremely suitable since the application was to be tested at two major events, where a lot of children are present, ergo the application will manage to attract and tempt a lot of participants. Therefore it constituted afterwards the base for our data analysis. The maze game consisted of six different maze levels, where the level of difficulty steadily increases by introducing more windy and narrow paths in an attempt to test how the different sensor devices will perform in accordance to the level of difficulty. Whenever the game player faultily hit against a wall, the game is restarted at the first level.

The application architecture could be simply described as a simple client-server application. The server side in this scenario is a locally installed server, where the various different events (starting new level, hitting a wall, hitting goal. . etc) are logged for detailed post-processing. The client side consists of both sensors as input modalities and the computer screen as output modalities. Input data are perceived by input modalities, and raw data are passed to the server. The server logs the interaction events and executes the application logic to move the mouse cursor accordingly. The output of the operation is displayed back to the user through the computer screen. The front end for the game prototype was implemented using the Adobe Flash SWF technology, whereas the back-end was developed using C# programming language with the help of libraries from the Software Development Kits supplied by both Micrsoft Kinect and LeapMotion. The main software model behind the application's logic could be explained clearly by stating that a general interaction handler with two simultaneously dedicated event listeners is defined at the application start. Both event listeners are working in the background of the application and have very similar working logics, i.e. they both take as input the raw 3D hand/finger positions and benefit from a transformation algorithm that maps these raw data into application coordinates. With the help of the default Windows library they both send mouse events to the operating system to move the mouse cursor to the newly calculated position. In case of multiple users are present in front of the public display, the Kinect event listener is responsible to track only the further most user to avoid any confusion. Once the tracked user has been determined, the second level of prioritization is resolved through tracking only the further most hand. Important to notice, that once a hand/finger is detected by the LeapMotion sensor, the Kinect event handler is being deactivated, allowing only the further most user to interact with the game. Hence completing the prioritization series of users in control. The hardware used for this user study consists of a 27 inch Apple iMac computer running 64-bit Windows 7 OS. On top of the iMac a Microsoft Kinect was mounted and in front of it the beta version of LeapMotion sensor was placed granting a full area coverage as depicted in Figure 3.1.

(a) The hardware setup for the hamster game

(b) A screenshot for the hamster game

Figure 3.1: The hamster game

3.2.1 Data Logging

Data were automatically logged on the deployed local server. Important to notice that the server was intentionally locally deployed in order to avoid any time delay that may occur due to data traffic congestion or other unexpected network problems. As our logging environment needed something light-weighted, fast and extremely scalable and needless to say cost-efficient, an open source document-based database system was seen to fit the requirements. Since Mongodb supplied all the above mentioned criteria, it was used as our main database. Seeing that data was logged in a JSON-like format, a JSON-like query interface was used to retrieve desired pieces of information. Events logged possessed the following attributes:

User: To keep track of the current user ID.

level: keeps track of the level where the event occurred.

date: Timestamp to make sure events occur in the right order.

type: Type of event occurred. Holds one of three values: Wall, goal or startNew.

x: The x-coordinate on screen where the event happened.

y: The Y-coordinate on screen where the event happened.

A sample snippet of the logged data and a sample query is given in Listing 3.1. Data were gathered separately for both events and merged together afterwards for a collective, more generic inspection. As results analysis was done offline a simple Java framework was programmed to extract the desired feed for the analysis.


```
{ "user" : 0 , "level" : 1 , "date" : { "$date" :
   "2013-06-20T13:49:04.266Z"} , "type" : "wall" , "x" : 576 , "y" :
   231 }
```


Listing 3.1: Sample event logged in Mongodb

```
*get all the events where level=0 or level =1 and user =101 and only
   from goal to goal
{ $or:
 [
       { "date" : {$gte:{ "$date" : "2013-06-20T15:20:05. 000Z"},
       "$lt":{ "$date" : "2013-06-20T16:30:00. 000Z"}},
       "level":0,
       "user":101,
       "type":"goal"
       }
       ,
       { "date" : {$gte:{ "$date" : "2013-06-20T15:20:05. 000Z"},
       "$lt":{ "$date" : "2013-06-20T16:30:00. 000Z"}},
       "level":1,
       "user":101,
       "type":"goal"
       }
 ]
}
```

Listing 3.2: Sample query in JSON format

3.3 Participants

As mentioned above, the user study took place in a public exhibition that mainly attracted families with their children to explore the modern innovations in the different technology sectors. A total of 113 participants were engaged in the study, of which 52 completed a qualitative study and hence will be the main focus group. Out of the 52 participants, 31 were males and 21 were females, and the average age group was 20-29. The majority(33) of the users had already gained some experience with gesture based console games (e.g. Nintendo Wii, Microsoft Kinect, PlayStation Move). Participants constructed a diverse community from proficient IT knowledge like computer scientists, graphic and media designers to basic IT knowledge like high school students, psychologists and workers in the finance sector. The ethnicity dimension was not very much explored by this experiment since it was only limited to the fair visitors who, more or less, live in Stuttgart area.

3.4 Procedure

The experiment took place over two days during the course of two different public exhibitions, videlicet the Sommerfest organized by Fraunhofer Institute and "Tag der Wissenschaft" organized by the state of Stuttgart held at the main public university campus. In the first day the experiment ran for five consecutive hours, while on the second day it ran for six consecutive hours. The study was deployed in the wild, which means it was deployed away from controlled, lab-based environmental conditions but the exact contrast it was arranged to run in the most natural and authentic, real world contexts to observe how well the sample application will integrate with user's natural behaviour towards the different sensing technologies, and therefore gaining our results extra validation and verification credibility. The experiment was structured as a between-subject study, where the participants were randomly assigned to one of three experimental groups. Each group was then introduced to one of the three different conditions under inspection. The first group used only the Kinect, the second group used only the LeapMotion whereas the third group had the LeapMotion and the Kinect both simultaneously in action. Both the first and second conditions were aiming at giving more accurate practical results about the comparison between both used motion sensing devices regarding efficiency and precision as well as task completion time aspects when each is used individually. The third condition intended to test the effects of merging both devices and how this will affect the user's performance besides evaluating how the optimization of the full 3D space could influence user's movement. For example players can use the further regions (Kinect) for easier maze levels and with increasing level difficulty approach a more near distance from the public display (LeapMotion) to achieve higher levels of accurateness. Booth visitors were welcomed to play the game as long as they liked, while our local server logged for each individual user the number of levels completed, the level completion time, whenever the hamster hit the wall (error) and which mode (Kinect or LeapMotion) was used. Upon finishing the experiments users filled out an SUS (System Usability Scale) [10] questionnaire to assess the perceived experience with the system. The questionnaire can be found under Appendix Section. A

3.5 Results and Implications

Before commencing into the details of the results and implications, it is valuable to give a brief introduction how the data is structured. Two types of data were gathered during the course of this study, namely quantitative and qualitative date. The quantitative data were mainly concerned with the performance evaluation of the application, whereas the qualitative data summarized the user experience and level of enjoyment and satisfaction that was delivered to the player.

3.5.1 Quantitative Analysis

Three different variables composed the main pillars for the quantitative data analysis. The condition under inspection (Kinect, LeapMotion, Leap+Kinect), the user, and the level (difficulty) the user is currently playing. Having these different perspectives into consideration , the Java framework furnished per experiential condition an excel sheet for each individual user indicating how the user performed in each level. Performance evaluation attributes included the total number of trials the user used in each level, the number of failed trials, the number of success trials and the average time the user used per successful and failed trial as well as the total engagement time with the system. Aggregating the results for all the users, made it available to asses the effectiveness for each condition from two major angles: the average Task completion time per level, and the average user progression through the different levels through inspecting the success and error rates. A sample individual table is given in Table 3.1

Condition: Kinect				**User: 101**	
Level	**# trials**	**# Failed trials**	**# Success trials**	**Avg. Time Success in sec**	**Accuracy %**
1	20	13	7	6	35
2	7	4	3	12	43
3	3	0	3	12	100
4	3	2	1	16	34
5	1	0	1	14	100
6	1	1	0	0	0
Total interaction time	1450				

Table 3.1: Sample data gathered for an individual user

Task completion Time

For all the different conditions the average task completion time per level over all users who participated in this level were accumulated and plotted in Figure 3.2. After careful observation to the plotted results, one can easily spot three different graphical regions. For level one and two, Kinect users scored the lowest task completion time(first region), whereas for the next difficulty, i.e. levels three and four users of the combination between LeapMotion and Kinect scored the lowest task completion time(second region), while LeapMotion users scored best in level five(third region).

In order to test the significance of the presented data, a one-way Analysis of Variance (ANOVA) [37] test was executed and the p value was inspected. Although the observations are still interesting , but no significance could be found. ANOVA results for the different levels came out [$F(2, 98) = 0.33$, $p = 0.72$], [$F(2, 82) = 0.776$, $p = 0.464$], [$F(2, 64) = 0.618$, $p = 0.542$], [$F(2, 33) = 0.015$, $p = 0.85$], [$F(2, 17) = 1.496$, $p = 0.252$] respectively.

