ABSTRACT

Giving the Users a Hand: Towards Touchless Hand Gestures for the Desktop

Alvin Jude Hari Haran, M.S.

Advisor: G. Michael Poor, Ph.D

Touchless, mid-air, gesture-based interactions have recently moved out of laboratories and Hollywood movies and into the hands of users. There is little difference in the interaction style and techniques used today from that of the 1980's, despite advances in the technology enabling this interaction. For this interaction to achieve mainstream popularity, and to be as ubiquitous as the keyboard or the mouse, common problems such as the "Gorilla Arm Syndrome" will have to be addressed. Additionally, the common use-case such as gestural navigation, selection, and manipulation will need to be improved and eventually standardized. This thesis presents solutions to existing problems and introduces possible interaction techniques that allows users to perform the actions above. This is expected to pave the way for touchless mid-air hand gestures to be a ubiquitous form of interaction on the desktop.

Giving the Users a Hand: Towards Touchless Hand Gestures for the Desktop

by

Alvin Jude Hari Haran, B.S.

A Thesis

Approved by the Department of Computer Science

Gregory D. Speegle, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree of

Master of Science

Approved by the Thesis Committee

G. Michael Poor, Ph.D, Chairperson

Erich J. Baker, Ph.D.

Alexander Beaujean, Ph.D.

Accepted by the Graduate School
August 2014

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES xi

LIST OF TABLES xiii

ACKNOWLEDGMENTS xiv

DEDICATION xv

1 Introduction 1

1.1 Motivation 1

1.2 Definition 2

1.3 Gestures: Past, Present and Future 2

1.3.1 A Brief History 2

1.3.2 Current Trends 3

1.3.3 Future Plans 3

1.4 Problems, Challenges, Limitations 4

1.5 Solution Scope 5

1.6 Structure 7

1.6.1 Manuscript 1 7

1.6.2 Manuscript 2 7

1.6.3 Manuscript 3 9

1.6.4 Manuscript 4 10

2 Literature Review 12

2.1 Beyond the Desktop 12

2.2 Interaction 14

2.2.1 Gesture-based Interaction 14
2.2.2 Barehanded Interaction 14
2.2.3 Gestural Manipulation 15
2.2.4 Multimodal 15
2.3 Fatigue 15
2.4 Pointing Device Evaluation 16
2.4.1 Fitts' Law 16
2.4.2 Gestural Pointing 18

3 Personal Space: User Defined Gesture Space for GUI Interaction 20
3.1 Abstract 20
3.2 Introduction 20
3.3 Related Works 22
3.3.1 Gesture-based Interaction 22
3.3.2 Barehanded Interaction 23
3.3.3 Ergonomics and Usability 23
3.4 Pilot 1 24
3.4.1 Approach 24
3.4.2 Experimental Design 24
3.4.3 Quantitative Results 26
3.4.4 Qualitative Results 26
3.5 Pilot 2 27
3.5.1 Approach 27
3.5.2 Experimental Design 27
3.5.3 Quantitative Results 28
3.5.4 Qualitative Results 28

3.6 Discussion and Future Works 29

3.7 Conclusion 30

3.8 References 30

4 An Evaluation of Touchless Hand Gestural Interaction for Pointing Tasks with Preferred and Non-preferred Hands 32

4.1 Abstract 32

4.2 Introduction 32

4.3 Related Works 34

4.3.1 Pointing Device Evaluation 34

4.3.2 Handedness 35

4.3.3 Gestural Pointing 36

4.3.4 Learning Effects 37

4.3.5 Motor Learning 37

4.4 Methodology 38

4.4.1 Task 38

4.4.2 Breaking the Standard Evaluation Method 39

4.4.3 Input Devices & Interaction 41

4.4.4 Participants 43

4.4.5 Procedure 43

4.4.6 Design 44

4.5 Results 44

4.5.1 Preferred Hand vs Non-Preferred Hand 44

4.5.2 Time 44

4.5.3 Throughput 45

4.5.4 Performance Improvement 46

4.5.5 Degradation 48
4.5.6 Fatigue and Discomfort 49
4.6 Analysis and Discussion 50
4.6.1 Literature Comparison 50
4.6.2 Metrics 51
4.6.3 Performance Improvement 52
4.6.4 Degradation 53
4.6.5 Gestures 53
4.7 Conclusion 54

5 Evaluating Touchless Hand Gestures and Speech for Point and Select Tasks 55
5.1 Abstract 55
5.2 Introduction 55
5.3 Related Works 57
5.4 Methods 58
5.4.1 Participants 58
5.4.2 Task 58
5.4.3 Input Methods 59
5.4.4 Performance Evaluation 60
5.5 Results and Analysis 60
5.5.1 Adapting Speech to Gestures 61
5.5.2 Hyperarticulation 62
5.5.3 Throughput 62
5.5.4 Completion Time 63
5.5.5 Errors 63
5.6 Discussion and Future Work 64

5.7 Conclusion 65

6 Grasp, Grab or Pinch? Identifying User Preference for Touchless Gestural Manipulation using Rule-Based Detection 66

6.1 Abstract 66

6.2 Introduction 66

6.3 Related Works 68

6.3.1 Grab Gesture 69

6.3.2 Pinch Gesture 71

6.3.3 Other Gestures 72

6.4 Design 72

6.4.1 Prototype Design 72

6.4.2 Interaction Design 73

6.5 User Study 75

6.5.1 Participants 75

6.5.2 Procedure 75

6.6 Results 77

6.6.1 Observations 78

6.6.2 Feedback 79

6.6.3 Variations in Gestures 79

6.7 Contribution 80

6.8 Discussion and Implication 81

6.8.1 Recommended Gesture 81

6.8.2 Gestures and Cultural Significance 82

6.8.3 Robust Gestural Recognition 82

6.8.4 Intuitive vs Learned Gestures 83

6.8.5 Usability 83
6.9 Future Works 83
6.10 Conclusion 84
6.11 Acknowledgments 84

7 Discussion, Future Work and Conclusion 85
7.1 Discussion 85
7.1.1 Calibration 85
7.1.2 Transformation 86
7.1.3 Z-coordinate Issues 87
7.2 Future Work 90
7.2.1 Modeling the user space 90
7.2.2 Elbow Position 91
7.2.3 Interaction Use-case 91
7.2.4 Gestures vs Mouse 91
7.2.5 Accessibility 92
7.3 Conclusion 92

APPENDICES 94

APPENDIX A Solution Design 95
A.1 Problem 95
A.2 Solution 95

APPENDIX B Perfomance Measurements 99

APPENDIX C Exit Surveys 102
C.1 Manuscript 2 102

C.2 Manuscript 3 104

BIBLIOGRAPHY 105

LIST OF FIGURES

1.1 Commonly used gestural interaction style 6

3.1 User defining his gesture space during calibration 22

3.2 De facto gestural input and Personal Space 23

3.3 Natural arm range and corresponding flat plane 25

3.4 Distribution of completion time for Pilot 1. 26

3.5 Distribution of completion time for Pilot 2. 28

3.6 Total distance travelled by the cursor during Pilot 2. 29

4.1 A scaled version of the targets used 39

4.2 Multi-Directional task implementation 40

4.3 Personal Space calibration:Constructing the user's space 42

5.1 Three types of targets and actions 59

5.2 Hover targets used in the task 60

5.3 Calibration process to model gesture space 61

6.1 The grasp gesture 68

6.2 The grab gesture 69

6.3 The pinch gesture 70

6.4 Rules to transition between open and close states of the palm 74

6.5 A more comfortable version of the *grasp* gesture 80

6.6 The *grasp* gesture with "Hitchhiker's Thumb" 81

7.1 Building the interaction space. 88

A.1 Commonly used gestural interaction method. 96

A.2 Comparison of the actual and calibrated space. 97

B.1 Target distance vs completion time. 99

B.2 Graphical representation of device and hand performance 100

B.3 Graphical representation of device and hand performance 100

LIST OF TABLES

4.1 Means of completion time 44

4.2 Mean movement time per trial in milliseconds. 45

4.3 Means and standard deviation of throughput 46

4.4 Performance increase as percentage of improvement 47

4.5 Performance increase as measured with Cohen's d 48

4.6 Degradation calculated as percentage of increase in movement time . 49

4.7 Degradation calculated as percentage of decrease in throughput . . . 50

4.8 Degradation between hands measured with Cohen's d. 50

5.1 Means and standard deviation of completion time and throughput . 64

6.1 Ranking of user preferences 78

ACKNOWLEDGMENTS

To Dr. G. Michael Poor, for giving me a free hand in choosing and running with my ideas, while still keeping me on path. Darren Guinness for his dedication in contributing to the code despite having his hands full. Dr. Greg Hamerly for expanding my view, Dr. Greg Speegle for guidance, Dr. Young-rae Cho for his help in clustering, Dr. Erich J. Baker for pointing me to a rheumatologist and Dr. Pablo Rivas in aiding in our search for participants.

Also Li Guo for helping me with the math, Ryan Henning for helping me with regression, Edward Collier for always being subject 0 in our experiments, Christian Marcantel for help with graphics, and all graduate students in Baylor University's Department of Computer Science, for their willingness to participate in our experiments and studies. To all our volunteers, without whom there will be nothing to report.

To all staff and faculty of Baylor University's Department of Computer Science, especially Pat Hynan, George Gonzales, Sharon Humphrey and Michelle Aars for solving the more tedious problems inherent to research, and therefore allowing me to focus.

Finally to my roommate Roman Smetana for putting up with me for 2 years.

To those devoted to teaching

CHAPTER ONE

Introduction

1.1 Motivation

Current methods of interacting with a desktop uses the age-old WIMP (Windows, Icons, Menus and Pointing devices) paradigm (Argyros and Lourakis 2006), where a pointing device such as a mouse or a touchpad is used to interact with on-screen widgets. An update to the WIMP paradigm which includes gesture-based interaction has been proposed as far back as 1993 (Nielsen 1993). This attempt to update the WIMP paradigm predicted amongst other things that gestures will play a large role in what was referred to as "post-WIMP interactions" (van Dam 1997).

In recent years, there has been some interest in developing "Natural User Interfaces" (NUI) as a post-WIMP interaction technique. NUI draws it's strength from incorporating interactions that are natural to most human beings, such as speech, vision, handwriting and gestures. While there is no concrete framework or definition on what constitutes a 'Natural Interaction', a good way to think about it is provided by Wigdor and Wilcox (Wigdor and Wixon 2011). They recommend that designers think of the NUI approach as a way to design interfaces that allows the user to leverage innate talents, and previously learned skills and experience in which the user is already an expert. This focuses on creating user interfaces that allows the user to feel and act "like a natural". Most users are experts at manipulating the real world with their hands. It only logically follows that interaction techniques which allow the user to manipulate virtual objects with their hands as they would in the real world, would come naturally to them.

Closely related to NUI is the Reality-Based Interaction (RBI) framework, which pushes for interfaces to be even closer to real-world interaction techniques (Jacob, Girouard, Hirshfield, Horn, Shaer, Solovey, and Zigelbaum 2008). The themes

proposed in the RBI framework helps designers to draw inspiration from the real world in designing interfaces and interactions, and to perform a tradeoff analysis wherever required. The authors of this framework suggest that new post-WIMP interactions draw it's strengths by exploiting users' pre-existing knowledge about their bodies and objects in the world. This allows users to interact with their digital world in a manner closer and more similar to interactions performed in the real (non-digital) world. Gesture-based interactions fit well into this assessment, it allows users to utilize their existing skills for controlling and coordinating their bodies, in the digital world as they would in the real world.

Both NUI and RBI gives us strong motivation to promote the use of gestures in our interactions with computers. Given the multitude of devices currently available and even more introduced each day, there are many different areas of computer interaction where gesture-based interactions can be used. This includes public displays, large displays, micro-interactions with personal devices, and the desktop. This study draws inspiration on previous work done in all these areas, but focuses strongly on introducing gesture-based interaction techniques to the desktop.

1.2 Definition

The term "gesture-based interaction" in HCI carries different meanings in different context. It could be used to refer to facial gestures or expressions (Lyons 2004), touch-based gestures on touchscreen displays (Sears, Plaisant, and Shneiderman 1991), or mid-air hand gestures. This study focusses only on the last item in the preceding list, with a focus on palm recognition as opposed to the full hand.

1.3 Gestures: Past, Present and Future

1.3.1 A Brief History

Touchless gestural interactions is not a new concept. Bolt introduced this concept in the seminal paper "Put-That-There" in 1980 (Bolt 1980) which also features

speech commands for manipulation of objects. This style of multimodal interaction received significant attention in the 1990's due to it's novel approach with exciting possibilities, but the interest dropped off as natural gesturing was deemed too hard for automatic processing systems. Additionally technical implementations that are over-simplified tend to end up being rigid and brittle (Kopp 2013).

Despite this, users were still introduced to the possibilities of gestural interaction through science fiction. Various movies utilized gestural interaction very similar to Bolt's "Put-That-There", including Minority Report (2002) and The Matrix Reloaded (2003). In *Chapter Three* of this study, we demonstrate why this exact form of interaction is not feasible due to quick onset of fatigue, and introduce a novel solution.

1.3.2 Current Trends

In recent years, commodity hardware such as the Leap Motion and the Xbox Kinect has introduced gesture-based interactions to the masses. The former catered towards interacting with the desktop and focuses solely on the user's hand, while the latter performs full body recognition and has generally been used for gaming. The Wii Remote is also a gaming-focused controller that preforms gesture recognition, but is not in the same family, as it requires the user to hold onto the controller, as opposed to a touchless interactions afforded by the Leap Motion and the Xbox Kinect. Gestures have also been incorporated into mobile phones and smart TVs to perform specific actions such as scrolling or flipping the channel.

This effectively took gesture-based interaction out of research labs and science fiction and into the hands of the users.

1.3.3 Future Plans

With gesture recognition technologies already in the market, and more set to follow, gestures are becoming a larger part of our interaction. Ensuring usability

and standardizing interactions becomes a stronger criteria. For example, users who have used gestures tend to be more confident and faster when using gestures with completely different devices or interfaces (Vatavu 2012). Going back to our definition of Natural User Interfaces, we will need to consider how to design interactions that allow users to feel and act "like a natural", regardless of their prior expertise (or lack thereof).

1.4 Problems, Challenges, Limitations

Grandhi, Joue, and Mittelberg (Grandhi, Joue, and Mittelberg 2011) identified two challenges to touchless gestural interactions: 1) the recognition challenge and 2) the vocabulary challenge. Nielsen et al (Nielsen, Störring, Moeslund, and Granum 2004) used the terms "technology based vocabulary" and "human based gesture vocabulary" which can be generalized to the recognition challenge and vocabulary challenge respectively. The recognition challenge deals with achieving accurate and meaningful gesture recognition while the vocabulary challenge deals with identifying natural, intuitive and meaningful gesture vocabularies appropriate for the tasks. (Grandhi, Joue, and Mittelberg 2011) claim the former challenge has received more research attention than the latter, causing gestures to be prescribed to users based on their ease of implementation. The attention traditionally given to the recognition challenge has resulted in commodity hardware such as the Leap Motion controller (Motion 2012) which is capable of hand recognition with sub-millimeter accuracy (Weichert, Bachmann, Rudak, and Fisseler 2013) and even comes bundled with a developers kit. This now allows us as interaction researchers to focus our research on the vocabulary challenge, with the recognition challenge a secondary consideration.

While commodity hardware does exist, these devices are targeted towards non-work, non-purposeful interactions often used in isolation (such as gaming), meanwhile the idea of the desktop was the exact opposite: to gather as much functionality as possible in one computer (Bødker 2006). But there is good reason to extend the use

of touchless gestural interaction to the desktop: they provide a natural and intuitive method of interaction by building on users' pre-existing knowledge of the everyday, non-digital world (Jacob, Girouard, Hirshfield, Horn, Shaer, Solovey, and Zigelbaum 2008), and could therefore be employed for productive use with minimal training or knowledge of existing devices. For example we can imagine using gestures to grab a file and move it to a different directory, or as an input device by users with hand impairments who are unable to use a mouse. Before these interaction techniques can be ubiquitous however, user research will need to be performed and standards will need to be defined. These new interactions could potentially be effective and even fun, however, in the rush to push interactions to the user, designers tend to ignore well established principles preveiously validated by user research (Norman and Nielsen 2010). Ignoring these principles is akin to ignoring the user, forcing them to adapt to the interface, instead of vice versa.

One glaring example of this scenario is the interaction style currently prescribed for gestural interaction where users are expected to hold their arms outstretched in front of them, as shown in figure 1.1. This is highly uncomfortable and is an infamous problem in HCI, commonly referred to as the "Gorilla Arm Syndrome" (Carmody 2010). Something known to HCI researchers since the 1980's.

1.5 Solution Scope

This study focuses on addressing known problems in gestural interaction, while identifying how gestural interaction can best be used on the desktop. Existing principles are heavily considered and the user is a pivotal point of the study. To this end, experiments were designed to identify users' performance and preference in using gestural interaction.

The problem addressed in this study revolves around everyday interactions with the desktop. As a result, the interaction style proposed focuses on the palm as input, as opposed to the entire arm which is perhaps more suited for interacting with

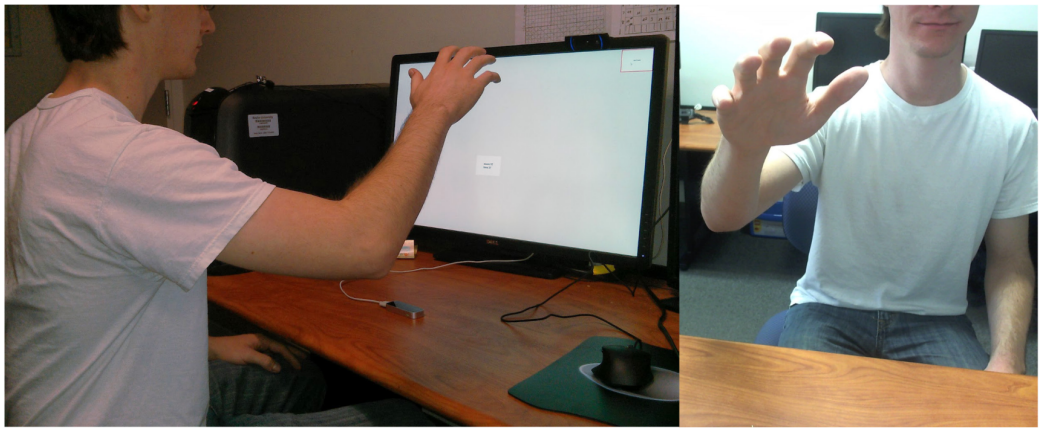

Figure 1.1. Commonly used gestural interaction style where the users arms are held outstretched.

large displays. Additionally, prolonged interaction is prioritized as opposed to short micro-interactions, as the former is the norm when interacting with the desktop. The experiments designed were designed with these factors in mind, as such there were subtle factors that were considered, for example participants would perform interactions from a sitting-down position as it is the norm in desktop interactions, as opposed to public displays where a standing-up position may be assumed.

As a result of the focus of this study, there is no claim made on the generalizability of the solutions and recommendations to other mid-air hand gesture interfaces. Public displays utilizing gestural interaction for example may not face the same challenges faced by the desktop. Likewise, interacting with a mobile device using mid-air gestures might have it's own set of challenges, neither of which is explored in this study.

The main purpose of this study is to introduce and evaluate gestural interaction methods for interacting with the desktop through user studies. In doing so, three main functions of gestures are considered: gestural navigation, gestural selection, and gestural manipulation.

1.6 Structure

This study contains four self-contained manuscripts, each of which adds to the overarching theme, which is to ease the introduction of gestural interaction to the desktop. The four manuscripts are included in this study as chapters 3–6 and are described below.

1.6.1 Manuscript 1

The first manuscript in *Chapter Three* presents the interaction designed to assist with fatigue. This interaction, referred to as the Personal Space approach allows the user to rest their elbow on a surface throughout the interaction. Resting the elbow in a regular linear-mapped interaction would make it difficult for users to reach the far corners of the screen, for example when using the right hand for interaction, the top left corner of the screen cannot be reached without lifting the elbow. To solve this issue, the users input space is modeled following a calibration phase. In our particular implementation, we build a flat quadrilateral surface which is then affine-mapped to the screen space. This approach allows for a rough approximation and does not accurately model the user's space which is a curved surface. Nonetheless through 2 pilot studies, we demonstrate how this approach allows for an interaction which is no worse in performance than the de facto method where the arm is held outstretched, but significantly reduces the fatigue problem.

This solution is the first and most important aspect of this study. The Personal Space approach allows an interaction that is sustainable across prolonged periods, making it suitable for desktop interaction.

1.6.2 Manuscript 2

Following the pilot study above where the interaction is shown to be feasible for prolonged interactions, a comprehensive analysis of this interaction is done and presented in *Chapter Four* which includes the second manuscript. The performance

of this interaction is benchmarked against the two most common input devices used in the desktop environment: the mouse and the touchpad.

The experiment presented in this manuscript is designed to test for two hypotheses:

(1) There will be minimum difference in performance between the preferred and non-preferred hands when using gestures compared to other devices.

(2) There will be more performance increase between rounds when using gestures compared to other devices.

These hypotheses are based on RBI framework and NUI interaction design. Since gestures are closer to the real/non-digital world, it has to be more natural to the user, therefore making it easier to be used even with the non-preferred hand. Additionally, this would also mean this interaction would be easier to learn, which would be reflected in higher performance increase between two rounds.

The results of the experiment were positive; there was least amount of difference between hands and highest increase in performance between rounds. There were also other interesting findings reported in this study. The performance increase between hands were almost equal between both hands on all devices. The mouse for example showed no performance increase on either hand. This is interesting as it implies the knowledge of the mouse has been transferred to the non-preferred hand, thus the lack of improvement. We also found that the performance increase when using the touchpad is closer to that of gestures than the mouse. This is despite the fact that the touchpad is an ubiquitous device, at least compared to gestures. We initially expected participants to be better skilled at the touchpad and therefore exhibit lower performance increase between rounds.

This experiment shows us the possible direction for the utilization of gestures in desktop computing. The fact that the performance of gestures is demonstrated to be significantly lower than both the mouse and the touchpad does not rule it out

of the desktop. It is entirely possible to utilize this interaction alongside the mouse. The fact that gestural performance has little degradation between hands opens up the possibility of two-handed interactions or as an input device that augments and complements the mouse. There is also a possibility of using an interaction where the user alternates between hands for pointer manipulation. This form of interaction has potential to be used as an assistive device, perhaps for those with carpal tunnel syndrome, arthritis, or any other hand impairments.

1.6.3 Manuscript 3

Following the experiment above, this study shifts to identifying application areas with gestures. *Chapter Five* augments the gestural navigation interaction with speech input to perform gestural selection to evaluate the possibility of multimodal interaction. This mode of interaction was suggested as it allows for selection action – analogous to a click or double click with the mouse.

Speech and gestural multimodal interaction is not a new novel technique as it was used as far back as 1980 in Bolt's "Put-That-There" (Bolt 1980). However, it has been mostly ignored in recent research as it was deemed too hard and unusable, with most implementations being rigid and brittle (Kopp 2013). It has also been suggested that co-speech and gestures interaction cause a problem as people find it more difficult to think and speak at the same time (Shneiderman 2000).

This experiment combined best practices in both speech and gestures to isolate methods of implementing a workable interaction. For example two-word commands were chosen as it had better recognition than single-word commands, while it has been previously demonstrated that most commands require no more than three words (Hauptmann and McAvinney 1993). This avoids common problems such as hyperarticulation which tends to lead to cycle-of-failures.

The experiment detailed in this study had positive results in combining these two modes of interaction despite the potential issues mentioned above. In was demonstrated that introducing speech did not cause a degradation in gestural performance. An interesting discovery found during this experiment was that users tend to begin utterance of commands before the cursor reached the destination. This form of preemptive utterance lowered the overall speech overhead of the interaction and demonstrated the possibility of further improvements that can be achieved by users in real world interaction.

1.6.4 Manuscript 4

The 3 manuscripts described above revolves around gestural navigation. To include gestures on the desktop however, interactions would have to be designed to allow users to perform selection and manipulation. These actions were initially examined as part of pilot 1 in *Chapter Three*, however they were found to be unsuitable. The selection gesture implemented with a mid-air finger-tap was found to cause severe fatigue in our test subjects. While the manipulation gesture implemented with a grab action was brittle, as predicted in the previous section.

The fact that the gesture recognition system is brittle can be attributed to a gap in the recognition challenge, and therefore warrants more research. The focus of this study however was not the recognition challenge but the vocabulary challenge, however a successful implementation requires researchers to look at addressing both the recognition challenge and the vocabulary challenge simultaneously. If researchers choose to focus on only the recognition challenge, it is possible to create interactions that can be recognized by the input device but not natural to the user. Meanwhile if only the vocabulary challenge is addressed, it is possible for interactions to be designed that are too difficult for the input device to correctly and consistently identify.

The final manuscript included in *Chapter Six* attempts to bridge this problem by first identifying user preference for gestural manipulation through a high-fidelity

prototype. This differs from the method generally used which is via a Wizard-of-Oz approach. The high fidelity approach implemented here is beneficial as it utilizes an actual system which allows participants to actually manipulate objects on-screen and obtain direct feedback. In addition, it serves as a proof-of-concept in terms of technical implementation, demonstrating viability of real-world implementation.

This experiment serves to identify user preference of the three gestures proposed for gestural manipulation. Two of these gestures are commonly used in gestural manipulation (grab and pinch) and a new gesture referred to as grasp is also introduced here. The grasp gesture was built based on observations of user behavior during pilot 1. This gesture turned out to be preferred by most users. This knowledge is significant as it breaks away from the two commonly prescribed gestures in favor of one that is lesser known.

With this knowledge, researcher working on the recognition challenge can direct their attention to ensuring this gesture is recognized, thereby building implementations that no longer brittle, more natural to users and have a higher preference.

CHAPTER TWO

Literature Review

2.1 Beyond the Desktop

Our interaction with the real world happens in three dimensions. Our interaction with technology however is generally done with a Graphical User Interface (GUI) in two-dimensions. The mouse cursor for example moves on a simple X and Y axis, completely ignoring the possibility of the Z-axis. In recent years, attempts have been made to utilize the third dimension with both input and output.

One such attempt at incorporating the third dimension is with Tangible User Interfaces (Ishii 2008a). With tangible user interfaces, the primary output is coupled tightly with the input mechanism. The work referenced above uses a projection onto existing input device to illustrate how the position of buildings will affect wind flow. To examine this effect, users simply move scaled versions of buildings directly instead of doing so on a computer program via the GUI. This form of interaction was demonstrated to be more consistent with our method of interacting with the real world, and takes advantage of the human ability to grasp and manipulate physical objects (Ishii 2008b).

Interactions that extend the desktop do exist as demonstrated above. However these devices and interactions are targeted towards non-work, non-purposeful interactions often used in isolation or for very specific purposes. The idea of the desktop was the exact opposite: to gather as much functionality as possible in one computer (Bødker 2006). Implementing touchless gestural interaction on to the desktop allows a bridge between the very specific interactions mentioned and the idea of a general purpose computer in the desktop. Gestures are suitable in this scenario as it provides a natural and intuitive method of interaction by building on users' pre-existing knowledge of the everyday, non-digital world (Jacob, Girouard, Hirshfield,

Horn, Shaer, Solovey, and Zigelbaum 2008), and could therefore be employed for productive use with minimal training or knowledge of existing devices. For example we can imagine using gestures to grab a file and move it to a different directory, or as an input device by users with hand impairments who are unable to use a mouse.

Moving beyond the typical GUI interaction is hot current topic. To quote Steve Balmer, then CEO of Microsoft: *"I believe we will look back on 2010 as the year we expanded beyond the mouse and keyboard and started incorporating more natural forms of interaction such as touch, speech, gestures, handwriting, and vision – what computer scientists call the 'NUI' or natural user interface."* (Rogers, Sharp, and Preece 2011) Dan Norman and Jakob Nielsen were however not convinced. In their publication, pointedly called "Gestural Interfaces: A Step Backward in Usability", they argue that designers tend to ignore well-established principles in their rush to develop gestural or natural interfaces, leading to usability disasters. They also state that while these new interactions can be effective and even fun, they should not be "inflicted" onto users until it has been validated by user research (Norman and Nielsen 2010). Norman also argues that natural user interfaces have become a marketing term, and many of the NUI designs and interactions are in fact not natural (Norman 2010). This publication suggests that gestures will eventually be a valuable form of interaction that will give users more control, empowerment, convenience and delight. But more research will have to first be done, including in standardization of gestures.

This study aims to first employ a more natural way of interacting with gestures by modeling the system according to the user, instead of the other way around. Interaction with speech: other interaction technique known to be "natural" is also looked into. In doing so, this study aims to make gestures truly natural to the user and not just another marketing buzzword. The preferred method for gestural manipulation is also looked into to provide recommendation for standardization of gestures. In doing so, this study adds to the literature of both natural interaction

techniques as well as standardization recommendations, the two necessary aspects to improving gestural interaction.

2.2 Interaction

2.2.1 Gesture-based Interaction

The application of gesture based interaction has been explored for over 3 decades, the first of which (Bolt 1980) used a multi-modal interaction style, where gestures were used only for pointing while actions analogous to pointing and clicking were done with voice commands. More recently, the application of gestures were shown to be a suitable mode of interaction with very large displays (Vogel and Balakrishnan 2005a), while (Segen and Kumar 2000) shows that gestures are useful for interacting with the everyday desktop, and (Poor, Tomlinson, Guinness, Jaffee, Leventhal, Zimmerman, and Klopfer 2013) shows how RBI themes used in gestural input can lead to higher accuracy. While gesture based research has traditionally revolved around the recognition of gestures and the various poses of the palm, the availability of commodity hardware with open SDKs such as the Leap Motion and Microsoft Kinect has allowed a shift towards interpretation and application of gestures such as (Lai, Konrad, and Ishwar 2012) and (Biswas and Basu 2011).

2.2.2 Barehanded Interaction

The "Come As You Are" design principle (Triesch and Von Der Malsburg 1998) states that users should not be required to wear a glove or specific markers to interact with the system (Wachs, Kölsch, Stern, and Edan 2011). This brings us to the barehanded interaction style, where the users interact with the computer without any device or wires attached. This has been explored and described in (Von Hardenberg and Bérard 2001) to be superior to traditional input devices. This approach was also utilized in (Freeman, Vennelakanti, and Madhvanath 2012) to demonstrate human-computer interaction with gestures that are natural and relaxed.

2.2.3 Gestural Manipulation

We use the term "gestural manipulation" to mean hand gestures used to hold or release which is in line with definitions currently used (Wachs, Kölsch, Stern, and Edan 2011, Karam et al. 2005, Quek, McNeill, Bryll, Duncan, Ma, Kirbas, McCullough, and Ansari 2002). As the focus of this study is desktop interaction, the focus is in the manipulation of two dimensional GUI elements and not three-dimensional objects, which is another area of gestural manipulation altogether. This form of manipulation is closer to that which is performed with the mouse with point-and-click interaction.

2.2.4 Multimodal

In an experiment which used simulation to recognize speech and gestures as input, Hauptmann and McAvinney reported that participants strongly preferred a combination of both inputs to either modality alone for graphics manipulation (Hauptmann and McAvinney 1993). The same results were reported by Mark Billinghurst (Billinghurst 1998). This formed the basis the investigation of multimodal interaction using gesture and speech in this study as a means to mouse replacement technologies.

2.3 Fatigue

The standard method of using gestures involves users holding their arms up to the display for long periods of time. This has been known to cause a fatigue problem known as the "Gorilla Arm Syndrome" (Yoo, Lee, and Ahn 2012, Carmody 2010, Wachs, Kölsch, Stern, and Edan 2011) and is considered to be a "known limitation" (Teixeira 2011) of gestural interaction. Segen and Kumar likewise found that one of the biggest issues with gestures was that it tended to cause fatigue after prolonged interaction (Segen and Kumar 2000).

A simple method of overcoming this issue was explored in (Freeman, Vennelakanti, and Madhvanath 2012) by allowing users to rest their elbows on a chair

armrest. Segen and Kumar observed the same solution, and both studies demonstrated this solution with a simulation or a Wizard-of-Oz experiment.

An experiment by Sambrooks and Madhvanath (Sambrooks and Wilkinson 2013) looked into comparing touch, gestures, and mouse interactions and reported that fatigue was not a factor in gestures and that there was no improvement in gestures over the course of the experiment. This study presents an argument to Sambrooks' study as it was believed to be counter-intuitive and contrary to existing literature. New devices generally do exhibit performance improvements (MacKenzie, Kauppinen, and Silfverberg 2001), and gestures being a more natural mode of interaction should too. The results are detailed in *Chapter Four*.

The Personal Space approach (Jude, Poor, and Guinness 2014) demonstrated reduction of fatigue with a similar approach and an actual experiment. This approach worked by building a model of the user's gesture space which was then affine-mapped to the output device through matrix manipulation. This solution reported to be equal in performance to the de-facto method where users hold their arms up, but managed to reduce the issue with fatigue.

This method forms the basis of this study. It is introduced and explained in *Chapter Three*, with a performance analysis in *Chapter Four*. The multimodal method in *Chapter Five* uses this technique for gestural input and adds speech-based input.

2.4 Pointing Device Evaluation

2.4.1 Fitts' Law

In 1954 Paul Fitts established a relationship between movement speed and accuracy in rapid motor movements (Fitts 1954). This relationship, generally referred to as Fitts' Law paved way to standardized pointing device evaluations used today. Fitts introduced what he initially termed Index of Performance, which has since been

referred to as throughput and calculated as the index of difficulty (ID) over movement time (MT) (Zhai 2004),

$$\text{Throughput} = \frac{ID}{MT}$$

where ID is twice the ratio of the distance(D) to width(W) and measured in bits:

$$ID = log_2\left(\frac{2D}{W}\right)$$

This definition has been recently updated (Soukoreff and MacKenzie 2004) to use the Shannon formulation which has been demonstrated to be a better measure of human performance, so:

$$ID = log_2\left(\frac{D}{W} + 1\right)$$

The original definition of distance(D) in the formula above measures the center point between two targets. This formula has been adjusted to use the distance between the starting point of the cursor to the ending point, known as effective distance (D_e). Width has also been updated to use effective width (W_e), which utilizes the idea that points in a normal distribution is $log_2((2\pi e)^{\frac{1}{2}} \times \sigma)$. $log_2((2\pi e)^{\frac{1}{2}}$ is equal to 4.133 and σ is the standard deviation of the actual endpoint distribution. The effective width is therefore calculated as 4.133σ (MacKenzie et al. 2013). It follows that current methods of calculating throughput uses an effective index of difficulty (ID_e) (Zhai 2004, Soukoreff and MacKenzie 2004) defined as:

$$ID_e = log_2\left(\frac{D_e}{W_e} + 1\right)$$

This adjusted definition of effective index of difficulty is used in our throughput calculation, where we calculate throughput as per recent recommendations:

$$\text{Throughput} = \frac{ID_e}{MT}.$$

Fitts' Law is widely accepted to be the standard method of evaluating pointing devices, and throughput has been established as the metric used to report performance. However, this study did not start with these methods in it's early stage. In *Chapter Three*, the interaction is evaluated with completion time, which is not a standard method, and has shown to be a naive metric (Zhai 2004). This was nonetheless deemed acceptable as it was a pilot study to examine the feasibility of this interaction.

Fitts' Law and throughput is heavily used in the subsequent stages of this study following the pilot. In *Chapter Four*, gestural interaction is evaluated against other pointing devices, throughput is used to report performance. The use of throughput and other evaluation standards are therefore discussed further here, and is recommended to readers interested in this aspect of the study. In *Chapter Five*, throughput is used to evaluate the impact of introducing speech to augment gestural interaction.

2.4.2 Gestural Pointing

Fitts' law has been shown to be applicable to gestural navigation in both two-dimensional and three-dimensional tasks (Pino, Tzemis, Ioannou, and Kouroupetroglou 2013). The aforementioned experiment utilized a 2D and 3D multi-directional using a Xbox Kinect, which found that Fitts' law was applicable in gestures similarly to the mouse in the 2D task and at a higher rate in the 3D task.

The Fitts' law task applied to pointing device evaluation generally uses clicks to trigger a selection action. Gestural input however does not have a simple method of selection by default. This study therefore uses a hover method for selection, where the selection action is triggered when the cursor hovers over a target for a predefined period. This method was done in other experiments (Sambrooks and Wilkinson 2013) with different durations. In the pilot studies the hover threshold is set to 1 second while subsequent experiments lower this duration to 500ms.

A different method of selection is introduced and evaluated in *Chapter Five*. Here, speech-based input is used for selection as first used in Bolt's experiment (Bolt

1980). This method is used in conjunction with the hover-based selection method used in *Chapter Four*.

CHAPTER THREE

Personal Space: User Defined Gesture Space for GUI Interaction

This chapter published as: Jude, A., G. M. Poor, and D. Guinness (2014). Personal Space: User Defined Gesture Space for GUI Interaction. In *CHI 14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA 14, New York, NY, USA, pp. 1615-1620. ACM.

3.1 Abstract

Reality-Based Interaction (RBI) [14] theorizes that realistic user interactions (UIs) are effective because they exploit users pre-existing knowledge about their bodies and objects in the world. Gesture based interaction allows users to relay information to a computer through body movement without physical contact with additional hardware such as a mouse or trackball. However, this interaction style requires the users to interact in a manner that is tailored for the system to recognize with very strict rules for bodily interaction, not toward a gesture space that is more natural for the user. In this paper we propose a natural method of gestural input through a user-defined 3-dimensional space. We conducted two pilot studies to assess the performance and usability of these augmented gestural pointing methods for cursor manipulation as compared to a standard mouse interaction as well as the current standard approach used in gestural input.

3.2 Introduction

Since the graphical user interface became the default user interface of personal computers, most input was done with users manipulating some input device with their hand (e.g. keyboard, mouse, etc). However, in the last twenty years, with the development of new interaction techniques, researchers have begun investigating using free-hands and gestures for user interaction. This type of interaction tends to

be an effective, intuitive, and natural interaction way for users to relay information to the computer [12].

One problem that plagues gestural input is referred to as the "Gorilla Arm Syndrome" [10]. Typically, designing a gestural input involves requiring the user to hold their arm(s) out in an unnatural position, causing shoulder and arm fatigue quite quickly. As interaction time increases, this fatigue can lead to poor user performance, physical discomfort, and decreased accuracy.

We propose an ergonomic interaction style that allows the users to manipulate the cursor, using their hand as a pointing device without the discomfort associated with Gorilla Arm Syndrome. This approach – henceforth referred to as "Personal Space" – involves having the user rest her elbow on a surface, while controlling the position of the cursor based on the position of her palm. In this position, the users range of motion is not naturally a rectangle (figure 3.1) and therefore does not directly map to the screen. Through a calibration phase, the user defines his preferred space to be mapped to his screen space. This space will not be a right-angled rectangle, and will not be an absolute map to the computer screen (figure 3.1). In our experiments, we prove that this non-linear mapping works just as well as a relative map in a right-angled rectangle space, but with significantly less reported fatigue.

In this paper, we describe two pilot studies to assess the performance (task completion time) and usability (self-reported fatigue) of this augmented gestural system for cursor manipulation as compared to a standard mouse interaction and the current standard approach for gestural input. Additionally we will discuss the implications and the direction of future research.

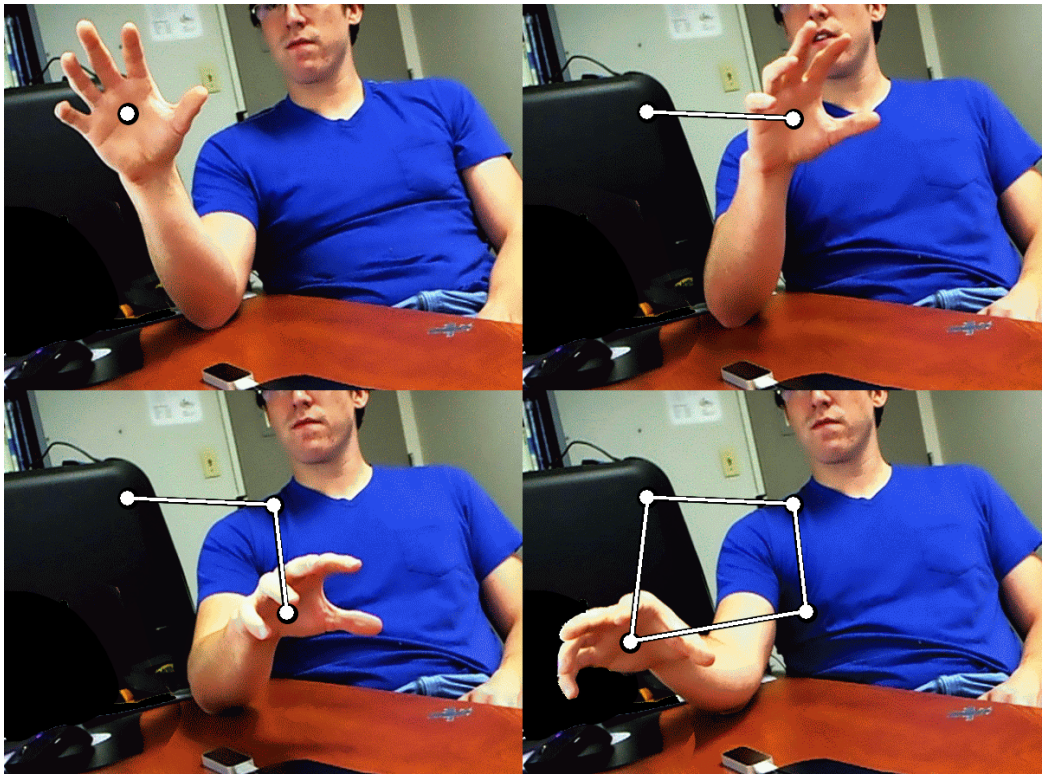

Figure 3.1. A user defining his gesture space during calibration.