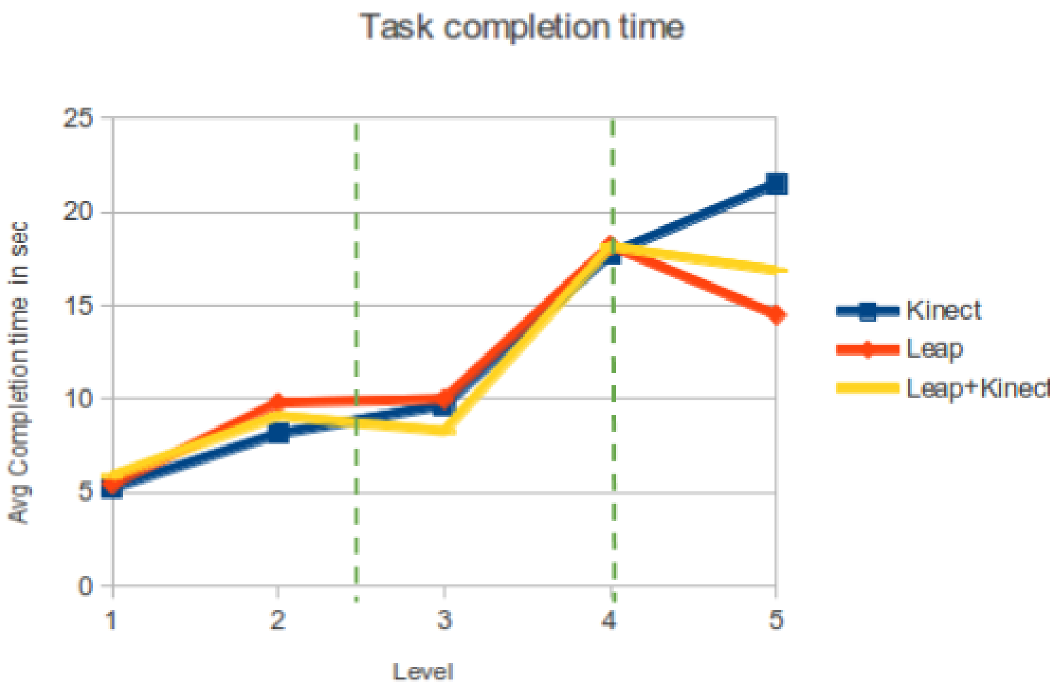


Figure 3.2: Average Task Completion Time for hamster game

User progression

The second aspect of quantitative data analysis was to focus on the average user progression through the inspection of the success and fail rates as well as the number of users who managed to reach a certain level. From the same individual tables presented in Table 3.1 that formed the basic feed for the Java analysis framework, the accumulative number of total trials per level as well as the summation of the successful and failed trials over all the participating users was aggregated. Looking at the mere numeric data of successful and failed trials wasn't much expressive or self explanatory. Hence we decided to work with success and fail rates, that is, the number of successful/failed trials with respect to the total number of trials at a certain level. As depicted in figure 3.3, three different graphical regions could be observed again. For levels one, two and three Kinect users recorded the highest success rates, whereas for level four Kinect users score significantly drops and LeapMotion users prevail, and for level five users of the combination managed to have the highest success rate.

Another interesting aspect we could draw from the raw data was the number

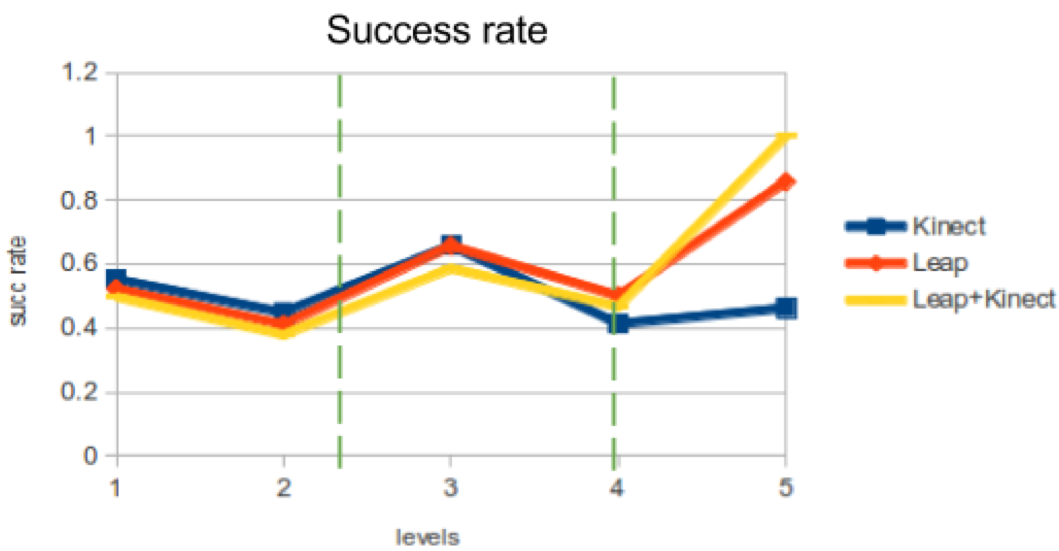


Figure 3.3: Success trials rate for hamster game

of users that managed to survive across the different difficulty levels. The graph shown in 3.4 shows clearly -and as expected- that generally the number of users who manage to reach the final levels drops in correspondence to the difficulty. But the interesting finding was to see how the medium used in interaction could affect this phenomena. The figure shows clearly how the data illustrated could be graphically divided into two adjacent regions. Levels one, two and three conform the first of the two regions, where the number of Kinect users who managed to reach those levels was clearly higher. However for levels four and five that constitute the second region, the number of users who used the combination apparently was higher.

After completion of the a game round -whenever it was seen suitable due to the age of the player- users were asked to fill out a standard System Usability Scale (SUS) questionnaire to evaluate the system usability, and they were also encouraged to add any valuable comments on the system in free hand form. An analysis of the SUS scores showed that users preferred LeapMotion (83.5) over Kinect (81) and the combined LeapMotion+Kinect approach (74.4).

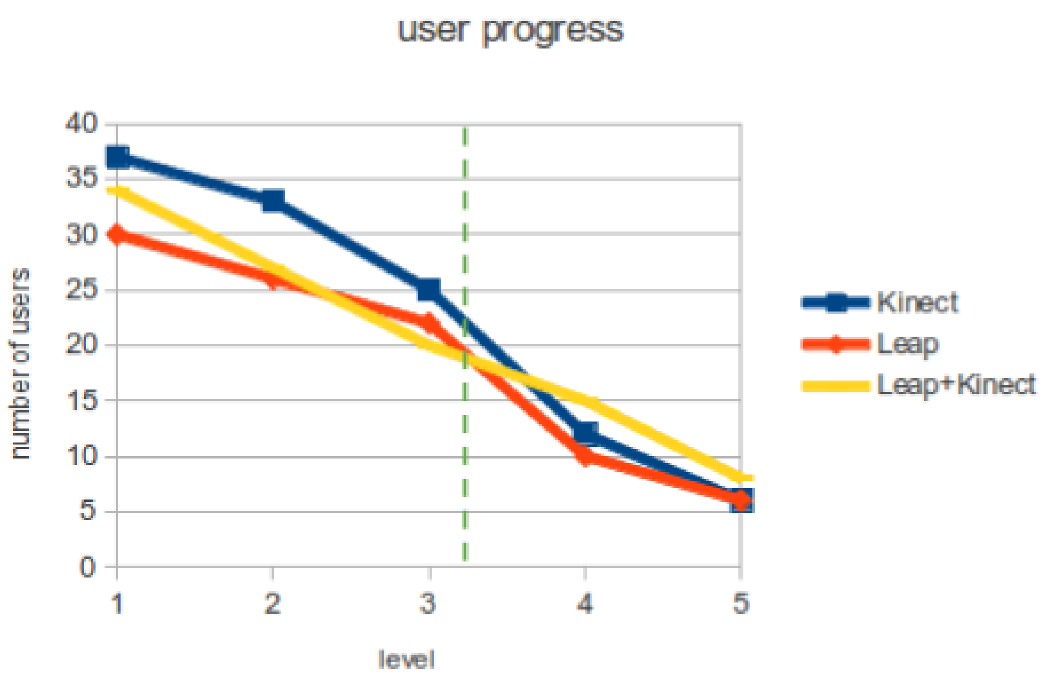


Figure 3.4: User progress across different levels of the hamster game

3.5.2 Qualitative Analysis

The quantitative data analysis section was mainly concerned with the inspection of application performance and effectiveness, whereas this section is primarily concerned with the different aspects of user experience and the levels of entertainment delivered to the player while using the system. Before getting into the details of this section it is important to explain how the qualitative data was gathered. A video recording camera placed on the opposite side of the public display in order to capture the entire scene including the player's position and movements with respect to the public display. Videos captured were then transcribed and matched against user's performance using the time stamp attribute that was logged while playing.

Since the Microsoft technology is already widely used and fairly known to a big portion of our user data set, the Kinect users video analysis was particularly interesting to observe at the more difficult levels, where the maze path was narrower and more curvy. It could be noticed that Kinect users faced some difficulties in precisely maneuvering the mouse pointer in critical edges, and at some cases it was even frustrating and caused the user boredom or loss of interest.
As for the LeapMotion, players seemed to be more entertained by using the new yet-undiscovered technology. However it was obvious to notice how the mental model they already gained through their previous experiences from other gesture-based game consoles (Kinect, Nientento Wii, Playstation Move) affected their behaviour using the LeapMotion. This was distinctively realized when players first tried to control the mouse cursor using the entire hand/hand palm (just like in Kinect case), instead of using just one finger.
The most interesting video analysis scenario appeared while transcribing the behaviour of the combined approach users. Users seemed to be intrigued by using the combination of different interaction zones and in some cases it could be clearly noticed how users used utilized the precision provided by the LeapMotion in order to overcome the hurdles in more difficult levels.

However one major problem certainly faced players from this user group, was the *"hamster jump"*. This happened when the player approached the screen and moved from the Kinect interaction spectrum to the LeapMotion interaction spectrum, where the mouse cursor suddenly jumps to a different position exactly after crossing the borders between both spectrums, which usually caused a wall hit and meant consequently increase in the error rate. The problem called for further inspection and hence a deeper look into the corespondent quantitative data using the time stamp attribute as a bridge between the video and data logs. As a result for the thorough inspection we managed to figure out that the reason behind the problem was due to the technological difference between the used devices. Kinect and Leap motion both take as an input a three-dimensional point and map it to the applications coordinate. However, both devices use two different coordinate systems whilst transforming the raw 3D input to application's 2D output, and thus the *"hamster jump"* appears. This problem called for an enhanced smoothing algorithm, that will be later discussed in the next section.