3.3 Related Works

3.3.1 Gesture-based Interaction

The application of gesture based interaction has been explored for over 3 decades, the first of which [4] used a multi-modal interaction style, where gestures were used only for pointing while actions analogous to pointing and clicking were done with voice commands. More recently, the application of gestures were shown to be a suitable mode of interaction with very large displays [3], while [5] shows that gestures are useful for interacting with the everyday desktop, and [16] shows how RBI themes used in gestural input can lead to higher accuracy. While gesture based research has traditionally revolved around the recognition of gestures and the various poses of the palm, the availability of commodity hardware with open SDKs such as

Figure 3.2. De facto gestural input (left) and Personal Space (right).

the Leap Motion and Microsoft Kinect has allowed a shift towards interpretation and application of gestures such as [6] and [7].

3.3.2 Barehanded Interaction

The "Come As You Are" design principle [8] states that users should not be required to wear a glove or specific markers to interact with the system [12]. This brings us to the barehanded interaction style, where the users interact with the computer without any device or wires attached. This has been explored and described in [1] to be superior to traditional input devices. This approach was also utilized in [2] to demonstrate human-computer interaction with gestures that are natural and relaxed.

3.3.3 Ergonomics and Usability

The standard method of using gestures involves users holding their arms up to the display for long periods of time. This has been known to cause a fatigue problem known as the "Gorilla Arm Syndrome" [9, 10, 12] and is considered to be a "known limitation" [11] of gestural interaction. A simple method of overcoming this issue was

explored in [2] by allowing users to rest their elbows on a chair armrest. In [5], the author observes the same solution, but was unable to utilize it during certain gestural interactions.

3.4 Pilot 1

3.4.1 Approach

A total of 7 (5 male, 2 female) participants were recruited, all volunteers from the Computer Science program at Baylor University. Participants ranged from 21-26 years of age with a median age of 22. Participants were technically skilled and reported to utilizing a personal computer between 28-84 hours per week, with a median of 49 hours per week. None of the subjects have had prior experience with gestural input. Personal Space interaction starts with a calibration phase where the user is required to select 4 points (top right, top left, bottom left, bottom right) while keeping her elbow rested and anchored to the table. These 4 points create a flat plane which defines her personal space (Figure 3.3). This flat plane will be mapped to the display device. The anchor point of the elbow has to remain static throughout a chosen calibration. All gestural input was captured with the Leap Motion. The de facto method currently used in applications developed for the Leap Motion involves the user moving his hands in 3 dimensions while the palm hovers over the device. We created a relative-mapped cursor manipulation program to benchmark the de facto interaction style. For Personal Space, we tilted the device 30 degrees forward. We found this tilting to reduce self-occlusion and improve palm tracking.

3.4.2 Experimental Design

A series of 7 experiments were designed: 3 to compare gestural navigation between the mouse, the de facto gestural input, and Personal Space, and 2 each to test gestural selection and manipulation between the mouse and Personal Space. The term "navigation" in [12] is broad, and includes 3D navigation, in our experiments

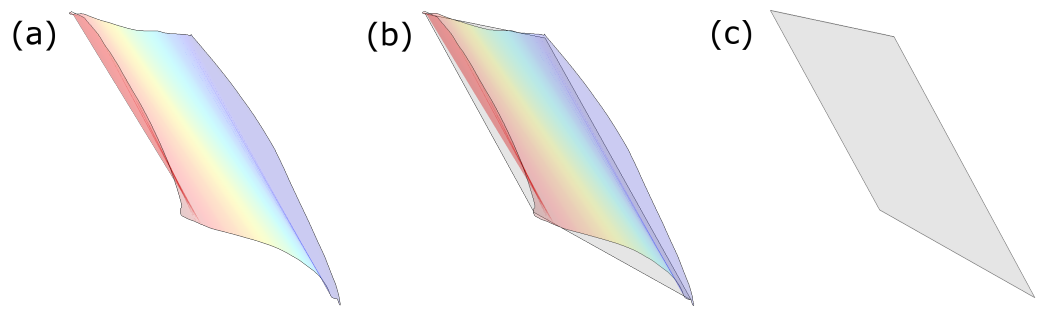

Figure 3.3. (a) Diagonal view of the arms natural range. (b) Building a flat plane for Personal Space based on the natural range. (c) Actual flat plane built for Personal Space.

we only use 2 dimensions, and will therefore use the term gestural pointing as per [3] to be precise. To assess the performance of the 3 input devices, the total time for each trial was recorded. This is described in [15] as a naive approach compared to throughput, however using Index of Difficulty would be redundant since target size and distance were both held constant.

3.4.2.1 Gestural pointing. To assess gestural pointing, a task was designed to benchmark performance and usability of the three interaction styles. The task in the experiment was a game similar to Whac-A-Mole; targets appear at seemingly random locations on the screen, user must hover on the target for one second before they would disappear and another would appear. Each user performs 63 trials per interaction style. The sequence of each trial is the same throughout the experiment, but designed to appear random to the users.

3.4.2.2 Selection. Selection tasks were done with a left click on the mouse and a finger tap on Personal Space.

3.4.2.3 Manipulation. Manipulation tasks were done with drag-and-drop on a mouse. With gestural interactions, the users were allowed to use any one of 3 different gestures to select: grasp, grab, and pinch.

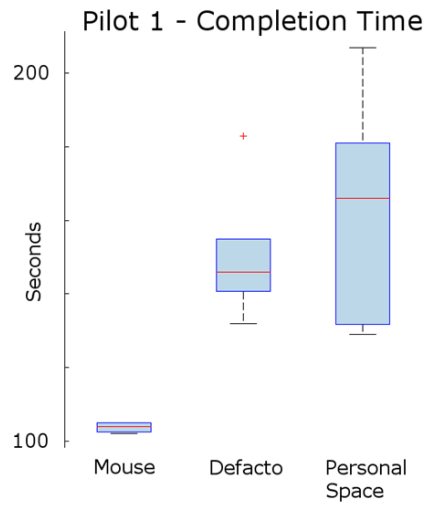

Figure 3.4. Distribution of completion time for Pilot 1.

3.4.3 Quantitative Results

Our test subjects required 7-28 minutes of calibration, with a median of 17 minutes. The distribution for the gestural pointing tasks across all 7 subjects is given in Figure 4. The mean completion times are mouse: 103s, de facto: 150s, and Personal Space: 161s. Our statistical analysis shows a significant difference between the mouse and both methods of gestural input, $F(2, 18) = 16.91, p < 0.01$, but none between the de facto method and Personal Space. The performance measures of selection and manipulation tasks were both significantly slower than the mouse. It will therefore not be given an in-depth analysis, nor will it be explored further in our research.

3.4.4 Qualitative Results

Subjects were encouraged to alert researchers to any issues during the experiment, particularly subjects mentioning fatigue. Of the 7 subjects, 4 mentioned that the de facto input method was causing fatigue during the experiment. The remaining 3 mentioned it after the task or at the end of the experiment. One mentioned fatigue with Personal Space after the experiment. Gestural selection received bad

reviews, with most subjects reporting pain in their joints. We concluded that the gesture proposed was a bad choice despite being conceptually similar to the actions of a left-click. Gestural manipulation was a difficult task for most users; the cursor tends to move during the selection process and self-occlusion causes the palm to be not accurately detected at certain points of the screen.

3.5 Pilot 2

The performance of gestures in the first pilot study led us to focus our research on gestural navigation only. The palm-flip gesture used to stop tracking was generally usable, but in the event of a false positive, completion time was delayed 5 seconds per false positive. We therefore designed pilot 2, which focused solely on navigation and disabling gestural input completely. Additionally, we captured another metric during this pilot: cursor distance travelled.

3.5.1 Approach

A total of 5 (4 male, 1 female) participants were recruited, all volunteers from the Computer Science program at Baylor University. Participants ranged from 24-28 years of age with a median age of 24. Participants were technically skilled and reported to utilizing a personal computer between 40-65 hours per week, with a median of 40 hours per week. None of the subjects have had prior experience with gestural input. None of the subjects involved with the first pilot study were recruited for Pilot 2.

3.5.2 Experimental Design

The same gestural navigation tasks from (3.2.1) were performed by the subjects. Tasks from (3.2.2) and (3.2.3) were not performed in this pilot study.

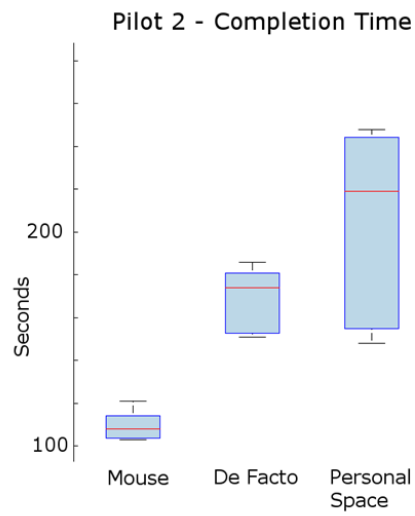

Figure 3.5. Distribution of completion time for Pilot 2.

3.5.3 Quantitative Results

As in pilot 1, the statistical significance between completion time using the mouse and gestural was significant, $F(2, 12) = 13.09, p < 0.01$, while the de facto gestural input method was not statistically different to Personal Space (figure 5). The means for completion times are mouse: 109s, de facto: 168s, Personal Space 203s. In terms of total distance travelled by the cursor, as shown in Figure 6, we surprisingly found no statistically significant difference between all 3 input devices. The means for distance travelled ($\times$100 pixels) are mouse: 96, de facto: 119 and Personal Space: 111.

3.5.4 Qualitative Results

We observed that four of the subjects started the de facto experiment with their elbows in a resting position and expected to be able to keep it rested throughout the experiment. Three subjects mentioned fatigue during the de facto method within 90 seconds and one subject asked to switch arms halfway through (this was allowed). Conversely, no subject mentioned fatigue while using Personal Space. However, one did mention discomfort in fingers and palm after the experiment.

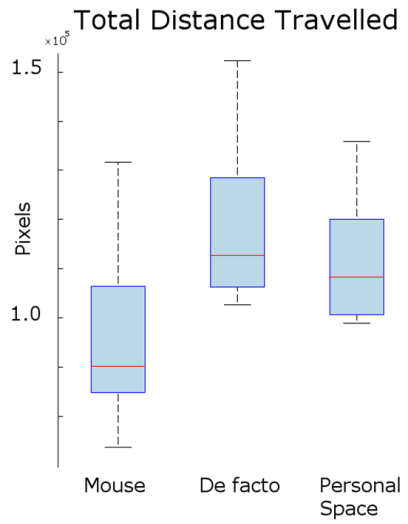

Figure 3.6. Total distance travelled by the cursor during Pilot 2.

3.6 Discussion and Future Works

The main contribution of our research so far is an interaction style that very significantly reduces the Gorilla Arm Syndrome, without sacrificing performance. This is despite the non-linear mapping used to map the users palm position and the cursor. While this method was done with a free-hands technique, we propose that this same technique can be used for any other gestural input device, including data gloves. The next step of our research will be to implement the users space in a curvilinear model instead of a flat plane. We believe this will map the natural angle of the arm much better based on Figures 3 and 4. Our next experiments will include an index of difficulty. This will include truly random target positions, and different target sizes. Performance will be calculated in throughput and evaluation will also be done with the ISO 9241-400:2007 specification. The statistical insignificance in travel distance leads us to believe there is a comparatively high accuracy between the Personal Space approach and the mouse. We intend to explore this hypothesis in the near future.

3.7 Conclusion

In this paper, we have introduced Personal Space, an interaction method for gestural pointing which allows users to define their own space captured through a calibration stage. This space uses a non-linear mapping between palm position and cursor position. This method has shown to be equal in performance to the de facto gestural method, but far superior in usability.

3.8 References

[1] Von Hardenberg, Christian, and Franois Brard. "Bare-hand human-computer interaction." Proceedings of the 2001 workshop on Perceptive user interfaces. ACM, 2001

[2] Freeman, Dustin, Ramadevi Vennelakanti, and Sriganesh Madhvanath. "Freehand pose-based Gestural Interaction: Studies and implications for interface design." Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on. IEEE, 2012.

[3] Vogel, Daniel, and Ravin Balakrishnan. "Distant freehand pointing and clicking on very large, high resolution displays." Proceedings of the 18th annual ACM symposium on User interface software and technology. ACM, 2005.

[4] Bolt, Richard A. "Put-that-there: Voice and gesture at the graphics interface". Vol. 14. No. 3. ACM, 1980.

[5] Segen, Jakub, and Senthil Kumar. "Look ma, no mouse!." Communications of the ACM 43.7 (2000): 102-109.

[6] Lai, Kam, Janusz Konrad, and Prakash Ishwar. "A gesture-driven computer interface using Kinect." Image Analysis and Interpretation (SSIAI), 2012 IEEE Southwest Symposium on. IEEE, 2012.

[7] Biswas, K. K., and Saurav Kumar Basu. "Gesture Recognition using Microsoft Kinect." Automation, Robotics and Applications (ICARA), 2011 5th International Conference on. IEEE, 2011.

[8] Triesch, Jochen, and Christoph Von Der Malsburg. "Robotic gesture recognition by cue combination." Informatik98. Springer Berlin Heidelberg, 1998. 223-232.

[9] Yoo, Juwan, Seungyup Lee, and Chieteuk Ahn. "Air Hook: Data preloading user interface." ICT Convergence (ICTC), 2012 International Conference on. IEEE, 2012.

[10] Carmody, T. "Why 'Gorilla Arm Syndrome' Rules Out Multitouch Notebook Displays." (2011). [online] http://www.wired.com/gadgetlab/2010/10/gorilla-arm-multitouch/

[11] Teixeira, Vtor. "Improving elderly access to audiovisual and social media, using a multimodal human-computer interface". Diss. Faculdade de Engenharia, Universidade do Porto, 2011.

[12] Wachs, Juan Pablo, et al. "Vision-based hand-gesture applications." Communications of the ACM 54.2 (2011): 60-71.

[13] Wang, Robert Y., and Jovan Popovi. "Real-time hand-tracking with a color glove." ACM Transactions on Graphics (TOG). Vol. 28. No. 3. ACM, 2009.

[14] Jacob, Robert JK, et al. "Reality-based interaction: a framework for post-WIMP interfaces." Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2008.

[15] Zhai, Shumin. "Characterizing computer input with Fitts law parametersthe information and non-information aspects of pointing." International Journal of Human-Computer Studies 61.6 (2004): 791-809.

[16] G.M. Poor, Brianna J. Tomlinson, Darren Guinness, Samuel D. Jaffee, Laura M. Leventhal, Guy Zimmerman, Dale S. Klopfer. "Tangible or Gestural: Comparing Tangible vs. Kinect^{TM} Interactions with an Object Manipulation Task." 7th International Conference on Tangible, Embedded and Embodied Interaction, 2013.

CHAPTER FOUR

An Evaluation of Touchless Hand Gestural Interaction for Pointing Tasks with Preferred and Non-preferred Hands

4.1 Abstract

Performance evaluation of touchless gestural interaction are generally done by benchmarking pointing performance against existing interactive devices, requiring the use of user's preferred hand. However, as there is no reason for this interaction to be limited to only one hand, evaluation should rightfully consider both hands. In this paper we evaluate the performance of touchless gestural interaction for pointer manipulation with both the preferred and non-preferred hands. This interaction is benchmarked against the mouse and the touchpad with a multidirectional task. We compared the performance between all devices, improvement in performance between 2 rounds, and the degradation of performance between hands. The results show the mouse has no performance increase between rounds but high degradation across hands, the touchpad has medium performance increase and medium degradation, and gestural interaction has the highest performance increase and the lowest degradation between hands.

4.2 Introduction

Touchless gestural interaction tends to be an effective, intuitive, and natural interaction for users to relay information to the computer (Wachs, Kölsch, Stern, and Edan 2011). Currently, commodity hardware such as the Leap Motion and Xbox Kinect is readily available to the general public. However, these devices are targeted towards non-work, non-purposeful interactions often used in isolation (such as gaming). This limited functionality is the exact opposite of the multi-functional concept of desktop computers (Bødker 2006). While touchless gestural interaction can definitely be fun (Norman and Nielsen 2010), there is no reason it cannot also be

functional. It should be possible to incorporate gestures into everyday devices used for work-related, purposeful interactions, such as pointer manipulation, 2-handed desktop interaction, or as a secondary input device.

The multi-purpose use of gestures should result in a highly usable mode of interaction, as it employs well-researched principles of Reality-Based Interaction. It provides a natural and intuitive method of interaction, by building on users' pre-existing knowledge of the everyday, non-digital world (Jacob, Girouard, Hirshfield, Horn, Shaer, Solovey, and Zigelbaum 2008). This gives us a theoretical foundation to believe that touchless gestural interaction should be a highly usable mode of interaction. However, little is known on how effective it would be in the real world, it's strengths and limitations, nor how it would compare relative to existing devices. Therefore we sought to develop a better understanding of the capabilities and potential of this interaction by performing an evaluation of pointer manipulation tasks on both the preferred and non-preferred hands.

There is no reason to limit touchless gestural interaction to using only one hand. Instead, it could potentially be used in 2-handed interaction, such as bi-modal interaction, as an optional secondary input device, or to allow users to alternate between hands based on preference or comfort levels. Our experiment was designed to benchmark gestural interaction using both hands against the mouse and the touchpad. We examined each device's improvement between rounds, and the difference in performance between hands – referred to as degradation.

Our results show lower performance with gestures compared to the other devices, which is generally the case with new devices. However, gestures had the highest performance improvements between 2 rounds, and the lowest degradation between hands. We believe the lower degradation will lead to better use of the device on the non-preferred hand, thus paving way for the 2-handed interactions mentioned above.

We also observed interesting findings in the other devices used in our experiment. For example, the mouse showed no significant performance increase on neither the dominant nor the non-dominant hand, leading us to believe that knowledge of the device is transferred between hands.

4.3 Related Works

4.3.1 Pointing Device Evaluation

In 1954 Paul Fitts established a relationship between movement speed and accuracy in rapid motor movements (Fitts 1954). This relationship paved way to standardized pointing device evaluations used today. Fitts introduced what he initially termed Index of Performance, which has since been referred to as throughput and calculated as the index of difficulty (ID) over movement time (MT) (Zhai 2004),

$$\text{Throughput} = \frac{ID}{MT}$$

where ID is the ratio of the distance(D) to width(W) and measured in bits:

$$ID = log_2\left(\frac{2D}{W}\right)$$

This definition has been recently updated (Soukoreff and MacKenzie 2004) to use the Shannon formulation, so:

$$ID = log_2\left(\frac{D}{W} + 1\right)$$

The original definition of distance(D) in the formula above measures the center point between two targets. This formula has been adjusted to use the distance between the starting point of the cursor to the ending point, known as effective distance (D_e). Width has also been updated to use effective width (W_e), which is 4.133σ where σ is the standard deviation of the actual endpoint distribution. Current methods of calculating throughput therefore uses an effective index of difficulty (ID_e) (Zhai 2004, Soukoreff and MacKenzie 2004) defined as:

$$ID_e = log_2\left(\frac{D_e}{W_e} + 1\right)$$

This adjusted definition of effective index of difficulty is used in our throughput calculation. Therefore, we calculate throughput as

$$\text{Throughput} = \frac{ID_e}{MT}.$$

4.3.2 Handedness

Handedness refers to the performance in the preferred and non-preferred hands; there are several reasons to study evaluate the non-preferred hand in pointing tasks. Studies have shown that with increased use of the mouse, we are exposed to higher risk of hand impairments such as carpal tunnel (Keir, Bach, and Rempel 1999). 2-handed gestural interaction allows a potential solution by allowing the user to use a more natural mode of interaction such as gestures, and by allowing users to alternate between hands. A smaller difference in performance between hands could better encourage this behaviour. Additionally multi-handed gestural interaction has demonstrated usefulness in areas such as robot control (Wu 2000) and neurosurgery (Goble, Hinckley, Pausch, Snell, and Kassell 1995).