3.5.3 Enhanced Smoothing Algorithm

Having described the problem of sudden cursor relocation, or how the system user's always referred to it as *"the hamster jump"*, the need for improving the smoothing algorithm emerged. Noticing that this problem appeared at the crossing borders between the Kinect interaction spectrum and the LeapMotion interaction spectrum, it was clear that the solution had to be deployed exactly at this transitional and pivotal point. After deep inspection for the LeapMotion API, the *touch emulation* feature was discovered and formed the first founding stone for the inspiration behind the advanced smoothing algorithm.

Enhanced Algorithm Concept

Having described how the touch emulation feature works in Section 2.2.2 forms a ground, fundamental and principal background knowledge to understand how the advanced smoothing algorithm's logic is constructed. Carrying on the analogy from the touch emulation model, the physical space in front of the public display is further divided into three different interactive regions. The first interactive region is dedicated for the *"Kinect"*, followed by a very narrow *"Transition"* region and concluded by the *"Leap"* region. Once the user is present at the *"Kinect"* region, the 3D Skeleton hand's position is tracked and mapped just as previously mentioned to the 2D screen output, thus the transformation formula follows Equation 3.1 for this region.

$$\begin{aligned} X_{application} &= X_{Kinect} \\ Y_{application} &= Y_{Kinect} \end{aligned} \tag{3.1}$$

As the user approaches the public display, and crosses the border between Kinect and Leap motion, the user enters the *"Transition"* region. In this region, three main transformation constants are calculated and the X-Y coordinates undergo a special mapping formula. Firstly the mathematical difference between the (X, Y) pairs extracted from Kinect and LeapMotion is calculated to determine the first two transformational constants, namely $\delta_X$and $\delta_Y$determining the positional shift in the vertical and horizontal directions. The third transformation constant calculated at this region is the Z_{max} value, that is the value extracted by the LeapMotion once the user entered this region, i.e. the maximum Z-coordinate value in the line of sight by the LeapMotion. In this region, the mapping formula is as follows:

$$\begin{aligned} X_{application} &= X_{Leap} + \delta_X \\ Y_{application} &= Y_{Leap} + \delta_Y \end{aligned} \tag{3.2}$$

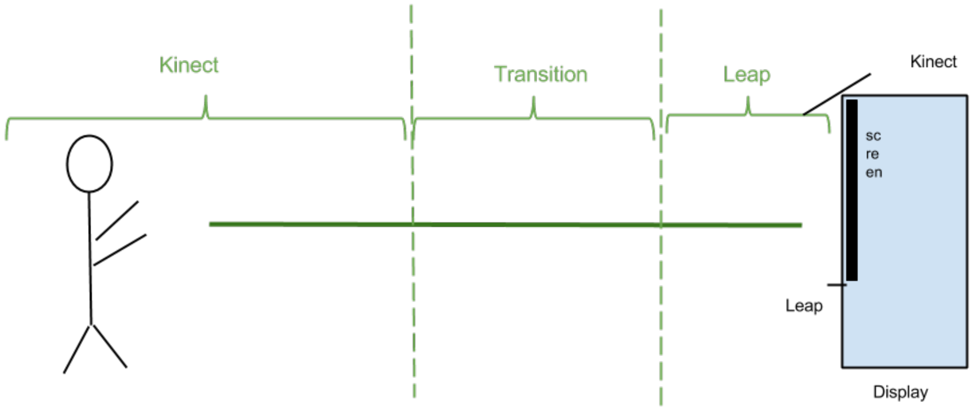
Starting from this point in the interaction the user is considered to be in the *"Leap"* region, and with the help of the previously calculated transformational constants together with two newly introduced transformation variable the real logic behind the algorithm starts to take over. The main idea behind the working logic was to utilize the fact that we could extract the Z-coordinate position, i.e. the user's relative proximity to the LeapMotion/public display and use it to gradually fade out the horizontal and vertical positional shifts caused by the transition from Kinect to LeapMotion's coordinate system. To further elaborate, once the user is past the *"Transition"* region, we try to constantly decrease the $\delta_X$and $\delta_Y$depending on distance between the user's finger and the LeapMotion/public display. The more the user approaches or nears the LeapMotion/public display the smaller should be the horizontal and vertical shifts, converging into a zero offset when the user is in a negligible distance away from the public display. In order to achieve this, we had to introduce two new transformational variables, that is $\delta_{X(t)}$and $\delta_{Y(t)}$. Both variables are initially set to equal $\delta_X$and δ_Y On each frame executed by the LeapMotion , these variables are updated to comprehend the new Z position according to Equation 3.3.

$$\begin{aligned} X_{application} &= X_{Leap} + \delta_{X(t)} \\ Y_{application} &= Y_{Leap} + \delta_{Y(t)} \end{aligned} \tag{3.3}$$

Once the deltas are updated, the 2D application's coordinates are accordingly calculated based on Equation 3.4. Figure 3.5 gives a graphical elaboration of the idea.

$$\begin{aligned} X_{application} &= X_{Leap} + \delta_{X(t)} \\ Y_{application} &= Y_{Leap} + \delta_{Y(t)} \end{aligned} \tag{3.4}$$

[1]Distance ratios of respective zones in the figure do not represent the actual distance ratios. For clear representation purposes and needed equation writing spaces all zones are represented to have almost equal spaces



Equation	$X_{application} = X_{kinect}$ $Y_{application} = Y_{kinect}$	$X_{application} = X_{leap} + \delta x$ $Y_{application} = Y_{leap} + \delta Y$	$X_{application} = X_{leap} + \delta X(t)$ $Y_{application} = Y_{leap} + \delta Y(t)$ where $\delta X(t) = \delta X(t-1) * \lvert Z(t)/Z_{max} \rvert$ $\delta Y(t) = \delta Y(t-1) * \lvert Z(t)/Z_{max} \rvert$

Figure 3.5: Side-view Graphical illustration for the enhanced smoothing algorithm [1]

3.5.4 Design Guidelines

After careful review to the previously stated outcomes in the quantitative and qualitative analysis sections, some important conclusions could be drawn. In the graphical illustrations for each of the observed performance evaluation criteria (Task completion time, Error rate, User progression), the graph could be spatially divided in 3 main -and almost constant- regions. In most of the cases the first spacial region consisted of levels one and two, whereas levels 3 and 4 formed the second region and level 5 the third region. While the qualitative data gave a clearer overview on how users preferred to use Kinect due to their former gesture-based control experience gained, but however preferred LeapMotion when the levels became harder and needed higher rigor. Having stated that, it is obvious that for different task difficulty or with a desired task performance precision a certain interaction modality is not only preferred but also performs better. And hence the denouement of mapping different services or level of information of a public display depending on the user's relative proximity arose. The results clearly pointed out a new perspective in our problem solving technique that we followed so far. Instead of fighting against the different interaction spectrums and instead of battling against the technological differences between gesture sensing devices, we could actually utilize the heterogeneity between different interaction spectrums to map them to different input and corresponding output modality-pairs. Thus we concluded that the spacial physical space could be conceptually divided into four main interaction spaces differentiating between different user proximities. Each space offers a different granularity of interaction fitness maintaining the idea to move seamlessly between the spaces to transition from rather coarse to high-precision gestures. Hence the 4 zones are not seen competitive, but rather complementary.

Touchscreen Area

This area is dedicated for direct-onscreen interaction. The on-screen content could be manipulated using common touch gestures including tap, drag, swipe pinch and rotation. Other interaction-related dimensions such as multiple hands, multiple fingers and the amount of pressure applied to the touchscreen are also comprised in this zone. The limitation of the *fat-finger* [5] [39], where the user's finger/hand covers a portion of the information displayed on the screen occurs while the touch event. When the user's finger/hand is lifted an automatic transition into the more general interaction zone takes place, and hence the occluded interaction space is seamlessly revealed and the interaction continues seamlessly. Feedback options for this interaction zone include visual, auditory and haptic responses.

Fine-grained Gesture Area

This zone is dedicated for interactions that occur in immediate proximity to - but not touching- the screen. It basically begins where at the moment where the touch is lifted and covers the general area where the mouse and keyboard are traditionally placed. To be more precise the interaction ranges from 0 m up till 0, 5 m and the deployed LeapMotion sensor is responsible for interpreting gestural operations occurring in this zone. Due to the technology used the available gesture set includes perpendicular tap, click, swipe, pinch, rotation in addition to custom recognized gestures like grabbing. All the previously mentioned gestures could be tracked when performed using multiple hands and up to 15 fingers simultaneously present. For the sake of the extreme close proximity of this zone a visual feedback given in a high resolution allowing a hovering effect before the tap gesture could be introduced. This could be of a great impact in some application scenarios including the display of tool-tips, preview effect, or pre-select objects on screen. In addition to the visual feedback, an auditory equivalent could be also mapped to this interaction zone, since the close proximity grants the user the ability to notice audible effects.

General Gesture Area

This area covers the space between the directly in front of the screen and the medium-far back of the entire space. The area stretches from 0, 5m up till 2 m away from the public display. Interaction starts effectively starts when the certain points in the user's skeleton are successfully tracked and thereupon user's movements are mapped to custom gesture set. A total of 20 skeleton joints could be tracked, that could be roughly divided as 10 for the upper body (including both hands) and 10 for the lower body(including both feet). The deployed Kinect sensor is responsible for tracking the interest points and interpreting the tracked gestures. Due to the medium proximity dedicated to this zone, it should be effectively used for content consumption where the user is comfortably positioned in front of the public display (for example for video watching or reading a longer document) rather than performing a fine-grained task. For an adequate user feedback it needs to take this further proximity into account and display information in a more adjusted way. The feedback could be presented in a medium-sized text and icons, whereas audio feedback should also be deployed.

Macro Gesture Area

This interaction zone is dedicated for the background activities that may occur in a public display setup. It covers the physical space in front of the display and beyond the directly presumed interaction area. This includes the back of the room where multiple people may be be present and hovering over some content shown on the screen. The sensing range stretches from 2m to 3m away from the display. Very similar to the previous zone, a full body skeleton could be tracked using the deployed Kinect sensor, where a lower priority is assigned to user's present in this interaction area. Visual feedback should be adjusted and rescaled to an enlarged form in terms of icons and text size, allowing the users to effectively interact from a relatively far distance. The visual output could serve as a main attraction technique to draw attention of regular passer-by users and invite them to engage with the application using further interaction zones. On the other hand this zone could serve as energy saving guard, that detects the absence of potential interested users, and hence turns off the display temporarily to decrease power consumption. Auditory feedback could also be seen suitable for this interaction area, where as a critical limiting factor will be the relatively high audio levels that are needed to reach the interacting user.