One of the first studies we found to examine handedness in human-computer interaction(HCI) (Kabbash, MacKenzie, and Buxton 1993) found that there were differences in performance between hands, which they referred to as degradation (Kabbash, MacKenzie, and Buxton 1993), defined as:

$$\text{Degradation} = \frac{\textsc{Time}_{\text{NPH}} - \textsc{Time}_{\text{PH}}}{\textsc{Time}_{\text{NPH}}}$$

In their definition of degradation, Time_{NPH} refers to the mean movement time of the non-preferred hand measured in milliseconds, while Time_{PH} refers to that of the preferred hand. They found that the mouse and stylus degraded at about the same

rate of 28% while the trackball did not degrade between hands. They also reported that for larger targets and larger distances, similar performance was observed between hands.

4.3.3 Gestural Pointing

(Pino, Tzemis, Ioannou, and Kouroupetroglou 2013) looked into the use of the Xbox Kinect for point-select tasks. The experiment utilized a 2D and 3D multi-directional task to evaluate the pointing device. The 2D task used 5 blocks of 15 selection targets resulting in 75 trials per task. The experiment reported 2.10 bits/s for the 2D tasks, and 1.06 for the 3D tasks. The Kinect's throughput was 39% lower for the 2D task, but outperformed the mouse in the 3D task by 9.7%. The study also found that Fitts' law extended into gestures similarly to the mouse in the 2D task and at a higher rate in the 3D task.

(Sambrooks and Wilkinson 2013) looked into comparing touch, gestures, and mouse interactions. This work focused on the performance of touch versus gestures, the relationship of gesture performances over time, and if gestural performance suffered from fatigue. The study used a task with 100 targets further broken into rounds of 20 targets with 10 second breaks between rounds. Selection was done with a left-click for the mouse and single-tap for touch. For gestures, the study used an animated 'selection circle' when hovering over a selection element. The study found that fatigue was not a factor in gestures and that there was no improvement in gestures over the course of the experiment. We designed our study to take a deeper look into this claim, as we found it to be counter-intuitive. New devices generally do exhibit performance improvements (MacKenzie, Kauppinen, and Silfverberg 2001) , and gestures being a more natural mode of interaction should too.

(Jude, Poor, and Guinness 2014) compared the mouse and two gestural interaction techniques. The study identified that the "Gorilla Arm Syndrome," a known

fatigue problem of touch screen interaction, extended to gestural interaction. The authors proposed a solution to this problem by allowing the user to choose the corners of the screen from a resting position – the elbow is rested on a surface – and used user-defined points to map to the screen coordinates. The experiment noted that in the standard gestural pointing technique, fatigue was mentioned within 90 seconds of use, while no fatigue was reported with their technique. The study also reported that there were no significant differences in terms of movement time between both methods of gestural interaction. Our interaction style was based on the approach introduced here.

4.3.4 Learning Effects

An evaluation of the mouse, trackball, joystick, and touchpad using a 2D multi-directional point-select task was performed in (MacKenzie, Kauppinen, and Silfverberg 2001). The experiment took the learning effects of the interface into consideration, and only performed analysis on the data where there was no learning present. In order to achieve this, each device was tested over 10 blocks each with 5 sequences of 15 target selections. Each device took about an hour to evaluate and each participant was evaluated with all devices. They found that learning effects were not significant after 5 blocks. In our experiment, we measure and compare the learning effects described here.

4.3.5 Motor Learning

As we were evaluating a novel interaction across both hands, we could not allow the user to perform the task until no performance improvement was observed. Therefore, we decided to investigate the learning effects of the device. Motor Learning, as defined by (Schmidt and Lee 1988), is described by four distinct characteristics: (1) Learning is a process of acquiring the capability for producing skilled actions. (2) Learning occurs as a direct result of practice or experience. (3) Learning cannot be

observed directly, as the processes leading to changes in behavior are internal, and are usually not available for direct examination. (4) Learning is assumed to produce relatively permanent changes in the capability for skilled behavior. (Schmidt and Lee 1988) indicated that while we cannot directly observe learning, we can measure and report performance improvements. From this we can infer learning. In this study we measured and reported performance improvements between rounds, with the commonly used measure of effect size (Lin, Sullivan, Wu, Kantak, and Winstein 2007, Sullivan, Knowlton, and Dobkin 2002, Seidler 2004).

(Lin, Sullivan, Wu, Kantak, and Winstein 2007) indicated that random-order practice (e.g. A-B-C, B-C-A, C-A-B) generally benefits motor learning more than block order practice (A-A-A, B-B-B, C-C-C). As we were evaluating performance improvement, it was in our best interest to optimize motor learning. We therefore used this method in designing our experiment, where each subject used 2 devices on both hands per block, and each block performed twice.

4.4 Methodology

4.4.1 Task

The task performed by the subjects consisted of 70 trials. In each trial, a square target appears at locations designed to look random. Subjects were required to hover on the target for 500 milliseconds, which caused the target to disappear and a new target to appear. This action is repeated until the task ends.

Four task profiles were created, each with 70 trials. These trials were specifically designed to incorporate different direction and distance between trials. All 4 task profiles had 15 small, 20 medium, 20 large and 15 extra-large targets, which were respectively 90, 120, 160, and 220 pixels in both height and width. All 4 task profiles had a similar ID rating, ranging from 200 to 210 bits.

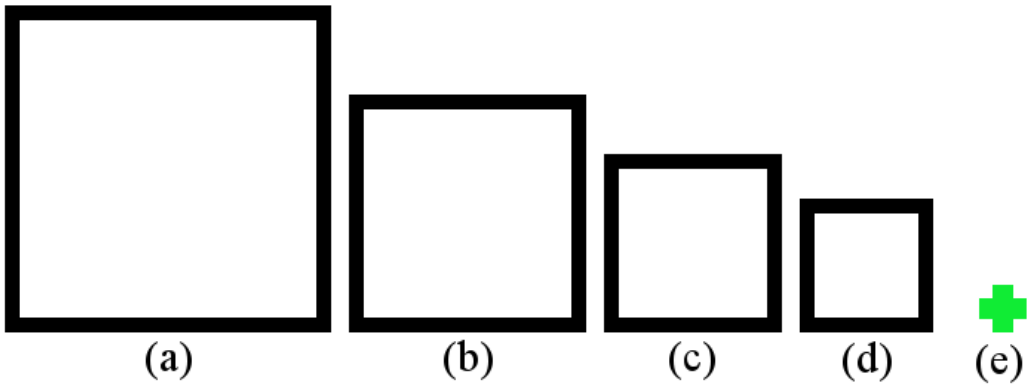

Figure 4.1. A scaled version of the targets used. (a) 220x220 pixels, (b) 160x160 pixels, (c) 120x120 pixels, (d) 90x90 pixels. The cursor shown in (e) was set to a high contrast green shown 32x32 pixels.

We sought to make the experiment fun and to encourage speed. We therefore attempted to gamify the task by asking the users to complete as quickly as possible, and included a status cell that denoted score and time per task.

During our pilot study, we identified the need for high contrast between the cursor, targets, and the background. We therefore used a black background, with white targets, and a green cursor which posed high contrast to both the background and the targets. The cursor was a 32x32 pixel image, the largest cursor size allowed by Microsoft Windows 7. These changes helped subjects identify their cursor easily on the screen, thus increasing the speed at which they could complete a trial.

4.4.2 Breaking the Standard Evaluation Method

The ISO 9241-9 standard has outlined an effective evaluation method for pointing devices. The 2D multi-directional task states the following are necessary (Soukoreff and MacKenzie 2004): (1) Circular or square targets may be used, (2) The effect of direction should be controlled, (3) The path should begin and end at the same target, and (4) The software must graphically indicate which target the subject should proceed to next.

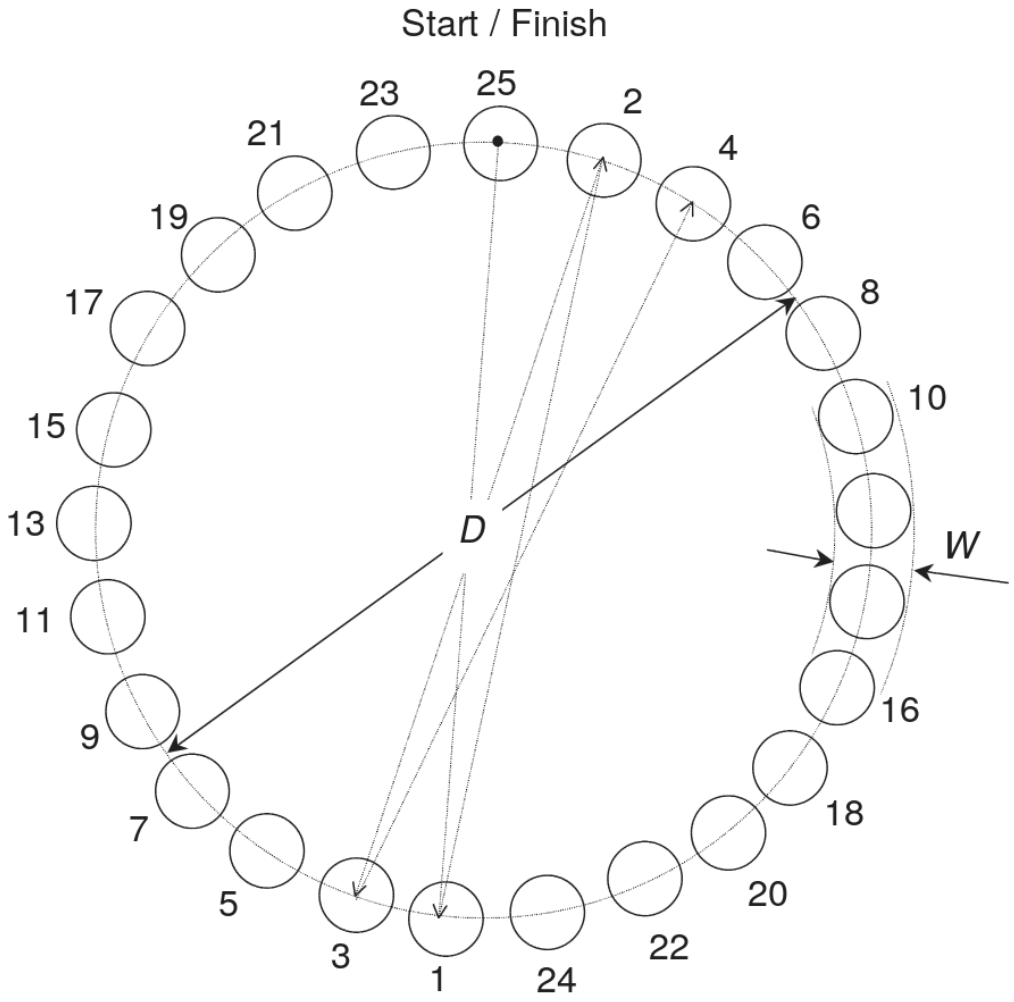

Figure 4.2. Multi-Directional task implementation from (Soukoreff and MacKenzie 2004) used with permission

This standard has been widely accepted by the HCI community and is highly recommended for pointing device evaluation (Soukoreff and MacKenzie 2004). However, we had to break away from this implementation as our pilot studies showed that the corners of the screen were the most difficult for users to reach with gestural interaction. Using the standard method would not allow us to assess this behavior, or to factor this behavior into the experiment, which would mean the results would be flawed, or based on a highly optimistic measure.

Other non-standard features were incorporated in our system. A hover action, similar to that used in another gestural study (Sambrooks and Wilkinson 2013) was used instead of a select action. This was done as we did not want to introduce any nuisance variable by implementing gesture-based selection. The targets were made to appear at random locations, with no graphical indication where the next would be. This controlled for muscle memory and promoted stronger learning (Grafton, Hazeltine, and Ivry 2002). Finally, the same location was not chosen for both the beginning and end. This allowed us to reach more areas of the screen with the same amount of trials. We did however take 2 of the suggestions into consideration: we controlled for the effect of direction, and created trials with ID between the recommended range (1-4.5 bits).

4.4.3 Input Devices and Interaction

The devices used in our experiment were selected for their equally usable nature on both hands without being biased to either. We used the a symetrical mouse (Logitech M-U0032-O), a standalone external touchpad (PERIPAD-702), and the Leap Motion controller for gestures. Participants were encouraged to place the peripherals in a position that was most comfortable for them. Researchers helped adjust the angle and position of the Leap Motion controller to increase recognition rates.

The standard method of interaction for gestural input involves the user holding their arms out, which is known to cause fatigue (Teixeira 2011, Wachs, Kölsch, Stern, and Edan 2011), commonly known as the "Gorilla Arm Syndrome" (Carmody 2010). We addressed this issue by using the Personal Space approach (Jude, Poor, and Guinness 2014) as it was demonstrated to heavily reduce fatigue without sacrificing performance, by allowing users to rest their elbow on a surface. The users were required to define their own interaction space in 3 dimensions through a calibration stage as shown in figure 4.3.

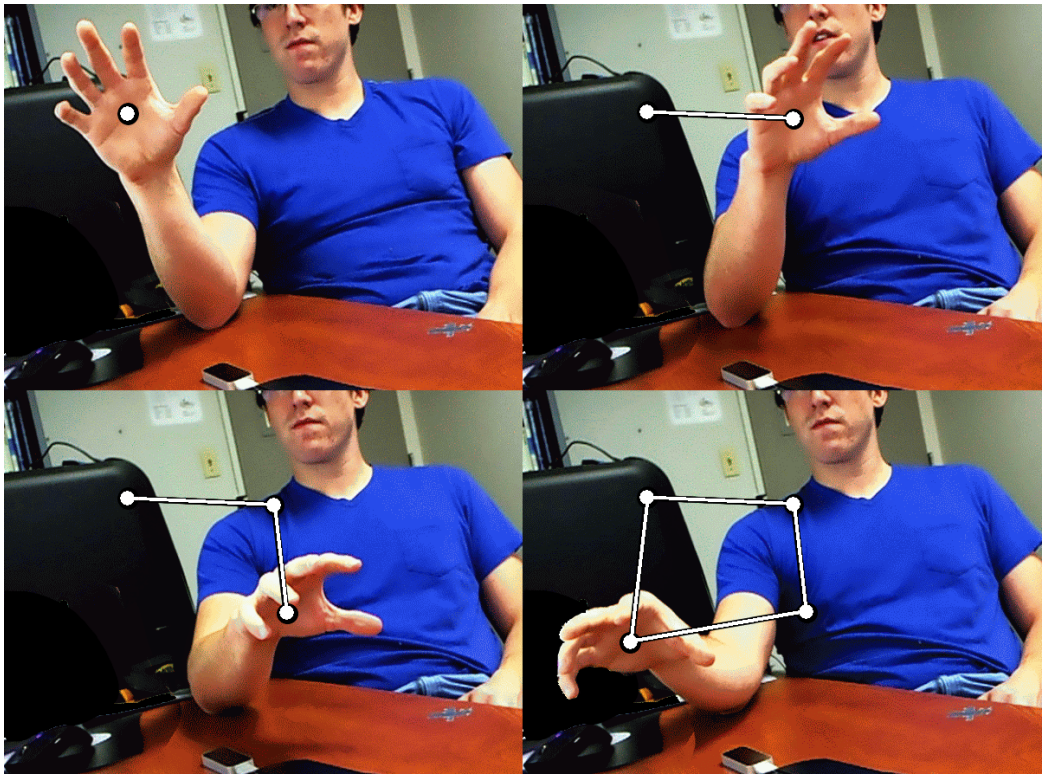

Figure 4.3. Personal Space calibration:Constructing the user's space

In this experiment, the software prompted subjects to position their hand where they would like each screen corner to be. Most subjects anchor their screen boundaries by pointing to the corner of the screen with their hand. This created a more intuitive calibration, which allowed the subjects to easily navigate in 3 dimensional space.

After the calibration, the subjects were asked to navigate around the screen. The 4 corners were given specific attention as we found this to be the most sensitive regions for recognition. Subjects were able to find a comfortable calibration by their second attempt during round 1 of the experiment, and on their first attempt on subsequent rounds.

4.4.4 Participants

A total of 36 subjects (22 male, 14 female) were recruited to participate in the user study. Subjects' age ranged from 18-36 years with a median of 21 years. Subjects' computer use ranged from 1 to 50 hours per week with a median of 30 hours. All subjects were undergraduate or graduate students who were compensated for their participation.

Subjects were randomly assigned to one of 3 groups, where each group would perform the given tasks with 2 different devices. The first group used gestures and mouse, the second used gestures and touchpad, the third used the touchpad and the mouse. This gave us 24 samples per device.

4.4.5 Procedure

The subjects were first shown a 1-minute video detailing the hover task. Before using each device, they were shown a short video explaining the the device being used. The gestural interaction had a longer video of approximately 3 minutes, including an explanation of the calibration process.

Subjects would perform the task once on each device on each hand, totaling to 4 tasks per round. Each task is done with a different task profile, therefore ensuring that the subjects would not be able to learn the task or guess the position of each target. A practice session consisting of 12 trials was performed before the full task of 70 trials. The 4 tasks in the first round was followed by a 2 minute break, after which the subjects were asked to perform the exact same tasks as before in the same order and on the same task profile. The training task was not present in the second round.

Upon completion, subjects were given an exit survey detailing their prior use of each device and their preference, comfort and fatigue experienced with those devices during the experiment.

4.4.6 Design

The experiment used a between groups design with 3 groups. A 2x2 latin-squares design was used to counterbalance the order in which the device and hands are used in an experiment. We had the option of performing the tasks in orders across 2 rounds (A-B-C-D, A-B-C-D) or in blocks (A-A, B-B, C-C, D-D), and we chose the former as it has been shown to generally benefit motor learning (Lin, Sullivan, Wu, Kantak, and Winstein 2007).

4.5 Results

4.5.1 Preferred Hand vs Non-Preferred Hand

All subjects reported their right hand as the preferred hand for using the mouse. As a result, we refer to the right hand as the preferred hand and the left hand as the non-preferred hand in the analysis. 2 of the subjects self-reported as left-handed while one self-reported as ambidextrous.

4.5.2 Time

Two time-related metrics are commonly used in literature: completion time and movement time (Zhai 2004). Completion time is shown in table 4.1. This metric provides good indication of the duration in which each participant took to complete each task, and a simple overview of the performance of each device. It is however not suitable for analysis as it includes the hover time of 500 milliseconds per trial or 35 seconds per task. Adding a constant factor to the metrics collected causes the differences between them to become much smaller than it actually is.

Table 4.1. Means of completion time in milliseconds for each device across both rounds.

Hand	Mouse	Touchpad	Gestures
Right	78360	93715	118475
Left	98245	109732	127835

Movement time refers to the time taken for the subject to move from the starting point to the target and therefore does not include the hover time of 500 milliseconds. Table 4.2 shows the mean movement time per trial across all devices and hands. This table is presented to facilitate comparison with existing literature such as (Kabbash, MacKenzie, and Buxton 1993), but has since been shown to be a naive metric (Zhai 2004) and therefore not used for further analysis. We did however notice improvements in movement time between rounds on all devices across hands, with the exception of the mouse used with the right hand. Movement time is a rather descriptive measure that can be used for performance comparison within a particular experiment, but it is highly dependent on the experiment itself. For example a different experiment which has a different configuration of target position, size, and number will yield a different movement time. Throughput, in contrast, uses target distance and width as weights and has been demonstrated to correctly model human behavior.

4.5.3 Throughput

We measure performance by throughput, defined as:

$$\text{Throughput} = \frac{ID_e}{MT}$$

The means of the throughput is shown in table 4.3 along with the standard deviation. The standard deviation is included in this table as it will be used in the analysis of

Table 4.2. Mean movement time per trial in milliseconds.

Hand	Mean	Mouse	Touchpad	Gestures
Right	Round 1	626	855	1232
	Round 2	621	823	1153
	Overall	624	839	1193
Left	Round 1	909	1097	1379
	Round 2	898	1038	1274
	Overall	904	1068	1327

performance increase between rounds as well as performance degradation between hands. An anova test performed on the pooled throughput between rounds shows a significant difference in the means, $F(5, 282) = 240.76, p < 0.001$. A Tukey's HSD test shows statistically significant difference in means across all devices and hands. We can therefore rank the devices based on the overall means from best to worst as such: right mouse, right touchpad, left mouse, left touchpad, right gestures, left gestures.

4.5.4 Performance Improvement

We performed a t-test to check for statistical significance between the performance in the first and second rounds which revealed no statistically significant performance improvement for the mouse between rounds on the right hand $t(23) = 0.34$, $p = 0.63$ nor the left hand, $t(23) = -0.94, p = 0.18$. We did however find improvements on the touchpad for both the right hand, $t(23) = -2.78, p < 0.01$ and the left hand, $t(23) = -5.41, p < 0.01$. There was also an improvement with gestures for both the right hand, $t(23) = -2.8542, p < 0.01$, and the left $t(23) = -3.81, p < 0.01$. To check for the performance improvements between the rounds, a simple method would be to calculate the difference between the mean of means of throughput between both rounds:

Table 4.3. Means and standard deviation of throughput for each device across both the right (R) and left (L) hands.

Hand	Source	Mouse		Touchpad		Gestures	
		Mean	Stdev	Mean	Stdev	Mean	Stdev
Right	Round 1	4.81	0.42	3.58	0.34	2.56	0.33
	Round 2	4.79	0.38	3.74	0.47	2.71	0.30
	Overall	4.80	0.39	3.66	0.42	2.64	0.32
Left	Round 1	3.36	0.38	2.86	0.44	2.33	0.36
	Round 2	3.40	0.42	3.03	0.41	2.48	0.33
	Overall	3.38	0.39	2.94	0.43	2.41	0.35

$$\text{Improvement} = \frac{\textsc{Throughput}_{R2} - \textsc{Throughput}_{R1}}{\textsc{Throughput}_{R1}}$$

Improvements measured this way are listed in table 4.4. This method while simple, is not a good measure as it ignores relevant information such as standard deviation, which is important to be used in calculation to contextualize the difference (Coe 2002). We therefore decided that a better way to measure performance increase is with effect size. A standard metric used to report effect size is Cohen's d (Cohen 1992) which is defined as the difference of the mean between 2 groups over the standard deviation:

$$\text{Effect Size} = \frac{\overline{x_1} - \overline{x_2}}{s}$$

Effect Size measured as 0.2 is considered a small effect, 0.5 signifies a medium effect visible to the naked eye, and 0.8 signifies a large effect size (Coe 2002). This description of effect size is assumed to be true but could possibly only holds true in the social sciences and not in human-computer interaction.

In the original equation above, the denominator is the standard deviation of the population, which would be impossible to obtain and was therefore replaced with the pooled variance of both samples. This is a common method used to calculate effect size and is used to assess performance improvement from one point to another (Sullivan, Knowlton, and Dobkin 2002, Lin, Sullivan, Wu, Kantak, and Winstein 2007):

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Table 4.4. Performance increase measured as percentage of improvement in throughput between rounds

Hand	Mouse	Touchpad	Gestures
Right	-0.4%	4.4%	5.7%
Left	1.2%	6.1%	6.6%

The first thing to notice about our effect size listed in table 4.5 is how the results for gestures can be interpreted differently based on the measures used. The simple performance improvement measured with percentage shows that left gestures has a higher improvement than right gestures, while the metrics calculated with effect size shows that right gestures is in fact higher.

A noteworthy observation is that the effect size differs very slightly between hands across all devices. For example the performance difference for the mouse is 0.05 and 0.11 for right and left hands respectively, touchpad is 0.39 and 0.41, while gestures' performance increment is measured as 0.46 and 0.45. This shows a consistent increase in performance using a particular device, regardless of the hand used to control the device.

We noticed that the performance improvement of the touchpad was closer to the gestures than the mouse despite the touchpad being a more ubiquitous interaction than gestures.

4.5.5 Degradation

To calculate degradation, we first used the formula used in (Kabbash, MacKenzie, and Buxton 1993):

$$\text{Degradation} = \frac{\textsc{Time}_{\textsc{nph}} - \textsc{Time}_{\textsc{ph}}}{\textsc{Time}_{\textsc{ph}}}$$

where TIME refers to means of movement time whereas NPH and PH refers to Non-preferred hand and preferred hand respectively. We found that this formula gives us

Table 4.5. Performance increase as measured with Cohen's *d*. This measure generally gives a positive number in our experiment, but we noted that performance with the mouse on the right hand actually decreases.

Hand	Mouse	Touchpad	Gestures
Right	(0.05)	0.39	0.46
Left	0.11	0.41	0.45

degradation of 45% for the mouse, 27% for the touchpad, and 11% for gestures based on overall means of movement time as shown in table 4.6.

Movement time however is a naive metric (Zhai 2004), we therefore also performed the degradation calculation with overall throughput means, with a slight modification to the original formula:

$$\text{Degradation} = \frac{\textsc{Throughput}_{\text{PH}} - \textsc{Throughput}_{\text{NPH}}}{\textsc{Throughput}_{\text{PH}}}$$

The results of this calculation shown in table 4.7 provides a better metric, standardized with current literature which recommends the use of throughput over movement time for benchmarking devices (Zhai 2004, Zhang and MacKenzie 2007).

In our calculation of performance improvement, we recommended the use of effect size and elaborated why it is a good choice in calculating differences between means of groups. We therefore used the same method to calculate degradation, which is listed in table 4.8. We observed that when using Cohen's *d*, degradation reduces between rounds for the mouse and the touchpad while it increases for the gestures. This is consistent with our analysis of performance which showed higher improvement with gestures on the right hand than on the left hand, the opposite of what happened with the mouse and touchpad.

4.5.6 Fatigue and Discomfort

Participants were asked in an exit survey whether or not they felt any fatigue or discomfort during the use of the gestural interface. 8 out of the 24 participant who

Table 4.6. Degradation calculated as percentage of increase in movement time between hands

Means	Mouse	Touchpad	Gestures
Round 1	45.2	28.3	11.9
Round 2	44.6	26.1	10.5
Overall	44.9	27.3	11.2

Table 4.7. Degradation calculated as percentage of decrease in throughput between hands.

Means	Mouse	Touchpad	Gestures
Round 1	30.1	20.1	9.0
Round 2	29.0	19.0	8.5
Overall	29.6	19.7	8.7

used the gestural interface reported more fatigue or discomfort than the mouse. This was measured through a 5-choice Likert scale.

4.6 Analysis and Discussion

4.6.1 Literature Comparison

Though our task deviated slightly from the ISO 9241-9 standard for evaluating pointing devices our results were in line with previous literature using the standard. The reported throughput of 4.8 bits/s for the mouse using the preferred hand was consistent with (MacKenzie, Kauppinen, and Silfverberg 2001) using the ISO 9241-9 standard for 2D multi-directional pointing evaluation which reported 4.9 bits/s. The throughput of the touchpad 3.66 bits/sec was much higher than the 2.9 bits/s previously reported (MacKenzie, Kauppinen, and Silfverberg 2001). It is possible that this is caused by hardware and software improvements made to the touchpad over the years, the difference in the dimensions of the device, as well as it's current ubiquity. Our gestural interaction had a reported throughput of 2.64 bits/sec, slightly better than previous reported gestural throughput of 2.10 bits/sec over an ID range 1-4 bits (Pino, Tzemis, Ioannou, and Kouroupetroglou 2013). The degradation of

Table 4.8. Degradation between hands measured with Cohen's *d*.

Means	Mouse	Touchpad	Gestures
Round 1	3.63	1.83	0.67
Round 2	3.44	1.60	0.71
Overall	3.57	1.69	0.68

the mouse calculated in percentage difference over movement time was 45%, much higher than the previous reported value of 27% (Kabbash, MacKenzie, and Buxton 1993). This could be attributed to our subjects using the computer and the mouse a lot more, as well as hardware and software improvements in the mouse since the study was conducted 21 years ago.

Our research showed a performance improvement between 2 rounds which contradicts a previous study (Sambrooks and Wilkinson 2013). This is likely due to the onset of fatigue in the aforementioned study, while in our research, we mitigate this issue by allowing subjects to rest their elbow on a surface.

4.6.2 Metrics

In this paper we focused on incorporating standard measures in reporting metrics, taken from peer reviewed literature outside and within the the realm of HCI. To start, we provided degradation between hands measured in throughput as opposed to movement time, as throughput has been accepted as the standard method of reporting performance of pointing devices, while movement time has been shown to be a naive metric (Zhai 2004).

We showed that reporting performance improvement and degradation with effect size as opposed to a percentage difference of means provides a more contextualized metric because it incorporates the standard deviation in it's calculation. This is demonstrated in the calculation of performance improvement, where using the percentage method shows a higher increase of performance on the left hand whereas the effect size shows a higher performance on the right.

Cohen's d has been used to measure effect size in other areas of research including motor learning (Lin, Sullivan, Wu, Kantak, and Winstein 2007, Sullivan, Knowlton, and Dobkin 2002, Seidler 2004). The use of this metric will allow for more meaningful interpretation and allow for integration into existing literature outside of HCI.

4.6.3 Performance Improvement

When evaluating a pointing device, it has to be expected that the user will be learning the device as the experiment continues. Our research showed that this was true of gestures and the touchpad, despite the subjects provided with a test run first. One way of reducing this effect is by making subjects perform the same task repeatedly as done in (MacKenzie, Kauppinen, and Silfverberg 2001). This was not feasible in our experiment as each round required 4 tasks for the user to perform: 2 devices on both hands. Making subjects perform the task until they have fully learnt the device could potentially take longer than the 1-2 hours recommended for usability studies (Olsen Jr 2007). To address this issue, we designed our experiment to cycle between tasks, thus allowing subjects to gain long-term motor-learning benefits (Lin, Sullivan, Wu, Kantak, and Winstein 2007), and we measured these improvements using the standard metric of effect size.

We observed that performance improvement is more related to the interface than it is the hand. This tells us that interfaces themselves have inherent performance improvement rates. We do not know if this is based on the subjects knowledge of the interface, or if it related to the design of the interface itself, however we do note that subjects have better improvement on interfaces that are foreign to them (i.e. gestures) and lesser to no improvement on the devices that they are familiar with (i.e. mouse).

An interesting observation is the lack of performance improvement with the mouse on the left hand. This has 2 potential explanations 1) that the mouse is an ideal design, allowing subjects to use this device to it's full potential immediately or 2) that knowledge from the subjects have acquired from long term use of the mouse on their preferred hand has transferred to their non-preferred hand. The second hypothesis is more likely as it has been shown in other studies that motor skill learned on one hand is transferable to the other (Grafton, Hazeltine, and Ivry 2002).

The fact that gestures had a higher performance improvement on both hands compared to the other devices validates the gestural interaction style (Jude, Poor, and Guinness 2014) used here and indicates that there is potential for the use of gestures for pointing devices, despite the lower throughput recorded.

4.6.4 Degradation

Devices that exhibit high degradation will behave noticeably different on the preferred hand than on the non-preferred hand. This difference could discourage the users from using the device on their non-preferred hand as they already have an expectation in performance. Since gestures had a much lower degradation than the other devices, we expect it to perform more similarly on the non-preferred hand, thus allowing for 2-handed interactions.

We have identified two potential explanations for the degradation in different devices, 1) touchless gestural interaction is natural enough to be used with the non-preferred hand or 2) subjects' pre-existing motor skills with regards to the device caused a higher degradation on interactions they were more familiar with. A concrete conclusion is hard to make as there is little known on the degradation with other interfaces used in HCI.

Performance degradation between hands is a useful metric to report when assessing pointing devices because it allows researchers to identify interfaces, interactions or tasks that may be suitable for the non-preferred hand. We report degradation in effect size as it allows for a much more meaningful interpretation; simply reporting the difference in percentage ignores the standard deviation and could result in different results, as illustrated by the difference between tables 4.7 and 4.8.

4.6.5 Gestures

Our gestural interaction design was based on the Personal Space approach (Jude, Poor, and Guinness 2014). This study showed that allowing the users to rest their

elbow on a surface would severely reduce fatigue while maintaining performance to a similar level as in a standard gestural interaction approach where subjects held their hand out. It was reported that the standard approach resulted in fatigue after approximately 90 seconds whereas none was reported in the Personal Space approach.

A different study in assessing gestural interaction showed no improvement in performance, a direct contradiction of our findings. This study also report that fatigue did not have any effect, despite the use of the standard approach to gestural interaction. There are a few explanations why these researchers did not observe any performance improvements in their experiments: 1) The interaction style used in our research which allows the subjects to rest their elbow allows for a better motor skills improvement 2) The performance increase facilitated by increase in motor skills is negated by fatigue, leading to zero net gain. We believe the second explanation is more likely given that it has been demonstrated that the standard interaction style would cause fatigue. The experimental design in (Sambrooks and Wilkinson 2013) broke 100 targets into 5 blocks of 20 target selection sequences with mandated 10 second breaks, which may have lead to no fatigue being reported, despite actually being present.

4.7 Conclusion

In this paper we examined gestural interaction for throughput, performance improvement and degradation, benchmarked against the mouse and touchpad. Although our study shows that gestures performs lower than the mouse and the touchpad, it does have much higher performance improvement between 2 rounds and much lower degradation between hands. The high performance improvement indicates that gestural interaction has a good potential for use in productive use such as on the desktop, as subjects will learn the interaction over time. While low degradation indicates that it is possible to use gestures in 2-handed interactions or to allow users to easily alternate between hands when used as a pointing device.

CHAPTER FIVE

Evaluating Touchless Hand Gestures and Speech for Point and Select Tasks

5.1 Abstract

Devices such as the mouse and keyboard are ubiquitous and present familiar interactions, but are far from natural. These existing interaction techniques are in fact believed to be the bottleneck to utilizing progress made in computing and communication technologies (Pavlovic, Sharma, and Huang 1997). Natural interactions such as speech and gestures have both achieved mainstream success independently, with consumer products such the Leap Motion controller and Microsoft Kinect popularizing gestures, while mobile phones have recently used speech as input. However, the idea of integrating these two modalities into one coherent interaction, has met little success despite significant interest in the 1990's. In this paper we designed an interaction style that combines both gestures and speech, in order to evaluate whether this technique can be used as a mouse replacement technology. Our results indicate that while gestures are slower than the mouse, the introduction of speech allows for selection to be performed without negatively impacting performance. We also found that users can adapt to this technology quickly and are able to improve their performance with minimal training.

5.2 Introduction

Touchless gestural interaction is an effective, intuitive, and natural interaction for users to relay information to the computer (Wachs, Kölsch, Stern, and Edan 2011). This makes it extremely well-suited for direct manipulation, such as cursor navigation. Commodity hardware such as Xbox Kinect and Leap Motion have managed to take gestural interaction out of research labs and Hollywood movies and into the hands of users. Going back in history, however, we note that the first gestural interaction

system – Bolt's "Put-That-There" introduced in 1980 (Bolt 1980) – was not purely gestural. Bolt's system integrated speech-based input for performing manipulation tasks.

Interest in gesture-speech interaction has recently dropped despite significant attention in the 1990's. This change was attributed to natural gesturing been deemed too hard for automatic processing systems. Additionally technical implementations that are over-simplified tend to end up being rigid and brittle (Kopp 2013). As a result, this multimodal interaction has been widely ignored, with little consideration on how it can actually be used in practice.

The aforementioned limitations are more technical, but as human-computer interaction (HCI) experts, we naturally also consider the strengths and limitations of this interaction from a human perspective. Ben Shneiderman (Shneiderman 2000) wrote that speech-based input has had limited success in HCI as it is more difficult for people to think while speaking, than it is for people to think while performing physical activities. People therefore find it easier to think ahead while using a mouse or a keyboard than when using speech. This causes higher error rates when using speech as input, leading to higher resistance from users. A key in allowing speech-based interaction would be to have realistic goals, and better models for multitasking.

In this paper, we addressed the issues above by designing an interaction method with realistic goals, a suitable multitasking model, and gestures that can be easily and accurately recognized for automatic processing. Specifically, we designed our system to explore the possibility of replacing the mouse, with hand gestures for navigation and speech for selection. With this interaction, the user could perform either speech-only, or gestures-only at any given time without the explicit need to perform both, therefore resulting in a better multitasking model. We hypothesize that this allows a more realistic goal, and would allow speech and gesture multimodal interaction to be better accepted and used in practice.

5.3 Related Works

In an experiment which used simulation to recognize speech and gestures as input, Hauptmann and McAvinney reported that participants strongly preferred a combination of both inputs to either modality alone for graphics manipulation (Hauptmann and McAvinney 1993). The same results were reported by Mark Billinghurst (Billinghurst 1998). This formed the basis for our investigation of multimodal interaction using gesture and speech input for everyday tasks. In our case, we chose to focus our research to mouse replacement technologies.

Sharon Oviatt in her speech and gesture study described the problem of hyperarticulation (Oviatt 1996). This refers to a stylized and clarified form of pronunciation following a recognition failure. It tends to cause a cycle of recognition errors due to the distance from the original training data. This problem is usually observed in commands that have more than one word.

Hauptmann and McAvinney reported that the speech processor would only need to understand a small number of total words, since most commands only require an utterance of 3 words (Hauptmann and McAvinney 1993). We designed our commands to have fixed 2-word utterance in order to observe this phenomenon and it's effect in the event of a speech recognition failure.

Segen and Kumar found that one of the biggest issues with gestures was that it tended to cause fatigue after prolonged interaction (Segen and Kumar 2000). Through a simulation, they found that a potential solution was to allow users to rest their elbow on a surface. The Personal Space approach (Jude, Poor, and Guinness 2014) demonstrated reduction of fatigue with a similar approach. We therefore chose to use this technique in our experiment.

5.4 Methods

5.4.1 Participants

A total of 7 participants (M=3, F=4) were recruited from a local university to participate in a within-subjects experiment. Participants were between 19 to 23 years of age with a median of 20 years. None had prior experience with gesture-based pointing devices. Participants reported a weekly computer usage between 20 to 50 hours with a median of 35 hours. All participants were compensated for their participation.

5.4.2 Task

The experiment consisted of 2 types of tasks: (1) navigation only task, and (2) navigation and speech multimodal task. In the latter, speech-based input is used to issue commands while mouse or gestures are used for navigation. Each task consisted of 70 trials, each task was performed with both the mouse and gesture-based input. This resulted in every participant performing 4 tasks per experiment. In each trial, a square target appeared at locations designed to look random. Participants were required to navigate to the target and perform a particular action that caused the target to disappear and a new target to appear. In the navigation-only task, a hover action was used, where users hover on the target for 500 milliseconds. In the multimodal task, participants would issue a speech command of "left click" or "right click" or perform a hover depending on the target. Multimodal tasks contained 40 hover targets and 30 selection targets. The targets shown in figure 5.1 were were labeled with different colours as well as with the initial of the action: "L" for left-click, "R" for right click while the hover target had no label.

A 1-minute training video was shown to the participants at the beginning of the experiment to explain the details of the task. In addition, participants were

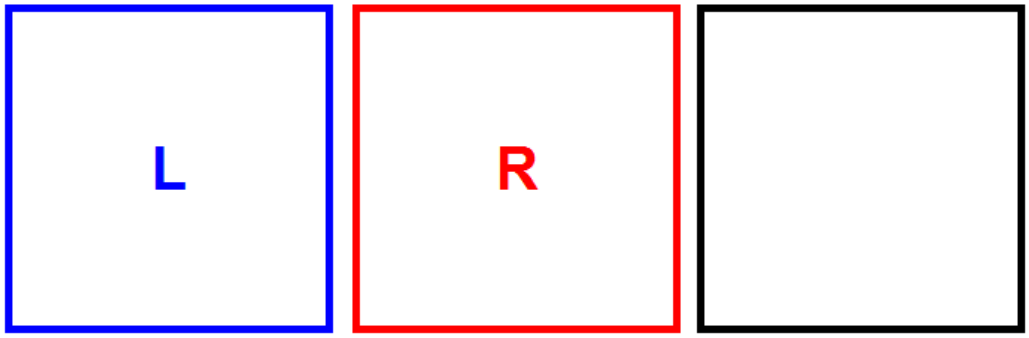

Figure 5.1. The targets expect 1 of 3 possible actions. From left to right: left-click, right-click, hover.

allowed a trial run before the actual task. The results of this trial run was excluded from the analysis.

Two possible errors can be observed in these tasks: misclicks and wrong clicks. Misclicks occurred when a left-click or a right-click happened outside of the target, wrong clicks occurred when a right-click happened when the target was expecting a left-click or vice-versa. These errors were displayed to the user in addition to the score and time in order to make the task seem more like a game. Participants were also encouraged to complete tasks as quickly as possible, which added to the gamification of the task.