After defining our framework for the different interaction zones differentiating between variable user proximities, a second user study was seen to be essential in order to validate the effectiveness of such spatial division and to give a deeper insight, how such a division could affect user's gestural behavior in different application scenarios. In the chapter to follow, a more thorough inspection for the resulting framework will be presented.

Chapter 4

Zone Evaluation study

The main spotlight of this chapter is the second prototype application implemented during the course of this thesis. The chapter describes the apparatus used, as well as the methodology followed to log and collect the data. Characterization for user demography is also included within this chapter. Finally the results and design guidelines that represent the outcomes of the study are given.

4.1 Concept

The central focus of this prototype is to further explore the feasibility of the four interaction zones, that represent the outcome of the previous study. The study aimed at explicitly emphasizing the division between the different zones while investigating how the user's gestural behavior evolves across the respective zones. The study also covers the second main archetype, specifically the mere gestural mode. In contrast to the first archetype, the mere gestural mode uses the motion sensing technologies to interpret the user's movements and accordingly map it to the desired functionality. This means, there is no real-time feedback on the screen that imitates user's movements like the moving cursor. Consequently some of the problems that appeared in the previous study, like the *"hamster jump"*, would have a negligible or unnoticeable effect in such a model. The real challenge that faces this archetype is to find the different gesture sets that suits each interaction zone best depending on the application scenario, and how efficient can the user perform a certain task within a certain zone or across two different zones. In order to address these desired goals, the study was structured as two consecutive phases. The first phase concentrated on the quantitative side of the study, where the performance within and across different zones was evaluated through a pointing task. The second phase contemplated on the qualitative aspect of the interactive zones through conducting semi-structured interviews while video recording user's gestural behavior for offline analysis.

4.2 Appratus

In order to study the previous questions we needed to have an application scenario that not only distinguishes between the four interaction zones, but rather forces the user to move from one zone to the other depending on the required task. We needed to have this sense of movement obligation in order to allow the user to wander across all the zones, so that the user realizes ,at least roughly, the physical boundaries for each of the interaction zones. Moreover it will allow us to determine later in the second study phase which gesture set suits best in each respective zone. In this manner the basic idea behind our prototype was developed. To be able to best describe how the prototype works, the hardware setup is described firstly. We used different sensing technologies to map to the different interactive zones. To represent Zone 1 (The touchscreen area) a 46 inch ekiosk [1] was used. Perpendicular to the ekiosk a wooden board was fasten to hold other sensor devices. LeapMotion was deployed to represent the second interactive zone(Fine-grained gesture area), whereas a Microsoft Kinect was placed to be responsible for zones 3 and 4 (General gesture area and Macro gesture area)respectively. In addition to the motion sensing devices, a Hero3 Go Pro wide-angle camera was placed on the other side of the laboratory to capture the entire scene including the user's movements and gestural acts with respect to the interactive zone and the public display. An illustrative hardware setup is shown in Figure 4.1 for further clarification.

[1]http://www.ekiosk.com/

Similar to the first prototype, the first phase of this user study included a cursor moving task with an additional, important modification. Instead of having to guide a hamster through a maze, the application presented the user with a circle-shaped target on the screen at a random position. Each target drawn on the screen was colored with one of the following colors: **red, blue, black, or green**. Each of the previous colors represented a different interaction zone according to this mapping scheme:

red: Touchscreen area.

blue: Fine-grained gesture area.

black: General gesture area.

green: Macro gesture area.

The user's task was mainly to move the cursor and hit the target, but the hit only counts when it occurs in the corresponding interaction area indicated by the target's color. The hardware sensors worked just like in the previous prototype, where the LeapMotion was responsible to track user's fingers and Kinect was used to track the user's hands. In addition, the ekiosk allowed the user to move the cursor directly by touching the desired location on the screen. In order to signal the interactive zone, in which the user is currently present, the cursor's color alternated in accordance to the previously mentioned color mapping scheme.Polluting the screen with a legend representing the coloring scheme was not desired by our research group, and as an alternative, the experiment's facilitator communicated the corresponding zone to the participant when a target appeared. The target acquisition task was inspired by the work of MacKenzie et al. in extending the Fitt's law into two dimensional space [25]. Our study presented a trial into extending the concept to 3 dimensional space.

To measure the input within zones and across zones, we designed the creation of targets in an alternating way: First, a target is created in a random zone according to the Balanced Latin Square algorithm at a random location on the screen.

Then, a target is created in the same zone at a random location to make a pointing task within a zone. This procedure was repeated 50 times per participant so that there were 50 interactions within zones and 50 interactions across zones. The algorithm also varied the diameter of the target between 10 and 100 pixels to alter the difficulty of the task. The algorithm balanced the size of the target and the distance from the starting point to provide the same difficulty as in the other transitions.

From the software architectural point of view, the prototype strictly followed the Observer design pattern [23]. Simply explained the system monitored the application's state, and upon a state change event, the appropriate event handler(observer) is called and the event handling process is delegated to it. In our application the system's state consisted of three main attributes: the user's interactive zone, the target's interactive zone, and the cursor's coordinate. For each of the motion sensing devices used, an event handling listener was implemented, and depending on which interactive zone the user is currently present, the corresponding event handler(observer) is called. The main task for all the event handlers was to decide the validity of the cursor's hit, and based on the result, how the application's logic moves on.

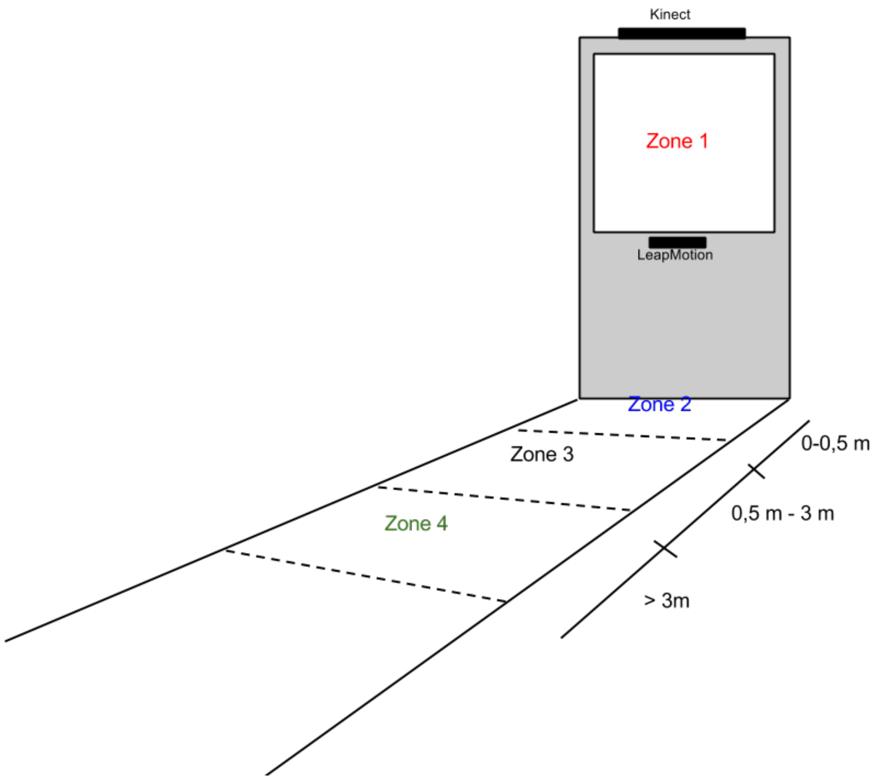


Figure 4.1: Illustrative visualization for zone evaluation hardware setup

4.2.1 Data Logging

To collect all the different data needed for the analysis, multiple data logging techniques were essential to use. The prototype implemented provided the required framework to gather basic quantitative data. For each individual user, the application generated a text file containing the interaction attributes regarding the 100 within and across-zones pointing tasks presented to the user. For each task the following attributes were logged:

start X: The x coordinate where the mouse cursor is currently present.

start Y: The y coordinate where the mouse cursor is currently present.

target X: The x coordinate where the target is present.

Distance to target: The shortest path between starting point and target.

Path length: The actual path taken by the user from starting point to target.

Target size: The size of the target in pixels.

Time: Time needed to reach from start to end .

start mode: Indicates the interaction zone where the user started the interaction.

End mode: Indicates the interaction zone where the the target is present, i.e. where the user needs to be to hit the target.

In addition to the data logged by the application, the video log is also collected by the experiement's facilitator and saved together with the corresponding data log.

4.3 Participants

For this study we recruited 14 participants (11 male, 3 female). The participants were between 23 and 34 years old (M=25., SD=3.33). Most of the participants were communication engineering or computer science students. In order to eliminate the risk of color confusion, participants were asked whether they suffered from any kind of color blindness and replies reported that no one suffered from it. To make sure that users would get acquainted to the setup of the experiment easily, they were asked about the type of their workspace.10 participants stated that their usual workplace is a sitting workplace in front of a monitor, and one participant sometimes used a standing-workplace. Users were also asked them about their previous experiences with gesture-based input technology. One participant was playing with a Nintendo Wii on a daily basis, and another participant reported to use a LeapMotion approximately once a week. To ensure the diversity factor of our implemented prototype, users were asked to give the preferred hand to use during the course of interaction. Two of the participants were left-handed, and hence our prototype was adjusted in accordance.

4.4 Procedure

The study was conducted at the user experience laboratory at the campus of University of Stuttgart. For this particular experiment, it was seen more appropriate to take place in a controlled environment rather than in the wild. The main reason for that was the nature of the application itself. Unlike the first prototype, this one was not masked in a game application, it was rather a straightforward task performing application. And hence, it would have been hard to attract a regular passer-by. Users were recruited from different faculties around the campus, and they were invited to participate in the 40 min study. As a compensation for user's contribution they were rewarded to choose from a diverse set of confection.
In contrast to the first study, we didn't want to compare different solutions against each other upon specific performance measures. In this study we had one proposed solution architecture, that we wanted to be tested by each and every participant, thus study was structured as a repeated measure or within-subject design with the interaction zone as the independent variable. One major disadvantage to the within-subject design is the learning effect. If participants are tested under condition A first, then under condition B, they could potentially exhibit better performance under condition B simply due to prior practice under condition A. To compensate for this, counterbalancing using a 4x4 balanced square algorithm technique is used. Figure 4.2 shows how the counterbalancing works. Each row represents a different combination in which targets appear on the screen. Note that each condition appears precisely once in each row and column. Furthermore, in each row, each condition appears before and after each other condition an equal number of times. For example, condition B follows condition A two times and it also precedes condition A two times. Thus, the imbalance is eliminated.

A	B	D	C
B	C	A	D
C	D	B	A
D	A	C	B

Figure 4.2: 4x4 Latin Square Algorithm [17]

In the first phase of this experiment, participants were asked to fill out a short questionnaire about demographics, their workplace, and their experiences in gesture-based input technology. After that, they were given some time to make themselves familiar with the setup and to perform some pointing tasks to get used to the prototype. When the participants felt comfortable with the prototype, they were asked to perform a pointing task. As objective measures, we collected the time a participant needed to reach the target, the distance from the start towards the target, and the target size.
In the successive phase, more light was shed on the possible gesture set that users saw most suitable for each interactive zone. Inspired by the work of Wobbrock et al. [43]. three main application scenarios were put under investigation. For each of the application scenarios, a set of actions were suggested by the experiment's facilitator and asked the users to invent and perform a 3D gesture for each action in each interactive zone. The entire gesture exploration done by the participants was video captured by the Hero3 wide-angle camera, and saved for later offline analysis. The application scenarios that were introduced to the users are:

Navigation in documents:

Actions: zoom, scroll, forwards, backwards, selection

Controlling applications:

Actions: selection, copy, paste, cut, drag, drop, exit

Navigation in maps:

Actions: rotate, zoom out, zoom in

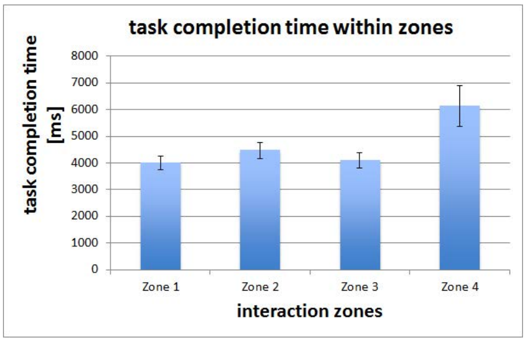
Participants were also encouraged to suggest any other application scenarios where they would imagine such a set up could be beneficial or might enhance quality of tasks performed. Users' replies were also video recorded for further inspection. After participants had defined a gesture for each action in each zone, we asked them to fill out a final questionnaire, where they had to rate the interaction in each zone for three categories on a Likert scale from 1 to 5, where 1 was considered as not suitable and 5 as very suitable.

4.5 Results and Implications

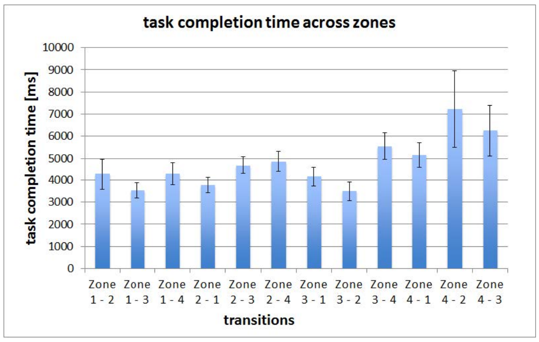
Just like the previous study, the results section for this experiment is also further structured into two parts, a quantitative results section and a qualitative one. Quantitative results focused on measuring how well the participants performed in the pointing task in the first phase of this experiment, whereas the qualitative data concentrated on the likeness factor and the usability of each interaction zone evaluated by the users.

4.5.1 Quantitative Results

The quantitative analysis for the data gathered focused on the task completion time to give a comparative view between the with-zone and across-zones performance. Based on the data logged and through the start mode and end mode attributes, the framework used for the analysis could extract and differentiate between the within zone and across zones interactions. For each user the framework produced two separate outputs, one for the within-zone and the other for the across-zones interactions. The output presented the average task completion time per interaction zone. The output was then aggregated over all the participants to end up with a general average task completion time. The general average task completion time for all the within-zone interactions is plotted in figure4.3. The



(a) Graphical representation for within-zone task completion time



(b) Graphical representation for across-zones task completion time

Figure 4.3: Results comparison between across and within zone interactions

task completion time for the within zones interaction ranged from 4007.ms and 6133.8ms, where the best results were scroed in Zone 1, Zone 2and Zone 3, i.e. Touchscreen, LeapMotion and Kinect interactive zones averaging a task completion time of 4200ms, whereas the worst was reported in zone 4. Results for the across zone interaction are illustrated in Figure 4.3. The task completion time for the within-zones interaction ranged from 3533, 5ms to 7223, 7ms. Important to notice that best results where recorded in transitions from zone 3 to zone 2 (Kinect to LeapMotion), from zone 2 to zone 1 (LeapMotion to Touchscreen) and from zone 1 to zone 3 (Touchscreen to Kinect) averaging in a task completion time of 3600 ms across the respective zones, while the worst results were present in transitions from zone 4 to zone 2 and from zone 4 to zone 3.

4.5.2 Qualitative Results

The qualitative results focus could be further structured in two main criteria, namely the usability of the prototype system and the innovation of 3D gestures done by the participants for each interactive zone. The questionnaire handed to the users by the end of the study served as the main data collection method for the system's usability. For each of the suggested application scenarios, the user was asked to give a rating for each interactive zone, indicating how suitable the zone is with respect to the application scenario The results of the questionnaire were accumulated over the total number of the participants in order to reach an average usability rating for the separate zones. Table 4.1 presents the collective outcomes for the average usability rating. Worth to note how the average ratings for Zone 2 prevailed in all the suggested application scenarios. In other words, it is obvious that participants preferred to use LeapMotion even if the application didn't require extreme precision or high efficiency.

Representation	Nav. in Document	Controlling applications	Map Navigation
Zone 1	1.571	1.500	1.714
Zone 2	3.357	3.142	2.500
Zone 3	2.285	1.785	1.28
Zone 4	3.142	2.785	2.500

Table 4.1: Average usability rating for Zone evaluation study

The second component of the qualitative analysis focused on the invention of 3D gestures based upon the experience users just gained through the pointing task. For each of the 14 participants the video capture was analysed offline and a transcription for each gestural act performed carefully studied, and then the observation was aggregated over all users detecting the most common suggested gesture per action. Central elements that were inspected during the gestural examination were:

1. Number of hands used
2. Number of fingers used per hand
3. Direction and duration of the motion
4. If gesture consists of multiple integrated gesture

Gestural transcription was done for each of the application scenarios separately and the collective summary for the invented gestures are presented in Figure 4.4.

Controlling applications:

For the selection, users preferred to use pointing a finger for Zones 1 and 2, whereas pointing with the entire hand palm for Zones 3 and 4. For copying and pasting (same operation in reverse order), users suggested to point and hold with one finger for Zone 1, while they chose to "grab" and "release" an object through an open/closed fist gesture for the rest of the zones. One of the most interesting and most intuitive gestures participants came up with, was the cutting gesture. Participants defined a scissor-gesture in Zones 2, 3 and 4. Another interesting observation is that participants defined an exit gesture by moving a cursor with their hand or finger to a desired upper corner of the screen. This finding strongly shows how users are influenced by already established mental models. Since the typical place for exit in WIMP interfaces in placed in an upper corner as well.

Document Navigation:

As shown in the figure 4.4, users chose one hand and two finger gestures for zooming in Zone 1, while they suggested an open/closed palm gesture mapped to zoom in/zoom out for Zone 2. For Zones 3 and 4, users preferred to use 2 hand gestures where the direction of hand movements whether towards each other or away from to each other should confine to zoom in or zoom out operation. Scrolling, forward navigation and backwards navigation were all treated the same by the users where the direction of the movement(up, down, left right) should be mapped to the operation. The only difference is that participants liked to use one finger for Zones 1 and 2, where as the whole hand palm for Zones 3 and 4. Regarding text selection, results indicated that for Zone 1 users promoted the usage of one finger to mark the beginning of the selection area, and then swiping till the end of the selection area. This could be directly mapped to the model users adopted from the usage of smartphones. For Zone 2 preferred using two hands with a pointing finger for each hand to mark the start and end of each selected text. Zones 3 and 4 are treated similarly, where users saw that pointing an open hand palm and swiping towards the end of the selected area was best suitable.

Navigation in maps:

In regard to the rotation task, participants favored for zone 1 to use a one hand gesture with two fingers to indicate the rotation's fixed point with one finger and the direction of rotation with the second finger. For the rest of the zones, users suggested to use a semi-closed palm while rotating the entire hand to decide the rotation direction. For zooming, the gestures were increasing or reducing the distance between two fingers indicating a zoom in and zoom out operation for Zone 1, while for Zone 2 people chose the suggested gesture was open/closed palm. Gestures favored for Zones 3 and 4 were the same as Zone 1, but with one modification, that is to use the whole hand instead of the fingers.