5.4.3 Input Methods

In tasks that use a mouse, a regular off-the-shelf mouse (LogitechM-U0032-O) was employed. Gesture-based navigation was performed using the Personal Space approach (Jude, Poor, and Guinness 2014), as it has been demonstrated to reduce fatigue, a common problem in gestural interaction. This approach does so by allowing users to rest their elbow on a surface and defining their own interaction space through a calibration step, as shown in figure 5.3. This step creates a quadrilateral flat plane in 3-dimensional space which is then affine-mapped to the display screen. A 2-minute training video detailing the use of gestures and the calibration was shown

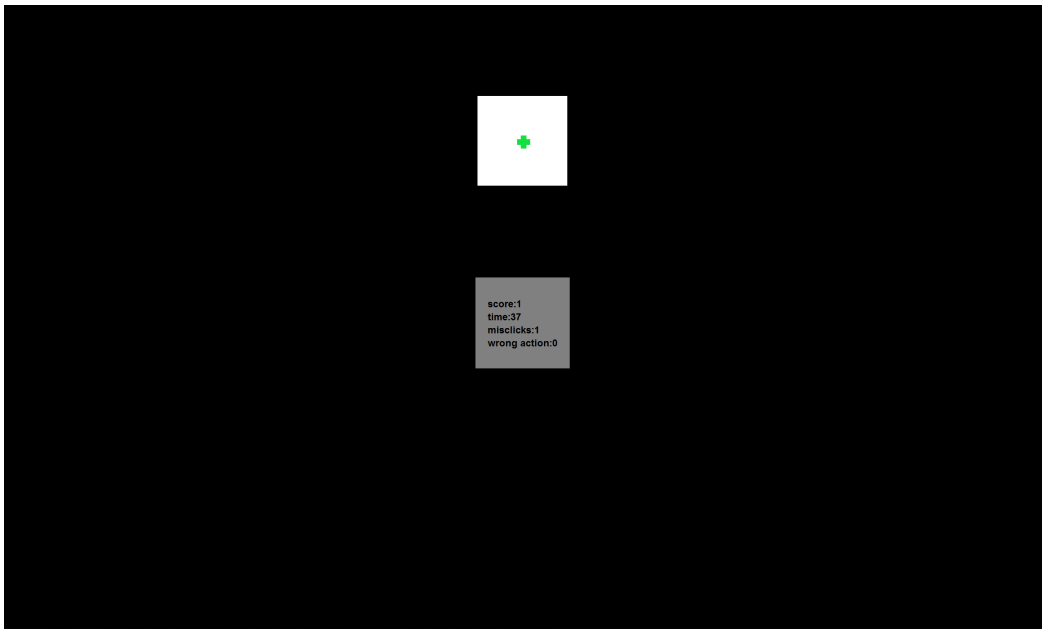

Figure 5.2. Target shown here expects a hover action. 70 targets are shown per task. The gray box at the center of the screen shows the users' score per task.

to the participants. Speech recognition was done with commercial software called "e-Speaking" version 4.1.1 (E-Speaking LLC) and a pre-trained recognition model.

5.4.4 Performance Evaluation

The software used to perform the task also collected metrics used for evaluation of the pointing technique. These metrics were collected during the task and evaluated post-experiment with MATLAB.

5.5 Results and Analysis

Analysis was performed on the data collected during the experiment. Our primary goal was to examine if introducing speech reduces performance in navigation. We also looked for error rates caused by failure in gesture and speech recognition, as well as errors made by users.

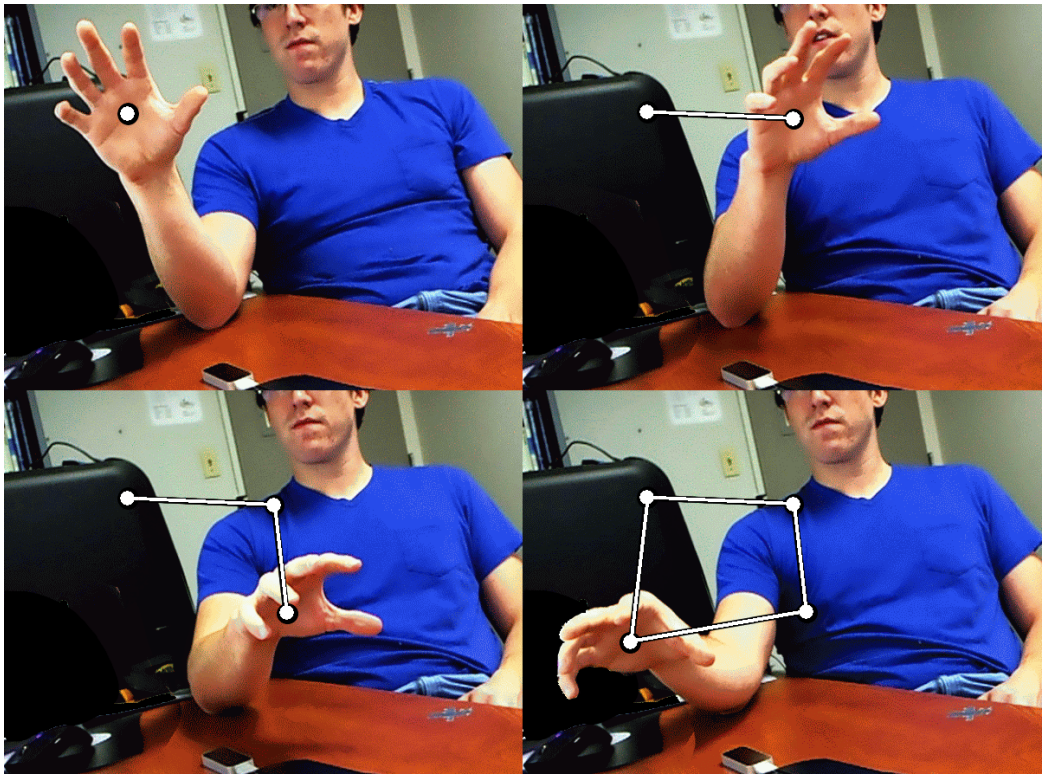

Figure 5.3. During calibration, the software guides users to position their hand at the 4 corners. Users can then rest their elbow on the table during the task, which reduces fatigue.

5.5.1 Adapting Speech to Gestures

We defined recognition latency as the time taken for the system to process a speech-based command. It is measured as the time between the completion of a command utterance and when the command was performed. In our system, the average recognition latency was 700 milliseconds. In addition to the recognition latency, speech-based input also had the overhead of utterance time, or the time it took users to complete the command utterance. In our experiment, the commands "right click" and "left click" took approximately 850 milliseconds. In total, this gave an expected overhead of 1550 milliseconds per trial for the speech based input.

We observed that when using the mouse with speech, participants had a tendency to move the cursor to the target before beginning utterance of the command.

This behavior resulted in a mean speech overhead of 1814ms, close to the 1550ms estimated above. When using gestures with speech however, participants eventually learned to optimize their movement by adapting for these overheads. 4 of the 7 participants used pre-emptive utterance, where they began the utterance of the speech commands before the cursor was on the target. These users had a lower mean speech overhead per trial of 856ms, 990ms, 1000ms and 1102ms respectively. While participants without pre-emptive utterances had a mean speech overhead per trial of 1438ms, 1453ms and 1344ms respectively.

5.5.2 Hyperarticulation

We observed that participants would hyperarticulate in the event of an error, which is consistent with previous work (Oviatt 1996). A normal utterance of the "left click" or "right click" command tended to take approximately 850 milliseconds. If the command was not recognized, subsequent utterances would be hyperarticulated, and therefore take a longer utterance time – approximately 1300 milliseconds. However, we did not observe a "cycle of recognition failure" as mentioned in the previous study. This could simply be attributed to the improvements in speech-recognition technology and fewer words uttered per command.

5.5.3 Throughput

Throughput has been acknowledged as the best way to evaluate performance of pointing tasks (Soukoreff and MacKenzie 2004) and is defined as the ratio between the Index of Difficulty (ID) over Movement Time (MT):

$$\text{Throughput} = \frac{ID}{MT}.$$

and ID is defined as the ratio of the distance traveled over the width of the target $\left(ID = log_2(\frac{D}{W} + 1)\right)$. Throughput means per device is listed in table 5.1. Our experiment had an ID range between 1 to 4.5 bits, well within the recommended range (Soukoreff and MacKenzie 2004).

We used throughput to measure the performance of the navigational aspect of our implementation as it only considers the movement of the cursor, without taking the overhead of the hover or speech into consideration. This allowed us to measure if introducing speech caused a deterioration in navigation. There was a minor increase in means of throughput when speech was added to the interaction. However, a t-test showed this to be statistically insignificant ($t(12) = -1.1766, p = .26$). Likewise, there was no statistically significant difference when speech-based commands were introduced to gestural navigation ($t(12) = -0.7357, p = 0.48$). This tells us that the introduction of speech into navigation tasks does not cause a degradation in performance measured in throughput.

As previously mentioned, one drawback of speech-based input is that it impedes performance. This is due to people finding it more difficult to think and speak at the same time, rather than to think and perform physical actions concurrently. This has been attributed to the lower adoption rates of speech-based input. Our experiment, however, demonstrates that usable gesture and speech multimodal interaction can be built given specific design considerations, such as fewer words per command, and a pointing task already familiar to the user.

5.5.4 Completion Time

The mean time taken for participants to complete each task shown in table 5.1 shows that the gestures and speech interaction time takes twice as long to complete compared to a mouse-only style. This was expected, as participants were much more familiar with the mouse than gestural interaction.

5.5.5 Errors

False negatives refer to instances where the utterance of a command was not recognized by the system. Our system recorded a false negative rate of 6.1% with the mouse, 4.7% with gestures, and an overall false negative rate of 5.4%.

Table 5.1. Means and standard deviation of completion time measured in seconds, and throughput measured in bits per seconds

Interface	Completion Time		Throughput	
	Mean	Stdev	Mean	Stdev
Mouse only	74s	5.19	4.33bps	0.35
Mouse & speech	114s	9.74	4.58bps	0.46
Gestures only	116s	14.76	2.66bps	0.40
Gestures & Speech	136s	14.82	2.83bps	0.49

Wrong clicks refer to instances where a left-click action occurred when a right-click was expected, and vice-versa. No such errors were observed.

Misclicks occurred when a click action was recorded outside of the target. Four such errors were each detected in mouse and gestures. Surprisingly, a total of 11 misclicks were detected in the navigation-only mouse task. We attribute this random error to participants intuitively clicking on the hover target when it appeared.

5.6 Discussion and Future Work

The idea that speech based input and output can be used by blind or motor-impaired users has been introduced in prior work (Shneiderman 2000). Meanwhile multimodal use of both gestures and speech has been shown to be preferred over either one alone (Billinghurst 1998). Future work should therefore assess the possibility of using gestures for cursor manipulation and speech for selection in assistive technology. We believe users with carpal-tunnel, arthritis, or muscle-dystrophy stand to benefit from this method of interaction.

During the course of this experiment, we found that our participants learned to adapt to speech processing latency with preemptive utterance of commands. This behaviour is likely easier to learn when the length of the commands are equal. The commands “right click” and “left click” used in our experiment are both two-syllable phrases which take equal time to vocalize. Future research should investigate the effects of commands with differing lengths (eg: “right”, “left”, “middle click”), and

if they interfere with this optimization learnt by users. Future research should also investigate why this behaviour was only noticed with gestures, but not the mouse.

The likelihood of speech-recognition failure in our experiment was relatively low despite the use of a pre-trained model. However, hyperarticulation caused the cost of this failure to be quite high as it doubles the speech overhead. We expect this cost of failure to be reduced when users are made aware that hyperarticulating the commands will not improve recognition, but could potentially make it worse. Hyperarticulation may simply be a natural human reaction, similar to when users click mouse buttons harder when expected actions or feedback is not observed. It is possible that given sufficient amount of usage, users will realise how hyperarticulation is in fact counter-productive. Alternately a simple training program could be helpful.

5.7 Conclusion

In this paper we evaluated the use of multimodal interaction using gestures and speech. Touchless hand gestures were used for cursor navigation and speech-based input was used for selection. We found that the inclusion of speech did not affect navigation performance which was counter to previous work performed. We also found that users adapted to recognition latency with pre-emptive utterance. This behaviour was only noticed while using gestures as input, not when using the mouse. We observed hyperarticulation in our experiment whenever the speech processor failed to recognise a command, this lead to overall reduced performance. However, we did not notice a cycle of errors caused by this behaviour as mentioned in previous works.

Our implementation draws it's strength by implementing the best features from gestural and speech interaction techniques. The Personal Space approach allowed gestural interaction without fatigue, while fewer pre-trained commands and fewer utterance per command allowed for better speech-recognition. We believe this will lead to multimodal speech and gesture system that can be immediately usable in the real-world.

CHAPTER SIX

Grasp, Grab or Pinch? Identifying User Preference for Touchless Gestural Manipulation using Rule-Based Detection

6.1 Abstract

Touchless gestural interactions are beginning to attract attention from the consumer market, fueled by the availability of commodity gesture-based input and the depiction of this interaction in the media. For this interaction to become a norm, standards must be employed for the most common actions performed. In this paper, we developed a robust rule-based hand gesture recognition system which recognized dynamic hand gestures based on transition between states. We then evaluated 3 different gestures for users to manipulate objects on a screen, which we referred to as *"grasp"*, *"grab"* and *"pinch"*. Finally, we performed a usability study to identify users' preference for gestural manipulation, which showed a strong preference for the *"grasp"* gesture.

6.2 Introduction

Touchless gestural interactions tends to be an effective, intuitive, and natural interaction for users to relay information to the computer (Wachs, Kölsch, Stern, and Edan 2011), making it extremely suited for direct manipulation. Since Bolt's "Put-That-There" interaction in 1980 (Bolt 1980), there has been significant interest in interactive hand gestures. However, this interest was relegated to mainly laboratory usage. Recently though, commodity hardware such as Xbox Kinect and Leap Motion have managed to take gestural interactions out of research labs or Hollywood movies and into the hands of users.

While commodity hardware does exist, these devices are targeted towards non-work, non-purposeful interactions often used in isolation (such as gaming), meanwhile the idea of the desktop was the exact opposite: to gather as much functionality as

possible in one computer (Bødker 2006). But there is good reason to extend the use of touchless gestural interaction to the desktop: they provide a natural and intuitive method of interaction by building on users' pre-existing knowledge of the everyday, non-digital world (Jacob, Girouard, Hirshfield, Horn, Shaer, Solovey, and Zigelbaum 2008), and could therefore be employed for productive use with minimal training or knowledge of existing devices. For example we can imagine using gestures to grab a file and move it to a different directory, or as an input device by users with hand impairments who are unable to use a mouse. Before these interaction techniques can be ubiquitous, however, user research will need to be performed and standards will need to be defined. (Norman and Nielsen 2010) argues that while new technologies require new methods, designers tend to ignore well-established principles in their rush to develop gestural or natural interfaces, leading to usability disasters. They also state that while these new interactions can be effective and even fun, they should not be "inflicted" onto users until it has been validated by user research.

Grandhi, Joue, and Mittelberg (Grandhi, Joue, and Mittelberg 2011) identified two challenges to touchless gestural interactions: 1) the recognition challenge and 2) the vocabulary challenge[1]. The recognition challenge deals with achieving accurate and meaningful gesture recognition while the vocabulary challenge deals with identifying natural, intuitive and meaningful gesture vocabularies appropriate for the tasks. (Grandhi, Joue, and Mittelberg 2011) claim the former challenge has received more research attention than the latter, causing gestures to be prescribed to users based on their ease of implementation. The attention traditionally given to the recognition challenge has resulted in commodity hardware such as the Leap Motion controller (Motion 2012) which is capable of hand recognition with sub-millimeter accuracy (Weichert, Bachmann, Rudak, and Fisseler 2013) and even comes bundled

[1]Nielsen et al (Nielsen, Störring, Moeslund, and Granum 2004) used the terms "technology based vocabulary" and "human based gesture vocabulary" which can be generalized to the recognition challenge and vocabulary challenge respectively

Figure 6.1. The grasp gesture

with a developers kit. This now allows us as interaction researchers to focus our research on the vocabulary challenge, with the recognition challenge a secondary consideration.

In this paper we investigate and propose standard vocabulary for touchless gestural interaction for manipulation tasks – analogous to a drag-and-drop action performed with a mouse – which can be easily performed by users and implemented by designers. We introduce natural and intuitive gestures, built on well established principles, validated by user research and easily recognized by input devices. We started by reviewing research that utilize gestural manipulation and created a generalized version of these interactions which we call *grab* (figure 6.2) and *pinch* (figure 6.3). We also introduce a gesture we call *grasp* (figure 6.1), which our user study shows to be the preferred gesture for manipulation.

6.3 Related Works

We use the term "gestural manipulation" to mean hand gestures used to hold or release which is in line with definitions currently used (Wachs, Kölsch, Stern, and

Figure 6.2. The grab gesture

Edan 2011, Karam et al. 2005, Quek, McNeill, Bryll, Duncan, Ma, Kirbas, McCullough, and Ansari 2002). In our case, we perform manipulation in a desktop environment with 2-dimensional objects, however we have examined literature performing manipulation in both 2D and 3D.

6.3.1 Grab Gesture

One of the earliest version of the grab gesture was used in (Davis and Shah 1994) with a data glove. The authors here revealed a high recognition rate but claimed that their implementation is less natural to the user, but there were no specific metrics regarding the grab gesture itself. This research however is good indication that the grab gesture can be reasonably detected by a monochromatic camera, with markers, and contributes to the recognition challenge. The authors used a Finite-State Machine approach which inspired our own detection methods.

Work done by (Pollard and Zordan 2005) uses motion capture to monitor users' hands when performing manipulation tasks, one of interest to us was holding a cube. While this work was not explicitly done for gestural interaction, the data collected in regards to hand pose, especially the open and closed positions, validate

Figure 6.3. The pinch gesture

the choice of using grab for manipulation tasks. The tasks performed were focused on object manipulation in 3-dimensions.

A *grab* gesture was used in (Freeman, Vennelakanti, and Madhvanath 2012) and called "bloom". The authors noted that user's gestures become more ambiguous and relaxed after prolonged usage. The absence of a rigid recognizer, due to a Wizard-of-Oz approach, allowed the user to be more natural, but raises the question if the hand poses could be discernible both by the system and the user. This influenced us to design a system that actually performed recognition of the *grab* gesture, and was robust enough to accept ambiguity when user's start to relax their hand postures.

The *grab* was also used in (Vogel and Balakrishnan 2005b) where it was referred to as "Grip Clutch" and used to disengage the pointer from the hand. The authors did remark that a clutching gesture such as this exerts some stress on the hands and as a result should not be used for longer periods of time. The authors instead used a static posture to perform manipulation. This brought our attention to evaluating our own gestures for ergonomics and comfort, and alerted us to the possibility that some gestures may not be ideal for interaction even if they seem intuitive and natural as long term use could cause discomfort.

6.3.2 Pinch Gesture

In G-Stalt (Zigelbaum, Browning, Leithinger, Bau, and Ishii 2010), a *pinch* gesture with the thumb and index finger allowed the user to perform similar manipulation tasks. The researchers used a data glove to perform this interaction, thus achieving high recognition rate. We believe this system contributes to the vocabulary challenge by providing a seemingly natural gesture for interaction, the researchers themselves reveal that the gestures in G-Stalt were said to be too complicated and difficult to learn. However, there was no formal user study beyond a demonstration and feedback. Additionally, there was little indication as to which particular gestures were difficult or non-intuitive.

PinchWatch (Loclair, Gustafson, and Baudisch 2010) contributed to the vocabulary challenge by using different *pinch* gestures for "microinteractions." These are interactions that occur in short bursts and therefore require a high amount of precision. The designers intended the system to interpret different actions based on which fingers are used to make contact with the thumb and the position in which the finger contacts the thumb. However, they concede that this is a remarkably difficult task to be performed barehanded. We noted that this gesture was designed to be natural and intuitive with little cognitive load.

A thumb-and-index *pinch* gesture was used in (Grossman, Wigdor, and Balakrishnan 2004) for manipulation of volumetric display, which their user study showed to be simple and usable for users with little prior training. The researchers made the choice to use a pinch gesture largely due to the tracking technology available to them at the time. The authors recommended full finger and palm tracking for future use. This work gave us a gesture that is applicable to both the recognition challenge and the vocabulary challenge.

6.3.3 Other Gestures

The authors of GripSee (Becker, Kefalea, Maël, Von Der Malsburg, Pagel, Triesch, Vorbrüggen, Würtz, and Zadel 1999) moved the index and middle finger to grip and release in a gesture akin to a scissors metaphor, which was based on gestural interactions from (Triesch and Von Der Malsburg 1998). This design was chosen due to limitations of the recognition challenge, with little consideration for the gestures to be natural or intuitive. The researchers concluded that using gestures to control a robot arm was easier than using traditional methods, even though the gestures themselves were unintuitive.