4.5.3 Implications

After a thorough inspection for both the quantitative and qualitative results, an access towards the true implication of the presented was granted. Implications drawn from the study represent a general set of observations, that could conform to be a general design guidelines for multi-modal public display application's with various interactive zones. Our realization for the interpreted results could be summarized in the following points:

Zone tranistions:

Careful examination for the qualitative results would yield that Zone 4 preformed worst in the within-zone interactions, and whenever it was involved in a zone transition operation it spoiled the results. Hence Zone 4 will be excluded from the discussion in this particular section, and will be separately discussed in the triggering action section. Avoiding zone transitions coming from or going to Zone 4, one could clearly recognize some transition paths that perform better than others. The transition path from Zone 3 to Zone 2 and then Zone 1 (3-2, 2-1 in figure 4.3) scored the best average task completion time. The average task completion time scored along this path was even close to the average within-zone task completion time for Zones 1, 2 and

3. This interesting phenomena could be observed when the reversed path is tracked. That is the path from Zone 1 to 2 and from Zone 2 to 3 (1-2, 2-3 in Figure 4.3). The reversed path yields also good results, but it is obvious that the task completion time increases along the reversed path.

What this could imply for the design guidelines, is that people tend to perform better when going from a general, more coarse interactions to a more specific and fine-grained ones. It could be also argued that the result implies avoiding interaction schemes that require backward steps. Such an aspect should definitely draw the attention of user interface designers when dealing with multi-modal interaction spectrums. The application's flow of control should be maintained in such an order that the first interactions are available in further interaction proximity, whereas the more detailed, longer interactions, should be allowed from a closer proximity. In all cases, avoiding transitions with gaps, such as from Zone 1 to 3, 1 to 4 ...etc should also be maintained to hold the applications' conformity. Having discussed how the quantitative data implied the most suitable transition across respective zones, the question now rises which gestures are best suitable for each zone. This was implied by the second phase of the study and will be further discussed in the following section.

Close proximity:

As the usability evaluation results showed, participants favored interacting in the second Zone regardless of the type of the application. Even in application scenarios where fine-grained gestures where not crucially needed, users yet preferred Zone 2. Based on such a result, it could be argued that most of application's functionalities should be mapped to gestures in close proximities. Due to the close proximity exhibited by this interaction zone, services implemented require a higher input precision as well as more accurate feedback when it comes to legibility of potentially small content changes. For example, when selecting text with high precision in a distant zone, a user will have difficulties viewing the selected text when displayed in small font.

When defining gestures for close zones, fine-grained hand and finger gestures should be used. Our study showed that participants defined mostly finger gestures when being close to the screen. Particularly when performing high precision tasks, finger gestures provide a higher granularity. Fine-grained actions such as selecting text in a document or cutting and pasting objects should be mapped to zones that are close to the screen.

Distant proximity:

We suggest for defining gestures to use the entire hand or other full body parts in distant zones. The results of our study show that the user-defined gestures of Zone 3 and 4 are always the same. All defined gestures in those zones are using the entire hand or both. According to Vladimir et al. [31] gestures can appear in an unintentional form that does not convey any meaningful information. Thus and in order to prevent input errors it is suggested to use hand gestures in distant zones in favor of finger gestures.

Triggering actions:

As mentioned before Zone 4 had the worst scores when it came to performance evaluation. Hence, it shouldn't be used for task execution purposes. According to the ten basic input modalities presented by Mueller et.al. [28], the presence of a person can be considered as application trigger. Therefore, just triggering an action, which does not require more complex interaction can be performed well in distant zones. Those actions could be exiting an application, or turning on/off the public display depending on the user existence for energy saving intentions.

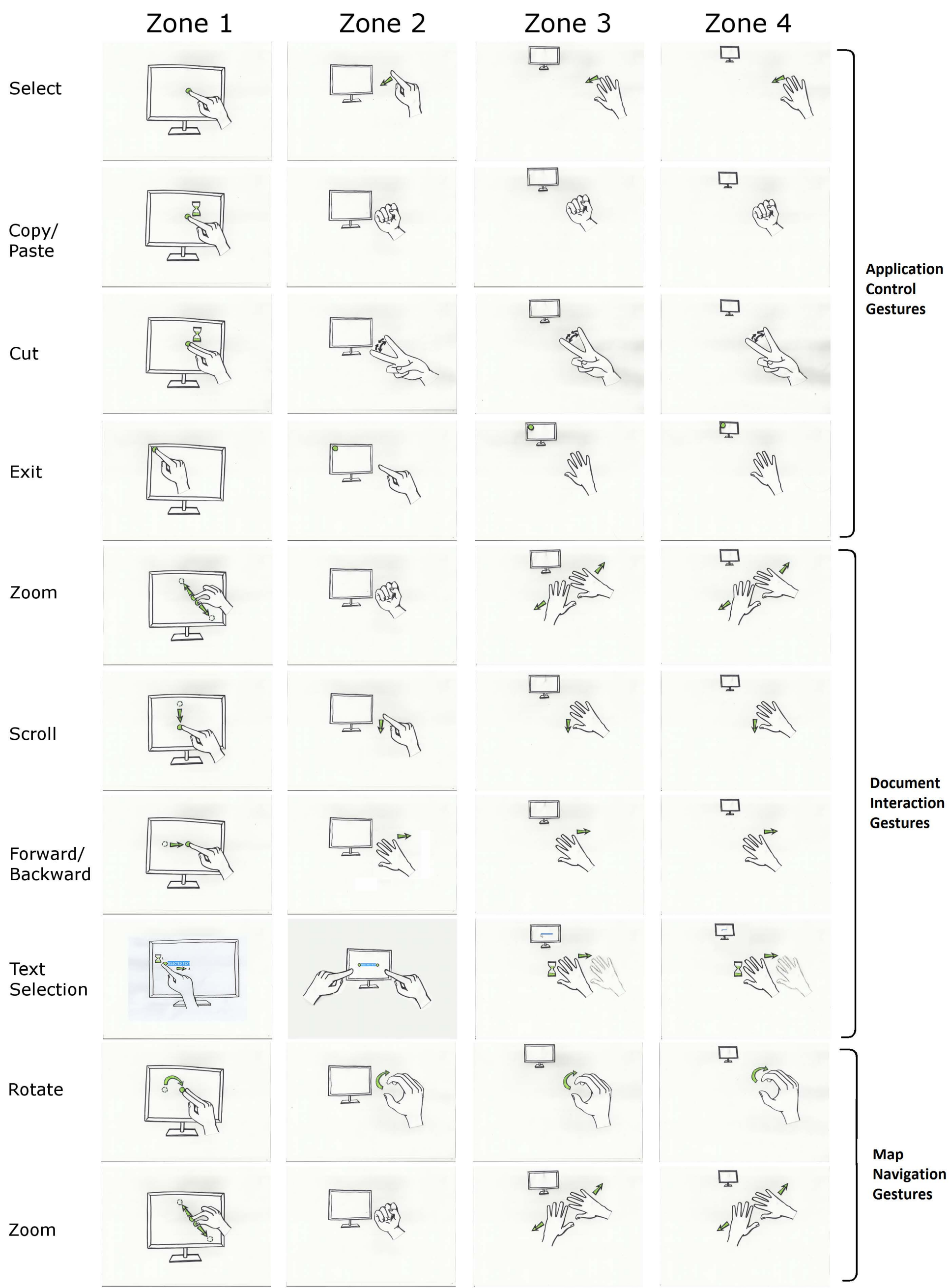


Figure 4.4: Cumulative summary for user-defined gesture sets in different application scenarios
Designed by: Carla Harris- Helmholtz Institute, Munich

Chapter 5

Discussion & Conclusion

The final chapter of this thesis summarizes the presented work. Once again, the chapter highlights the motivation behind the study and emphasizing on how our approach and implementations are structured towards finding an answer to the research question. Furthermore, the chapter discusses the limitations that were encountered during the research and sheds the light on potential future research areas where the study outputs are expected to have positive impact.

5.1 Conclusion

For the last decade Hollywood movies have always introduced futuristic forms of touch-less user interfaces, where the user would just wave and use mid-air gestures to control a computer system. Nowadays, especially with surfacing new hardware sensing technologies like the LeapMotion, the question arose whether such interfaces were mere fantasies or science made it so far for it to exist in reality? Having such an intriguing problem statement tempted our research group to further investigate the matter in a trial to bring the once seen-ahead-of-time technology to today's realm. Our research aimed at finding an answer for computer systems in general and public display systems in particular.

Before commencing with our approach to finding an answer for the above stated question, a brief inspection for the current state of the art gestural recognition models, as well as previous work done in this area, was essentially needed in order to form a solid cornerstone for our research. The next step in the research path was to extensively explore the market of hardware motion sensing technologies to be able to construct the appropriate hardware setup. Consequently, the combination of Microsoft Kinect and LeapMotion was seen to deliver an adequate solution for the presented problem. In our realization of the research problem, we inspected two main application archetypes, namely continuous real-time feedback applications and purely gesture-based applications. In order to gain a deeper understanding for the requirements and problems that faced each archetype, our research was structured to investigate each separately.

In our examination for the first paradigm, we constructed a maze-type game where the user controlled a hamster through a labyrinth across different levels. The game was deployed in public events to be put under test and collect our initial user feedback. Results of the study showed that user's performance as well as behavior, varied depending on the proximity of interaction and the level of difficulty for the task in hand. We could conclude that the physical space in front of the public display could be divided into different interactive zones, where each zone is better suitable for a different type of interaction or better suitable for delivering different level of precision. Results for this study were twofold: The first part of the outcome represented an enhanced algorithm was developed to maintain a smooth transition across the different interactive zones without disturbing the user experience. The second part of the results was the introduction of a framework for structuring the physical space in front of the display to four main interactive areas: a touchscreen area, fine-grained gesture Area, a general gesture area and a macro gesture area. As these zones needed further exploration, the outcome for this study presented the input for the second study.