Another research (Kim, Park, Kim, Do, Song, and Bien 2000), used a static gesture where the user curls their fingers in a claw-like posture. This system used a data glove for input and reported a high recognition rate, however there was no justification for the choice of gestures nor user feedback on the gesture's usability ratings. This was interesting to us as a few of our participants used a similar gesture when asked to perform gestures without any training provided.

6.4 Design

6.4.1 Prototype Design

In designing the experiment to assess gestural manipulation, we first designed a task to assess behavior. We decided that the users should be able to select or hold an object on screen, then move it, followed by release. We also decided that a Wizard-of-Oz experiment would not be suitable as it could not be consistently correct or wrong, a crucial factor in allowing people to believe they were actually working with a computer (Dahlbäck, Jönsson, and Ahrenberg 1993). It would also get in the way of participants' ability to adapt (Tennant 1980), and is less effective at building a mental model (Hare, Gill, Loudon, and Lewis 2013), thus possibly limiting our data to only what we expect to find, thereby increasing confirmation bias.

A fully functional software, with high accuracy rates would be ideal for input given sufficient training data, however we decided that it would get in the way of assessing user behavior as they will require more assistance (Hare, Gill, Loudon, and Lewis 2013). This would affect the outcome of our study as we will not be able to assess the robust nature of the gesture nor the users' ability to adapt.

There is a trade-off between the vocabulary challenge and the recognition challenge (Nielsen, Störring, Moeslund, and Granum 2004), similar to the robust vs accuracy trade-off (Wachs, Kölsch, Stern, and Edan 2011), and can be controlled for in the fidelity of the prototype. We therefore designed the prototype to be a balance between the two, which means it was a higher fidelity prototype than the Wizard-of-Oz approach, but without a high recognition rate for robustness. The unfinished nature of the system had little impact in stopping the user from using the device successfully (Hare, Gill, Loudon, and Lewis 2013). In this way, participants were able to operate and receive immediate feedback from the interface with little need to adapt to the device, and allowed us to collect real world data including user adaptation, and preference.

6.4.2 Interaction Design

The design for gestures used for manipulation was based on the 4 usability principles for gestures proposed by Nielson et al (Nielsen, Störring, Moeslund, and Granum 2004): (1) easy to perform and remember, (2) intuitive, (3) metaphorically and iconically logical towards functionality and (4) ergonomic. The authors of the aforementioned paper themselves recommended a static pointing-like gesture for manipulation, however a later study by Grandhi, Joue, and Mittelberg (Grandhi, Joue, and Mittelberg 2011) indicate that a dynamic gesture is preferred over static gestures for manipulation.

The design of the *grab* and *pinch* gestures were inspired from existing implementations as outlined in our Related Works section, while the *grasp* gesture was

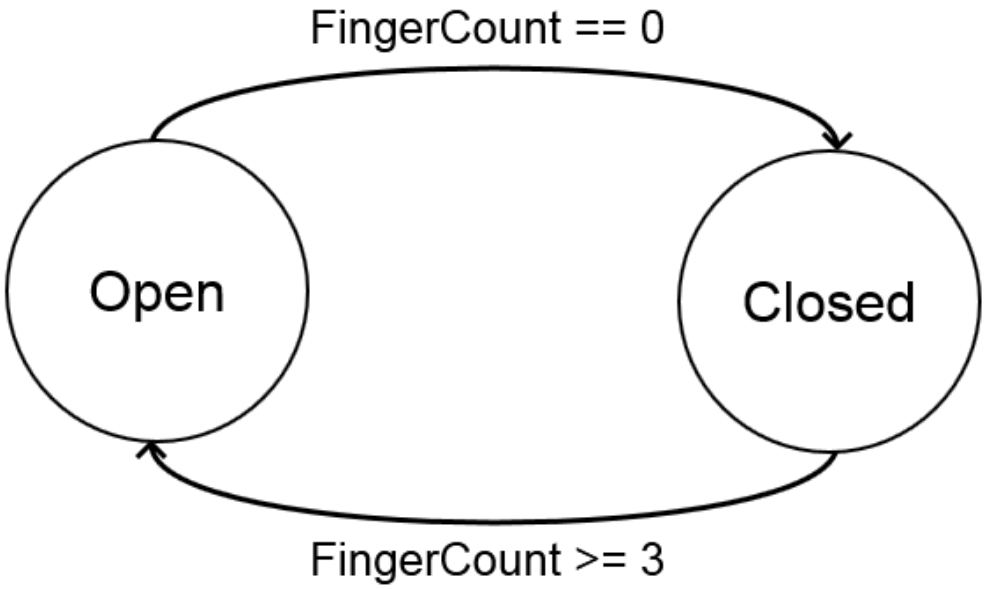

Figure 6.4. The rules that causes the transitions between the *open* and *closed* states of the palm

designed based on the R-family of gestures which is deictic in nature and used in real life for precision grip (Kendon 2004). The R-family of gestures involves the index finger and thumb making contact, but in our implementation, the thumb makes contact with all 4 fingers. This was done to improve recognition and to introduce a hard separation between the *grasp* and *pinch* gestures.

We built a rule-based system that recognizes gestures as the transition between static poses. This prototype was able to handle the prescribed gestures as well as variations by the participants. As a result, our proposed manipulation gestures seemed dynamic to the user but was really the effect of transitions between the "open" and "closed" states of the palm. These states were identified based on the number of fingers detected in the frame. We detected the palm as closed if the palm is visible but no fingers are visible. The transition to an open position from a closed position is only triggered when 3 or more fingers are detected. This can be generalized to a

rule-based system, where the Rule R_1 performs transition from State S_1 to S_2 and Rule R_2 performs transition from S_2 to S_1. But R_1 is not the simple negation of R_2.

We used a regular off-the-shelf Leap Motion controller (Motion 2012) on a Microsoft Windows 7 computer. The software was written in Java and utilized the SDK provided. The controller was adjusted to fit the participant's input method for better recognition. For example we angled the device forward about 30° for participant who decided to rest their elbows on the table.

One drawback of our gesture recognition system was with false negatives. There were occasions when gestures performed were not recognized by the system. In our experiments, we specifically found that this was caused by the user's thumb sticking out too far, which caused the system to detect one finger and therefore did not transition to a closed state. This problem was of little consequence as the users would quickly learn and adapt with some coaching.

6.5 User Study

A user study was conducted to gather feedback from users preference for hand gestures used to perform gestural manipulation. 2 tasks were designed to perform this evaluation: manipulation of a 3-dimensional object and manipulation of a 2-dimensional object. Qualitative feedback such as comfort and usability concerns were also collected from participants.

6.5.1 Participants

A total of 17 participants (F=7) between the ages of 18 to 22 were recruited to participate in the user study. All participants were undergraduates who were compensated for their participation.

6.5.2 Procedure

Participants were first shown that they are able to control the mouse cursor on the desktop with their palms. Once the participant got accustomed to moving

the cursor with their palm, they were given the 3D manipulation task; a cube was presented on screen, which could be moved in 3-dimensions on a fixed axis, equivalent to a pitch, yaw or roll. The participants were asked to "hold the cube and move it around." No specific instructions were given as to what "hold" means, and no gestures were demonstrated to them. participants therefore used their own form of gestures, which the researcher observed and noted. The researcher then demonstrated the 3 proposed gestures to the participants and they were asked to manipulate the cube again, this time with the proposed gestures. After allowing participants to attempt all 3 gestures, the participants were then asked to rank the gestures by preference. This experimental design was inspired by (Wachs, Kölsch, Stern, and Edan 2011) which identified several techniques of identifying intuitive gestures including "teaching by example." The author also recommended that gestures should not be selected during the design stage but by interacting with the user. We took both recommendations into account by first observing the user's natural way of performing gestural interaction, and then demonstrating our prescribed interaction styles.

After the 3D manipulation task, participants were then asked to move a window from one position of the screen to another, again with the 3 gestures proposed. There was no additional training provided, nor were the participants given any time limit to perform the action. participants had to first navigate the mouse cursor to hover over the title bar and then perform the gestures. Since the cursor position was dependent on the palm position, and performing any hand gestures also causes the palm to move, the action of performing hand gestures will therefore cause the cursor to move. The target size was 40 pixels in height, which is the height of a Windows 7 titlebar. The small target size and natural hand-tremors (Benko, Wilson, and Baudisch 2006), made this task more difficult. The 2D gestural manipulation task therefore required a higher level of precision by the participants.

A think-aloud protocol (Lewis 1982) was used during the experiment. participants were encouraged to describe the gestures they were attempting and whether they felt any discomfort. After the final tasks, participants were asked to rank the 3 gestures by preference for each individual task and overall preference. Each experiment ran for a maximum of 30 minutes but most participants would complete in 20 minutes.

6.6 Results

Participants ranked their preference with a score of 2 for their first choice, 1 for their second choice and 0 for their third choice. Although 3 different preferences were collected: large 3D object manipulation, small 2D object manipulation, and overall preference, our primary interest was in the overall preference and 2D task preference as it was more representative of the tasks performed on a WIMP based system. The results of the overall preference provided in table 6.1 shows the *grasp* gesture as a strong preference with a total of 11 of the 17 participants (64.7%) selecting it as their first choice. A weighted preference rank gives the grasp gesture 27 points of the possible 34 (17 x 2) and therefore a z-score of 0.79 which places it in the fourth quartile, while the grab and pinch gestures are z-scored at 0.2 and 0.5 respectively. A Friedman's test shows this results to be statistically significant, $\chi^2(2) = 11.7647, p = 0.002$.

It is also notable that none of the participants chose the *grab* action as their first choice. Additionally, there was no difference between the overall preference and the 2D interaction preference. The difference between the 3D interaction preference and the overall preference however was slightly different, with two users preferring the *grab* over the *grasp*. Due to the robust nature of the rule-based recognition system, participants were able to adapt quickly, modifying the prescribed gestures to fit their own comfort level. The *grasp* gesture was especially prone to adaptation whereby the participants would use a much more relaxed posture. During the think-aloud,

Table 6.1. Ranking of user preferences; higher is better.

Participant	Grasp	Grab	Pinch
1	2	1	0
2	2	1	0
3	1	0	2
4	2	0	1
5	2	1	0
6	2	1	0
7	2	0	1
8	2	0	1
9	2	0	1
10	2	1	0
11	0	1	2
12	2	1	0
13	1	0	2
14	2	0	1
15	1	0	2
16	1	0	2
17	1	0	2
Total	27	7	17

three of the participants remarked that the prescribed *grasp* posture required some amount of stress on their fingers but were able to find a more relaxed version. All three nonetheless chose *grasp* as their preferred gesture. All participants pointed out that the *grab* gesture did not feel as precise during the 2D task, whereas the *grasp* (or *pinch*, depending on their preference) felt very precise. None of the participants reported fatigue or discomfort.

6.6.1 Observations

We observed that when instructed to "hold the cube and move it around" without much context nor training, the users were understandably confused, but quickly learnt what it means, and performed gestures based on their understanding of the request. We consider this a validation of our prototype design, as the users will notice consistent feedback from the system based on their input, which allowed them to understand the effects of their actions, and allowed them to explore better.

Most of the users performed gestures that were variations of our *grab*, *pinch*, or *grasp*, interestingly 2 participants performed used the claw-like posture as mentioned in (Kim, Park, Kim, Do, Song, and Bien 2000). However, after the 3 prescribed gestures were demonstrated, these participants switched preference, noting that any of the 3 were more natural than the gesture they initially performed. This is interesting to us because it implies that certain gestures might be intuitive to a user despite not being natural.

Participants themselves noticed the impact that hand tremors had on the accuracy of their manipulation task. Participants would naturally rest their elbows on the table in an attempt to stabilize their hands and achieve higher accuracy.

6.6.2 Feedback

Upon completion of the experiment, participants were asked for subjective qualitative feedback. All s stated they were impressed with the interaction, with the word "cool" frequently used to express the fact. Two participants expressed their desire to use the interaction style immediately while two others were more cautious, saying that while interesting, they did not see this interaction style replacing their mouse just yet.

6.6.3 Variations in Gestures

The gestures prescribed were designed to be strict and rigid. As we expected, participants managed to find variations to fit their own comfort level. The robust nature of our recognition system allowed them to experiment until such a comfortable gesture was found, such as that found in figure 6.5. Our system was also capable of recognizing participants with the "hitchhikers thumb" (figure 6.6) whose *grasp* gesture was not significantly different from the prescribed version, but was not explicitly catered for in design.

Figure 6.5. A more comfortable version of the *grasp* gesture was eventually discovered and used by most participants.

6.7 Contribution

Our focus was to contribute to the vocabulary for gestural manipulation in a desktop environment. We demonstrate how the *grasp* gesture is an excellent choice as it draws strength from the following:

- Natural & Intuitive - The *grasp* gesture is based on the R-family of gestures which is used in everyday life for precision gestures (Kendon 2004).
- Robust to adaptation - We show that participants in our research will find a comfortable interpretation of the *grasp* gesture during the experiment itself and within the recognition capabilities of the system.
- Built on well established principles - We use dynamic gestures for manipulation as recommended by (Grandhi, Joue, and Mittelberg 2011), while interactions were assessed with a high fidelity prototype, which adds validity (Dahlbäck, Jönsson, and Ahrenberg 1993, Tennant 1980, Hare, Gill, Loudon, and Lewis 2013) to our results.
- Validated by user research - We conducted a user study to collect feedback and to obtain user preference ratings which gives the *grasp* gesture a high preference rating.

Figure 6.6. The *grasp* gesture as performed by a participant with "Hitchhiker's Thumb"

- Easily recognized by user input - This gesture and its variations can be recognized with a simple rule-based system.

6.8 Discussion and Implication

Apart from the results and contributions mentioned above, there are some discoveries in regards to our implementation and gestural design that should be considered, which are listed below.

6.8.1 Recommended Gesture

We found that there were a few reasons the *grasp* gesture was preferred for 2-dimensional object manipulation. The first is the fact that this gesture felt more precise: it allowed users to visualize their target by simulating a pointing action with the tips of their fingers. In addition, the gesture itself was robust enough such that users could modify it to fit their own comfort level. As pointed out by (Vogel and Balakrishnan 2005b), the *grab* gesture requires more stress, but the *grasp* allows a more comfortable posture.

6.8.2 *Gestures and Cultural Significance*

It is important for designers to know that gestures are not universal, a seemingly innocent hand gesture in one culture may be insulting in another. For example, the "thumbs-up" gesture is generally used to indicate a "good job" and has been identified in studies as a strong user preference to issue the "OK" or "Confirm" action to the computer (Nielsen, Störring, Moeslund, and Granum 2004). This experiment was conducted in a number of different locations, including one study conducted in Vietnam (Mai, Hai, and Son 2011) and one in the United States (Aigner, Wigdor, Benko, Haller, Lindlbauer, Ion, Zhao, Koh, and Redmond 2012). However, this gesture is considered vulgar in parts of North Africa and the Middle East (Foster 2002). A literature review revealed no indication of the *grasp* or *pinch* gesture violating any social norms, but the same cannot be said about the *grab* gesture (Foster 2002). *Grab* is considered rude in the Middle East and North Africa. Allowing users to select or calibrate their preferred gestures that they find offensive is necessary in designing a globally acceptable software.

6.8.3 *Robust Gestural Recognition*

Machine learning techniques may potentially allow more robust gestural recognition models, however sufficiently large data sets are required (Mitra and Acharya 2007). Multiply that by the vocabulary challenge and we have a much larger problem. There is a robust-accuracy trade-off that will need to be made; allowing more gestures to be recognized by the system will reduce accuracy (Wachs, Kölsch, Stern, and Edan 2011) and is something we experienced ourselves even with our rule-based system. We scaled the system to allow more false positives than false negatives, favoring recall over precision, which we considered acceptable in our setting.

6.8.4 *Intuitive vs Learned Gestures*

The "walk up and use" criteria assumes the users should be able to use the interface with minimal training (Olsen Jr 2007). In gestural interaction, this requires a prescribed gesture that takes into consideration the cultural significance mentioned above, as well as being robust enough to handle input variation. We propose that the *grasp* gesture is a good candidate in this area.

6.8.5 *Usability*

Hand tremor makes precise selection of small targets a difficult task (Benko, Wilson, and Baudisch 2006). Our participants found it more difficult to manipulate the smaller object which was 40 pixels in height. We expect this to be a factor in any similar touchless hand gesture interaction and recommend allowing users to rest their elbows on a surface to achieve higher stability and accuracy.

6.9 *Future Works*

Based on this experiment, we expect the results of our findings to be utilized in a few different ways:

- Application - Touchless hand interaction has been considered especially useful to the medical community (Wachs, Stern, Edan, Gillam, Feied, Smith, and Handler 2006, Johnson, O'Hara, Sellen, Cousins, and Criminisi 2011, Wachs, Stern, Edan, Gillam, Handler, Feied, and Smith 2008) who require a sterile method of interacting with a computer. We expect the grasp gesture to be particularly useful here.
- Machine Learning - Given the fact that we now have reliable results for users' preference for gestural manipulation, it would be easier to build training data to create better machine learning models which can recognize the *grasp* gesture and variations with higher accuracy.

- Accessibility - We propose that this interaction can be used by users with hand impairments which limit their ability to use a mouse, such as those with arthritis or carpal tunnel syndrome.
- HCI Research - The rule-based approach used in this paper demonstrates the possibility of building high-fidelity prototypes to identify users' preference. We expect this approach to be used to better identify users' preference in performing other gestures such as gestural selection. One possible use-case is to completely replace the mouse with gestures.

6.10 Conclusion

The emergence of commodity hardware could potentially improve our current interaction with the desktop and WIMP based interfaces by incorporating gestures. But to do so, user preference has to be understood and interaction styles standardized. In this paper we present the *grasp* gesture for gestural manipulation which was designed based on well-established principles and backed by user research. We showed that this gesture itself is natural, intuitive and robust enough for users to modify and adapt to fit their own individual interpretation. A rule-based system can be easily built to recognize this gesture, which allows it to be immediately used in gestural interfaces or for further research.

6.11 Acknowledgments

We thank all volunteers for their participation and valuable feedback.

CHAPTER SEVEN

Discussion, Future Work and Conclusion

7.1 Discussion

7.1.1 Calibration

The key to the Personal Space interaction is to allow the users to rest their elbow on a surface, and then define their interaction space. Defining the interaction space was performed through a calibration process, where users specify the points in real-world, three dimensional space, that they can actually reach from the rested position. These points were then mapped to on-screen coordinates.

The exact implementation of the calibration and translation is subject to implementation. In this study, a planar model is used, whereby a flat plane is built based on the four corners reachable by the user from the rested position. The software guides users to position their hands in these four corners and collects points representing the center of the users palm for 2.5 seconds. The Leap Motion was configured to sample at 90 points per second, and therefore a total of 225 points were collected per corner. Each point represents a coordinate in three dimensional space in a right-hand coordinate system.

These points represent a cluster of points per corner, which may or may not contain outliers. The first thing to do is therefore to build a tight cluster, and prune out outliers. This is done through the DBSCAN algorithm which works in O(n log n) time complexity (Ester, Kriegel, Sander, and Xu 1996). This takes approximates 5 seconds per corner and was therefore done in parallel, with 4 threads or one thread per corner.

Four points are then built – one per corner – based on the users' calibration to mark these four corners. These points are built rather than selected from the cluster

to allow for the smallest interaction space possible based on the calibration. This guarantees that the user can definitely reach these points during their interaction. The points are defined based on the following criteria:

- Top Right: Lowest value of X, lowest value of Y, lowest value of Z
- Top Left: Highest value of X, lowest value of Y, lowest value of Z
- Bottom Left: Highest value of X, highest value of Y, highest value of Z
- Top Right: Lowest value of X, highest value of Y, highest value of Z

This study does not evaluate the correctness of building these four points. Preliminary tests however indicate that it is a good approach, or at least better than selecting the center point of the cluster as the input into the matrix.

Using the DBSCAN algorithm was an effective method of removing outliers. However, the complexity of the algorithm was an issue as it is the sole bottleneck in the calibration process despite being parallelized. Improvements will definitely need to be considered in order to make pruning of outliers more efficient.