The second study was concerned with quantitatively as well as qualitatively inspecting each of the four interactive zones both in within-zone and across-zone

manners. The study was conducted under a fully controlled lab environment to gather the required data. It was further structured into two successive phases. In the first phase a pointing task application scenario was developed in order to assess the performance and to give a comparative view on the user's efficiency within and across zones. The second phase of the study focused mainly on user's gestural behavior in each zone separately in an attempt to appraise a gestural design guideline. The outcomes of the study showed that some of the interactive zones are not best-suited for task performing usage due to the far proximity and the existing imprecision. For the other interactive zones, results proved that some transition paths perform better than others and implied that transition gaps should never be encountered in application development in order to maintain the delivered user experience. A set of the most commonly suggested user-defined gesture set for each interactive zone was also included in the outputs of this study.
As this thesis was developed with cooperation of Fraunhofer IAO research facility, the chance to display and demonstrate our implemented prototypes was made possible in multiple public exhibitions and events. The next sections will shed the light on some of the most important appearances of the system together with the feedback received. Limitations that faced us during the course of this study are presented and discussed as well in the next sections. Our recommendations and implications for further research directions based on the results presented are also demonstrated in the future work section.

5.2 Public exposure

Having an industrial partner in the supervision of this master thesis has gained us the benefits of public exposure to gain extra validation and credibility regarding user feedback.

Tag der Wissenshaft :

The event took place on the 22nd of June 2013, was organized by the state of Stuttgart, and held at the main university campus in Vaihingen, Stuttgart [1]. The event aimed at giving its visitors a first insight on the latest technology advancements and the current state of research in progress in order to raise the technological awareness and knowledge of the audience. Around 120 research institutes demonstrated their recent research results and provided help whenever it was sought by visitors. Visitors consisted mainly of families with their children who lived nearby the Stuttgart area. The event was used as a wild environment to test our first implemented prototype and the feedback composed our main data for analysis.

Sommerfest at Fraunhofer:

The event was a privately held event for the employees of the Fraunhofer IAO together with their families. It was held at the Fraunhofer main campus in Vaihingen, Stuttgart on the 20th of June 2013. The event fell under the category of entertainment, where the visitors got introduced to the work done at Fraunhofer through enjoyable and engaging context. Our first implemented prototype perfectly fit this criteria, since the hamster game attracted a large number of participants. Results gathered during this event were also used for the data analysis.

IT & Business Messe:

The exhibition took place from 24. 09.2013 - 26. 09.2013 was organized by LandesMesse Stuttgart GmbH, and was held at the main exhibition hall in Echterdingen, Stuttgart Messe [2]. The event aims at bringing IT companies and business seekers together in one place to establish contacts. Four main areas are the main focus of this fair:Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), Enterprise Content Manage-

[1]www.uni-stuttgart.de/tag/2013

[2]http://www.messe-stuttgart.de/where-it-works/

ment (ECM) and Output Management. Fraunhofer was presented in this exhibition as an exhibitor, were the interactive wall project was used to demonstrate information about services Fraunhofer provided for its business counterparts. Positive feedback was received by the visitors and the requests to rent the public display were submitted.

"Gestensteurung im Alltag" conference:

The conference took place on 4.12.2013, and was organized by the Bundesministerium für Bildung und Forschung. It was held at Geriatrisches Zentrum, Esslingen [3]. The conference aimed at exchanging technical knowledge regarding gestural interaction in every day life applications. Participants from research institutes and companies like Microsoft GmbH and DLR were present. Fraunhofer was invited to participate in the conference to demonstrate the interactive wall project, whereas the hamster game was presented during the break for entertainment purposes. Positive feedback was received regarding the integration of LeapMotion and Kinect.

Türöffner-Tag der Sendung mit der Maus:

The event took place on the 3rd of October 2013. It was organized by Fraunhofer and it was held at Fraunhofer main campus [4]. The event is part of a series of events that followed a popular TV program for kids in Germany. It aimed at giving the children the latest technology improvements in an amusing and delightful context. Our hamster game was used to introduce the idea of gestural interaction to the visitors as a demo application. It received great success during the day and press coverage in the local newspaper, *Stuttgarter Zeitung* , was also present.

[3] http://www.geni-aal.de/
[4] http://www.wdrmaus.de/

5.3 Limitations

During the course of the research, a lot of limitations have been encountered. Reasons for these limitations varied between resources, time, technology, or other external limitations. The most important ones are emphasized in this section.

Application Scenarios:

Despite all the benefits that our study group gained through the valuable cooperation with Fraunhofer research facility, some constraints were introduced due to the nature of application scenarios that were desired and deployed on the interactive public display project. The greatest limitation was the fact that all the applications were restrained to confine to the presentation form to be applicable for the various exhibitions and fairs where the public display was hosted. The setup presented throughout the study was seen to be beneficial in some application scenarios that were not directly applicable to the hosting events. An application scenario for a public library with different functionalities mapped to different proximities, (for example flipping pages, zooming into text passages, and highlighting text), was desired by our research group, but however incompetent to the deployment environment.

Mental Models:

Throughout the study it was generally noticed that users are biased by established metaphors and mental models. The interesting phenomena probably emerged due to the recent years of interaction with touch technology that invaded the smartphones and tablets markets. The mental models established through such an exposure influenced users' mental models for the mid-air technology. This was specially noticeable in the second phase in our second study where users were asked to invent their own 3D mid-air gestures. It would be interesting to repeat the study with users with no previous experience with multi-touch technology.

Positioning and Screen size:

The interactive wall project used by Fraunhofer consisted of four adjacent segments, each with a 40 inch wide and 23 inch height screen, while LeapMotion's covering area only extends to 23 inch semi-spherical area. When deploying the LeapMotion exactly at the center of the screen, the very far ends of the segment's screen are not entirely covered. However a conceptual solution for the problem is could be approached through horizontally combining multiple devices. The suggested solution is further discussed in the Future work section. Another limitation that was addressed specially in public events, is that in order to have a full coverage for the entire physical space in front of the public display, the LeapMotion sensor and Kinect has to be carefully positioned at the correct angels.

5.4 Future work

We would like to think of our study results to be the first cornerstone in the path towards making public display distance-aware. Hence, we present in this section some of the basic ideas that would support further innovations in this area.

Horizontal integration:

In our study we mainly focused on the integration of multiple sensing technologies in the horizontal dimension perpendicular to the public display. However an interesting research room is still vacant in the horizontal combination in the direction adjacent to the public display. We would imagine an application scenario where the content is stretched over two adjacent public displays and a combination of multiple adjacent LeapMotions/Kinects would be beneficial. As discussed in the limitation section above, the proposed scenario with multiple LeapMotion to gain double the coverage area is suggested to be a valid solution.

Enhanced smoothing algorithm:

In the course of the study we have presented an enhanced smoothing algorithm, that solved the problem caused by the two different technologies used.

However the algorithm was only tested in lab environments under various test cases but with no real user study. We would definitely encourage further inspections in this area to gain extra validation or enhancement for the implemented logic.

Across zones gestures: All the defined and tested gestures presented in both user studies were entirely within zones gestures. In the second study only the pointing task was executed and inspected across zones. Thus we embody the necessity to examine and define gestural interaction that goes cross zone. In other words, further research should be conducted to test and evaluate gestural interaction that starts in one zone and ends in another during the course of one single actio.

Appendices

Appendix A

System Usability Scale Questionnaire

Interaktives Display-Spiel

- Benutzerstudie –

(Juni/Juli 2013)

Universität Stuttgart

Fraunhofer IAO

Information

Vielen Dank für die Teilnahme an dieser Studie.

Der Zweck dieser Studie ist die Erforschung verschiedener Interaktionstechniken mit öffentlichen Großbildschirmen. Speziell interessiert uns der Einfluss verschiedener Technologien für Gesten-Interaktion auf Usability und User Experience.

Die im Rahmen dieser Studie gesammelten Daten werden ausschließlich für Lehr- und Forschungszwecke eingesetzt. Daten werden ausschließlich anonym erhoben.

Die Teilnahme ist freiwillig. Teilnehmer können jederzeit ohne Grund ihre Teilnahme an der Studie beenden. Bei weiteren Fragen zur Studie wenden Sie sich bitte an

Dr. Florian Alt
Universität Stuttgart
Lehrstuhl für Mensch-Computer Interaktion
Email: florian.alt@vis.uni-stuttgart.de

Interaktives Display-Spiel

- Benutzerstudie –

(Juni/Juli 2013)

Universität Stuttgart

Fraunhofer IAO

Zustimmung zur Teilnahme an der Studie

❑ Ich habe das Informationsblatt gelesen und verstanden.

❑ Ich habe den Zweck der Studie verstanden und bin bereits daran teilzunehmen.

❑ Ich habe verstanden, dass ich meine Teilnahme an der Studie jederzeit beenden kann.

Teilnehmer ID ______________________ (vom Studienleiter auszufüllen)

Falls wir Sie für weitere Fragen kontaktieren dürfen, geben Sie bitte ihre Email / Telefonnummer an.

____

Teilnehmer ________________________ Datum _____________

____

Studienleiter ________________________ Datum _____________

This study is funded in part by European Union Seventh Framework Programme (FP7/2007-2013) under the project "pd-net - Towards Future Pervasive Display Networks"

Interaktives Display-Spiel

- Benutzerstudie –

(Juni/Juli 2013)

Universität Stuttgart

Fraunhofer IAO

Fragebogen

A. Demographie

Geschlecht: ❑ männlich ❑ weiblich

Alter: ____________ Beruf / Studiengang: _________________________________

Besitzt du eine Spielekonsole (z.B. Xbox, Wii, etc.)?

Wenn ja, welche? ___________________________________

Wie oft spielst du Computerspiele?

❑ täglich

❑ wöchentlich

❑ monatlich

❑ seltener / nie

B. User Experience

Bitte bewerte das System das du eben benutzt hast anhand der folgenden Kriterien.

furchtbar	❑	❑	❑	❑	❑	❑	❑	❑	❑	wunderbar
schwierig	❑	❑	❑	❑	❑	❑	❑	❑	❑	einfach
mangelhafte Kontrolle	❑	❑	❑	❑	❑	❑	❑	❑	❑	angemessene Kontrolle
langweilig	❑	❑	❑	❑	❑	❑	❑	❑	❑	stimulierend
starr	❑	❑	❑	❑	❑	❑	❑	❑	❑	flexibel

Vom Studienleiter auszufüllen!

Condition: ❑ L ❑ K ❑ LK

Interaktives Display-Spiel

- Benutzerstudie –

(Juni/Juli 2013)

Universität Stuttgart

Fraunhofer IAO

C. Benutzbarkeit

Bitte beantworte die folgenden Fragen zum System welches du gerade verwendet hast.

1. **Ich denke, dass ich das System gerne häufig benutzen würde**

   stimme voll und ganz zu ❑ ❑ ❑ ❑ ❑ stimme überhaupt nicht zu

2. **Ich fand das System unnötig komplex.**

   stimme voll und ganz zu ❑ ❑ ❑ ❑ ❑ stimme überhaupt nicht zu

3. **Ich fand das System einfach zu benutzen.**

   stimme voll und ganz zu ❑ ❑ ❑ ❑ ❑ stimme überhaupt nicht zu

4. **Ich glaube, ich würde die Hilfe einer technisch versierten Person benötigen, um das System benutzen zu können.**

   stimme voll und ganz zu ❑ ❑ ❑ ❑ ❑ stimme überhaupt nicht zu

5. **Ich fand, die verschiedenen Funktionen in diesem System waren gut integriert.**

   stimme voll und ganz zu ❑ ❑ ❑ ❑ ❑ stimme überhaupt nicht zu

6. **Ich denke, das System enthielt zu viele Inkonsistenzen.**

   stimme voll und ganz zu ❑ ❑ ❑ ❑ ❑ stimme überhaupt nicht zu

7. **Ich kann mir vorstellen, dass die meisten Menschen den Umgang mit diesem System sehr schnell lernen.**

   stimme voll und ganz zu ❑ ❑ ❑ ❑ ❑ stimme überhaupt nicht zu

8. **Ich fand das System sehr umständlich zu nutzen.**

   stimme voll und ganz zu ❑ ❑ ❑ ❑ ❑ stimme überhaupt nicht zu

9. **Ich fühlte mich bei der Benutzung des Systems sehr sicher.**

   stimme voll und ganz zu ❑ ❑ ❑ ❑ ❑ stimme überhaupt nicht zu

10. **Ich musste eine Menge lernen, bevor ich anfangen konnte das System zu verwenden.**

    stimme voll und ganz zu ❑ ❑ ❑ ❑ ❑ stimme überhaupt nicht zu

Bibliography

[1] Florian Alt, Thomas Kubitza, Dominik Bial, Firas Zaidan, Markus Ortel, Björn Zurmaar, Tim Lewen, Alireza Sahami Shirazi, and Albrecht Schmidt. Digifieds: insights into deploying digital public notice areas in the wild. In *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia*, MUM '11, pages 165–174, New York, NY, USA, 2011. ACM.

[2] Florian Alt, Stefan Schneegass, Michael Girgis, and Albrecht Schmidt. Cognitive effects of interactive public display applications. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays*, pages 13–18. ACM, 2013.

[3] Seok-Hyung Bae, Takahiro Kobayash, Ryugo Kijima, and Won-Sup Kim. Tangible nurbs-curve manipulation techniques using graspable handles on a large display. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 81–90. ACM, 2004.

[4] Keith Bain, Sara H Basson, and Mike Wald. Speech recognition in university classrooms: liberated learning project. In *Proceedings of the fifth international ACM conference on Assistive technologies*, pages 192–196. ACM, 2002.

[5] Patrick Baudisch and Gerry Chu. Back-of-device interaction allows creating very small touch devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1923–1932. ACM, 2009.

[6] Gilbert Beyer, Florian Alt, Stefan Klose, Karsten Isakovic, Alireza Sahami Shirazi, and Albrecht Schmidt. Design space for large cylindrical screens. In *Proc. 3rd Workshop on Pervasive Advertising and Shopping, Pervasive*, 2010.

[7] Sebastian Boring, Dominikus Baur, Andreas Butz, Sean Gustafson, and Patrick Baudisch. Touch projector: mobile interaction through video. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 2287–2296. ACM, 2010.

[8] Maged N Kamel Boulos, Bryan J Blanchard, Cory Walker, Julio Montero, Aalap Tripathy, Ricardo Gutierrez-Osuna, et al. Web gis in practice x: a microsoft kinect natural user interface for google earth navigation. *International journal of health geographics*, 10(1):45, 2011.

[9] Harry Brignull and Yvonne Rogers. Enticing people to interact with large public displays in public spaces. In *Proceedings of INTERACT'03*, pages 17–24, 2003.

[10] John Brooke. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189:194, 1996.

[11] Andreas Bulling, Florian Alt, and Albrecht Schmidt. Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 3011–3020. ACM, 2012.

[12] Claude Cadoz. Le geste canal de communication homme/machine: la communication «instrumentale». *TSI. Technique et science informatiques*, 13(1):31–61, 1994.

[13] Robert Stephen Feldman and Bernard Rimé. *Fundamentals of nonverbal behavior.* Cambridge University Press, 1991.

[14] Sidney Fels and Geoffrey Hinton. Glove-talkii: an adaptive gesture-to-formant interface. In *Proceedings of the SIGCHI conference on Human factors in*

computing systems, pages 456–463. ACM Press/Addison-Wesley Publishing Co., 1995.

[15] Barak Freedman, Alexander Shpunt, Meir Machline, and Yoel Arieli. Depth mapping using projected patterns, April 2 2008. US Patent App. 12/522,171.

[16] Thomas Fysh and JF Thompson. A wii problem. *JRSM*, 102(12):502–502, 2009.

[17] Anthony G Greenwald. Within-subjects designs: to use or not to use? *Psychological Bulletin*, 83(2):314, 1976.

[18] Edward Twitchell Hall and Edward T Hall. *The hidden dimension*. Anchor Books New York, 1969.

[19] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568. ACM, 2011.

[20] Michael RM Jenkin, Allen D Jepson, and John K Tsotsos. Techniques for disparity measurement. *CVGIP: Image Understanding*, 53(1):14–30, 1991.

[21] Jong-wook Kang, Dong-jun Seo, and Dong-seok Jung. A study on the control method of 3-dimensional space application using kinect system. *International Journal of Computer Science and Network Security*, 11(9):55–59, 2011.

[22] Christian Küblbeck and Andreas Ernst. Face detection and tracking in video sequences using the modifiedcensus transformation. *Image and Vision Computing*, 24(6):564–572, 2006.

[23] Anthony Lauder and Stuart Kent. Precise visual specification of design patterns. In *ECOOP'98—Object-Oriented Programming*, pages 114–134. Springer, 1998.

[24] Annie Luciani, Matthieu Evrard, Damien Couroussé, Nicolas Castagné, Claude Cadoz, and Jean-Loup Florens. A basic gesture and motion format for virtual reality multisensory applications. *arXiv preprint arXiv:1005.4564*, 2010.

[25] I Scott MacKenzie and William Buxton. Extending fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 219–226. ACM, 1992.

[26] David McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago Press, 1992.

[27] Daniel Michelis. *Interaktive Großbildschirme im öffentlichen Raum*. Springer, 2009.

[28] 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.

[29] Jörg Müller, Robert Walter, Gilles Bailly, Michael Nischt, and Florian Alt. *Looking Glass: A Field Study on Noticing Interactivity of a Shop Window*. ACM, Austin, Texas, 2012.

[30] Ara V Nefian, Luhong Liang, Xiaobo Pi, Xiaoxing Liu, and Kevin Murphy. Dynamic bayesian networks for audio-visual speech recognition. *EURASIP Journal on Advances in Signal Processing*, 2002(11):1274–1288, 1900.

[31] Vladimir I Pavlovic, Rajeev Sharma, and Thomas S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):677–695, 1997.

[32] Peter Peltonen, Esko Kurvinen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, John Evans, Antti Oulasvirta, and Petri Saarikko. It's mine, don't

touch!: interactions at a large multi-touch display in a city centre. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 1285–1294, New York, NY, USA, 2008. ACM.

[33] Thorsten Prante, Carsten Röcker, Norbert Streitz, Richard Stenzel, Carsten Magerkurth, Daniel Van Alphen, and Daniela Plewe. Hello. wall–beyond ambient displays. In *Adjunct Proceedings of Ubicomp*, pages 277–278. Citeseer, 2003.

[34] Lawrence Rabiner and Biing-Hwang Juang. Fundamentals of speech recognition. 1993.

[35] Jun Rekimoto. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*, pages 21–27. IEEE, 2001.

[36] Carmen Santoro Riccardo Dini, Fabio Paternò. An environment to support multi-user interaction and cooperation for improving museum visits through games.

[37] Henry Scheffe. *The analysis of variance*, volume 72. Wiley. com, 1999.

[38] Thomas Schlömer, Benjamin Poppinga, Niels Henze, and Susanne Boll. Gesture recognition with a wii controller. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*, pages 11–14. ACM, 2008.

[39] Katie A Siek, Yvonne Rogers, and Kay H Connelly. Fat finger worries: how older and younger users physically interact with pdas. In *Human-Computer Interaction-INTERACT 2005*, pages 267–280. Springer, 2005.

[40] Dorothy Sparks, Daniel Chase, and Lisa Coughlin. Wii have a problem: a review of self-reported wii related injuries. *Informatics in primary care*, 17(1):55–57, 2009.

[41] Norbert A Streitz, Carsten Roecker, Thorsten Prante, Richard Stenzel, and Daniel van Alphen. Situated interaction with ambient information: Facilitating awareness and communication in ubiquitous work environments. In *Tenth International Conference on Human-Computer Interaction (HCI International 2003)*. Citeseer, 2003.

[42] Daniel Vogel and 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*, pages 137–146. ACM, 2004.

[43] Jacob O Wobbrock, Meredith Ringel Morris, and Andrew D Wilson. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1083–1092. ACM, 2009.

[44] Justin Young, Michael Lin, Alexander Bick, Ammar Sarwar, and Jack Dennerlein. Gestural workspaces for computer interaction configuration and performance. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 424–428. SAGE Publications, 2013.

[45] Thomas G Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, and Young Harvill. A hand gesture interface device. In *ACM SIGCHI Bulletin*, volume 18, pages 189–192. ACM, 1987.

Declaration

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

---

place, date, signature