7.1.2 Transformation

The transformation from the input in three dimensions to the output in two dimensions is done with matrix multiplication: Source $\times$ Transformation = Destination. Following the calibration process there were four points in right-three dimension cartesian coordinate system which were then used to build the source matrix:

- Top Right: $(x_{s_1}, y_{s_1}, z_{s_1})$
- Top Left: $(x_{s_2}, y_{s_2}, z_{s_2})$
- Bottom Left: $(x_{s_3}, y_{s_3}, z_{s_3})$
- Bottom Right: $(x_{s_4}, y_{s_4}, z_{s_4})$

The next matrix that was built is the Destination matrix, which is dependent on the resolution of the monitor. Assuming a output device with height H and width W, and a coordinate system which starts from the top left corner, the coordinates corresponding to the four corners can be defined as:

- Top Right: $(W, 0)$
- Top Left: $(0, 0)$
- Bottom Left: $(0, H)$
- Bottom Right: (W, H)

The values needed to create the Source and Destination matrices were then available, with which the Transformation matrix was calculated as: Transformation = Source$^{-1}\times$ Destination. Inversing the source matrix is easily performed with a square matrix, as such the Source and Destination matrices were both padded with ones to make it a 4-by-4 matrix:

$$Source = \begin{bmatrix} x_{s_1} & y_{s_1} & z_{s_1} & 1 \\ x_{s_2} & y_{s_2} & z_{s_2} & 1 \\ x_{s_3} & y_{s_3} & z_{s_3} & 1 \\ x_{s_4} & y_{s_4} & z_{s_4} & 1 \end{bmatrix} \; Destination = \begin{bmatrix} W & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & H & 1 & 1 \\ W & H & 1 & 1 \end{bmatrix}$$

The Transformation matrix built was persisted throughout the interaction. With every movement of the palm, the input coordinates were multiplied with the Transformation matrix to obtain the corresponding on-screen coordinates. The new interaction Source matrix is built based on the current palm position: $[\ x_i \quad y_i \quad z_i \quad 1\ ]$. Multiplying this value by the Transformation matrix gives us the corresponding on-screen coordinates: $[\ x_o \quad y_o \quad A \quad B\ ]$. Only x_o and y_o from the resulting row vector is used, as the screen is in 2D wheres the input is in 3D. The remaining items in the vector: A and B are both ignored.

7.1.3 Z-coordinate Issues

The Personal Space approach takes the z-coordinate as a factor in the input as it rightfully should. Simply ignoring the z-space will not create a correct model of the interaction as the user's hand moves in all 3 dimensions. In this approach where the model built is a quadrilateral plane, the interaction space is in fact a 2-dimensional structure built in 3-dimensional space, as shown in figure 7.1. In other

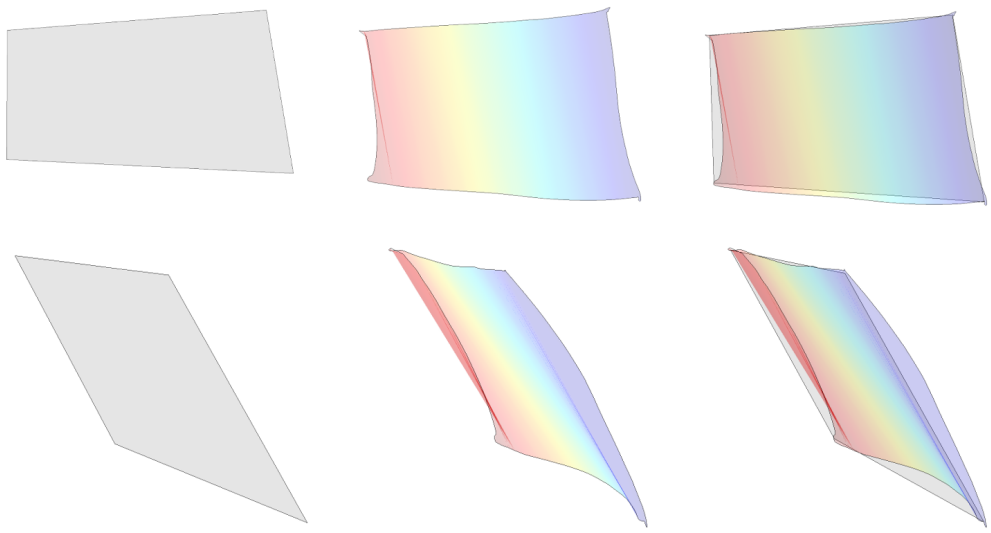

Figure 7.1. Building the interaction space. Top row shows the view from the front while the bottom row is the view of the space from a diagonal view. The first column shows the interaction space actually built based on the matrix calculation,the second column shows the actual hand motion in three dimensions, the third column shows the difference between the quadrilateral plane built and the actual hand movement.

words, although the interaction space is in three dimensions, it is not volumetric. Meanwhile, the matrix based transformation above requires input in three dimensions which means the z-space is a factor in building the output.

There is therefore a problem if the z-coordinate of the palm is not exactly on the flat quadrilateral plane. This is in fact not achievable in real life for 2 reasons:

(1) The quadrilateral plane is not an accurate representation of hand movement whereby the curvature of hand motion is ignored. As a result the palm will almost always never "touch" the plane on the z-axis except for the four corners, as illustrated in figure 7.1. This issue is specific to this particular model; a more accurate representation of the model which accounts for the curvature in the interaction will solve this issue.

(2) The elbow and palm is not in a static position. Some movement is expected to occur during the interaction by the elbow or the palm. As a result, there has to be some correction performed by the system to adjust the Z-coordinate

and "push" (or "pull") the z-coordinate values back onto the plane. This issue will exist in any model built to represent the interaction space as it is a user-related problem.

In order to correct for the above z-coordinate issues, some preprocessing was required. In this study, the Z value is pre-calculated based on the X and Y positions with a separate matrix. Based on the calibration, there were four points in three-dimensional space representing the four corners of the interaction space. From these four corners, a new Transformation matrix was built such that Transformation = Source$^{-1}\times$ Destination where:

$$Source = \begin{bmatrix} x_{s_1} & y_{s_1} & 1 & 1 \\ x_{s_2} & y_{s_2} & 1 & 1 \\ x_{s_3} & y_{s_3} & 1 & 1 \\ x_{s_4} & y_{s_4} & 1 & 1 \end{bmatrix} \; Destination = \begin{bmatrix} z_{s_1} \\ z_{s_2} \\ z_{s_3} \\ z_{s_4} \end{bmatrix}$$

This preprocessing step was built in this manner as it would be a lot more modular than simply ignoring the Z values without any preprocessing. In this way, it is possible to correct or fix the preprocessing step whenever needed.

The preprocessing method used was not explored in depth in this study despite being an important part of the interaction, as the matrix preprocessing step used was demonstrated to be usable, compared to when the interaction was performed without the z-space correction. A potentially more accurate correction method could be to find the z value of the interaction space perpendicular to the input coordinate. In doing so, this correction would fix the coordinate in all three-dimensions, therefore changing the X, Y and Z values rather than only the Z value. This method of correction could potentially have the extra benefit of being usable in a surface fitted model of the interaction space which accounts for the curvature, as opposed to the matrix correction method used which only solves the problem when the interaction space is built based on the planar model, as is done used in this study.

To consider which correction method to use, researchers should consider both the mathematical correctness as well as human factors. For example, we know that the length of the forearm is static per user. The distance between the palm and the elbow will never change. The elbow is the anchor for this interaction, therefore if the movement of the elbow can be detected, the interaction space can be moved to account for the difference.

While the length of the forearm is always static, the distance between the palm and the elbow also depends on the angle of the palm. The participants in this study were all asked to hold their palm in a stable position, however, there is a very high probability of the palm position moving, if just so slightly. The size of the palm could potentially be collected during the calibration process and then used as input to building the model of the palm. This way, the expected position of the palm can be calculated based on the palm angle during input, which would allow better input during the matrix transformation process as the input would be the position of the palm if the palm were held straight, rather than the input of the palm regardless of the angle.

7.2 Future Work

As each manuscript above has it's own section on future work, this section focuses on the overarching theme of making gestures a ubiquitous interaction on the desktop. To do so, a few questions to be addressed have are detailed below.

7.2.1 Modeling the User Space

The key to transforming the user's palm position to an on-screen position is an accurate modeling of the user's input space. The approach used here is rather naive, as it uses a flat plane. A better version would take the curves into consideration, as shown in figure 3.3. It is worth noting that a sphere fit method would be able to

model the user's space quite well, with the arm length being equivalent to the radius of the sphere.

7.2.2 Elbow Position

The Personal Space approach used here requires for the user to keep their elbow in the same position throughout the interaction. While this works for experimental purposes, it can be safely assumed that this does not make for a good form of interaction. Future researchers would have to look into methods of moving the interaction space defined by users based on where the elbow is rested.

7.2.3 Interaction Use-case

The experiments described above focuses on quantitative results to indicate that gestures are in fact possible for inclusion in desktops. However designers and researchers alike would have to consider how to best utilize this interaction. Is it best used in a multi-modal setting? If so what other modality can be considered. This study demonstrates the possibility of including voice, but what about other devices? Perhaps combining it with the keyboard or the mouse or even a touchscreen.

7.2.4 Gestures vs Mouse

This study uses gestures in a manner similar to a mouse, and yet it is demonstrated that the performance of the mouse and the touchpad were both much better than gestures. Does that mean we can never replace the mouse? How can we design interactions that utilize the best of both worlds. Perhaps shorter interactions with larger targets might be more suitable for gestures. For example, moving a window from one screen to another may be easier done with gestures as users will be able to *grasp* anywhere on the window. This could take a shorter time than moving the hand from the keyboard, to the mouse, then moving the cursor to the target, and then moving the window.

7.2.5 Accessibility

This method of interaction could potentially replace the mouse on the desktop for users with impairments affecting the hand. Users with carpal tunnel or arthritis who find it difficult to use a mouse might find it easier to simply use gestures. In addition, the ability to use both hands without significant performace impact could encourage users to alternate hands in pointing tasks, thus reducing the risk of developing repeated-strain injuries.

7.3 Conclusion

In this study, we demonstrate approaches that can be taken to ease the introduction of gesture-based interaction to the desktop.

In *Chapter Three* a novel method of interaction was introduced, which modeled the interaction space of the user through a calibration space. This interaction was demonstrated to be no worse that existing de facto methods used with gestures, but with reduced fatigue. This allows prolonged interactions for gestures, a key factor in desktop interaction.

The interaction introduced in *Chapter Three* is further explored in *Chapter Four*, and benchmarked against other pointing devices currently used for desktop interaction. Here it is shown that while performance of gestures is lower compared to the mouse and the keyboard, there is minimal degradation between hands and high amounts of performance improvements between both hands. This shows that gestures is suitable to be used on both hands, and the fact that users were learning this interaction more than they were learning the mouse.

Speech-based input is introduced in *Chapter Five* alongside gestures to perform gestural selection. This experiment shows that this form of multimodal interaction is in fact possible under certain constraints such as fewer words. While performance is not better than the mouse, it is demonstrated that the introduction of speech does not hinder the performance of gestures. In addition, users tend to learn to

utter commands earlier than needed, potentially further improving interaction over prolonged periods.

Chapter Six evaluated users preference for gestural manipulation and introduced the *grasp* gesture which was preferred by a majority of the participants. This gesture can be easily recognized by the recognition software and can potentially be used in desktop interaction in real-world software immediately.

More work will definitely need to be performed before gestural interaction becomes as ubiquitous on the desktop as the mouse currently is. It is hoped that this study provides a starting point to that end.

APPENDICES

APPENDIX A

Solution Design

A.1 Problem

The first problem with gestural interaction, is the position of the arm. A common method prescribed for gestural interaction involves the elbow being elevated as in figure A.1.

This introduces fatigue very quickly, approximately 90 seconds as per our study. This is a problem in itself, but even more so on the desktop where prolonged interaction is expected.

A.2 Solution

A simple solution would be to allow the user to rest their elbow, but this then creates another problem as the interaction space is no longer a rectangle, as the user will not be able to reach the far edge of the screen. To solve for this, we create a model of the users interaction space, or specifically we map the regions that are reachable by the user without lifting their elbow.

In our case we build a model through a calibration phase. The software directs the user to move their arm to the 4 corners of the screen and builds a suitable model which is a flat quadrilateral plane.

The quadilateral plane is shown in figure A.2 The top row is a front view while the bottom view is a diagonal view of the interaction space. The third column illustrates the difference between the two, and we see that the most amount of difference is in the center and the bottom of the model.

The main algorithm for building the interaction is in the matrix multiplication. It has been shown that Source $\times$ Transformation = Destination, and the

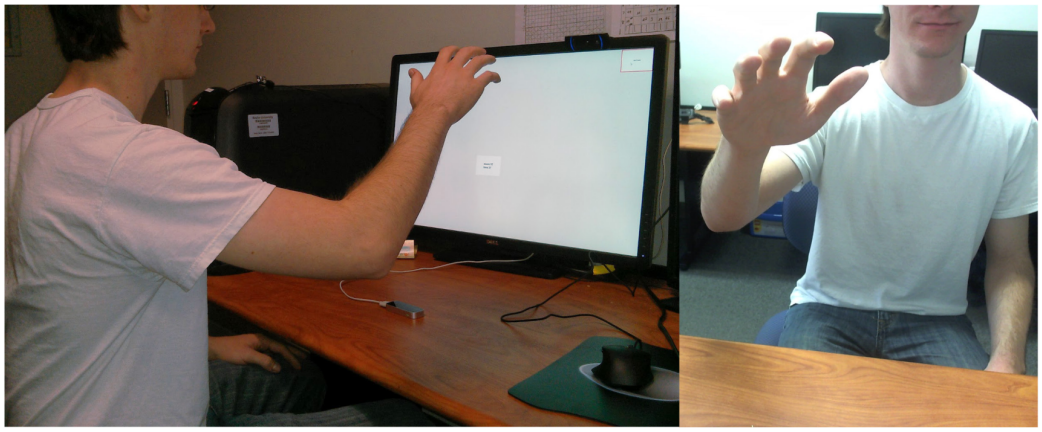

Figure A.1. Commonly used gestural interaction method.

Transformation matrix is calculated during the calibration stage. The minimum necessary variables for the Source matrix are four coordinates in three dimensions, which means the smallest matrix that can be built is a 4×3 matrix:

$$Source = \begin{bmatrix} x_{s_1} & y_{s_1} & z_{s_1} \\ x_{s_2} & y_{s_2} & z_{s_2} \\ x_{s_3} & y_{s_3} & z_{s_3} \\ x_{s_4} & y_{s_4} & z_{s_4} \end{bmatrix}$$

Likewise, the minimum number of variables needed in the Destination matrix are three coordinates in two dimensions:

$$Destination = \begin{bmatrix} x_{s_1} & y_{s_1} \\ x_{s_2} & y_{s_2} \\ x_{s_3} & y_{s_3} \\ x_{s_4} & y_{s_4} \end{bmatrix}$$

Based on the information above, there are three ways in which we can solve for Transformation:

The main algorithm for building the interaction is in the matrix multiplication. It has been shown that Source $\times$ Transformation $=$ Destination, and the

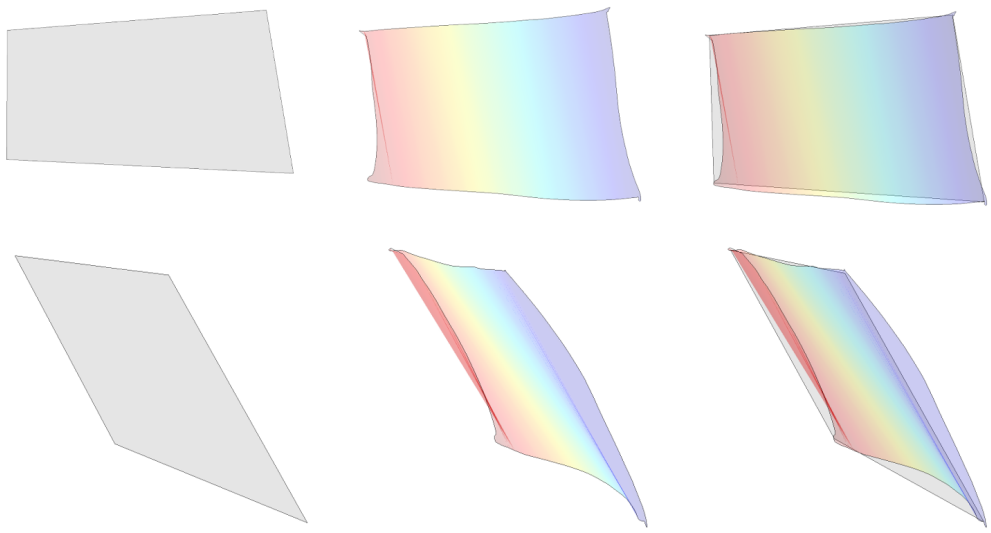

Figure A.2. Comparison of the actual and calibrated space.

Transformation matrix is calculated during the calibration stage. The minimum necessary variables for the Source matrix are four coordinates in three dimensions, which means the smallest matrix that can be built is a 4×3 matrix:

$$Source = \begin{bmatrix} x_{s_1} & y_{s_1} & z_{s_1} \\ x_{s_2} & y_{s_2} & z_{s_2} \\ x_{s_3} & y_{s_3} & z_{s_3} \\ x_{s_4} & y_{s_4} & z_{s_4} \end{bmatrix}$$

Likewise, the minimum number of variables needed in the Destination matrix are three coordinates in two dimensions:

$$Destination = \begin{bmatrix} x_{s_1} & y_{s_1} \\ x_{s_2} & y_{s_2} \\ x_{s_3} & y_{s_3} \\ x_{s_4} & y_{s_4} \end{bmatrix}$$

Based on the information above, there are three ways in which we can solve for Transformation with a few methods of linear solvers:

(1) One-padding and inverse – This is the implementation described in *Chapter Seven*. It is usable primarily as there is only one column that is padded. This will not work, however if the Source matrix only uses the X and Y coordinates, ignoring the Z and therefore requiring 2 columns to be padded as this will cause the matrix to be invertible.

(2) Pseudo inverse – Does not require a square matrix, thus allowing inversion without first performing any padding. This method yields a Transformation matrix that is very close to the former method, but there will bi loss of significant bits. As a result, the Transformation matrix will not be accurate. Our initial tests show that this causes the interaction to be approximately 100 pixels off.

(3) Right-hand division – This method is available in MATLAB (denoted by $A\backslash B$) and a number of implementations of Matrix arithmetic libraries, including EJML which is used in this study. This solution algorithmically selects the ideal algorithm to solve the system of linear equations. This method is not used as the exact implementation is not clearly documented, making it difficult to fully understand the exact method in which the solvers work.

APPENDIX B

Perfomance Measurements

In figure B.1 we see the performance analysis comparing the distance of the target against the completion time. The trend line is calculated based on a robust regression method using the bisquare algorithm, and shows minimal difference between the Personal Space approach introduced in *Chapter Three* and the defacto method used in commercial products today.

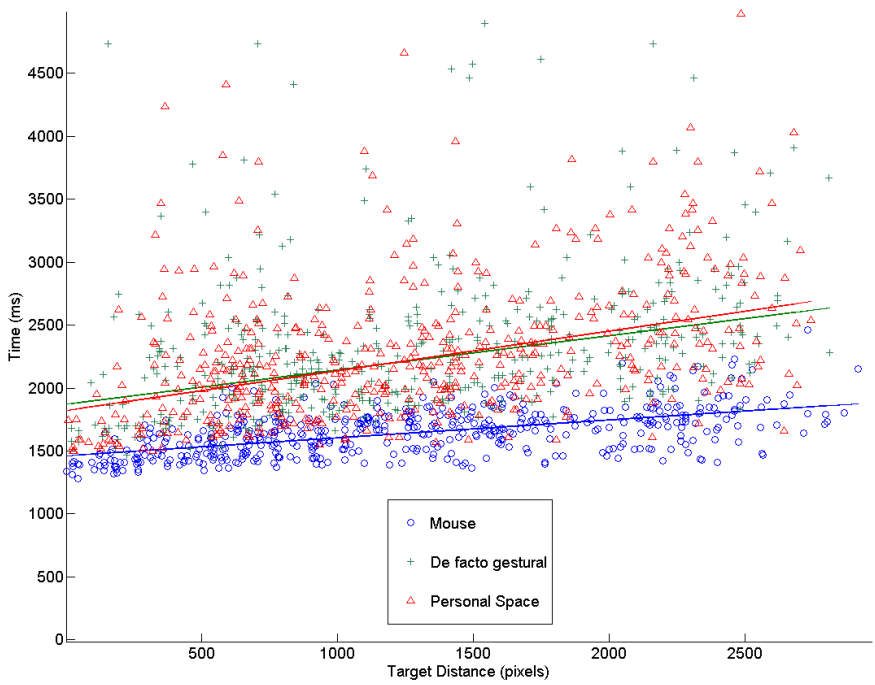

Figure B.1. Target distance vs completion time.

By comparing performance measured in completion time against performance measured in throughput, we see the reason the latter is more suited for pointer device evaluation. Completion time is highly sensitive to outliers, with a possibility of

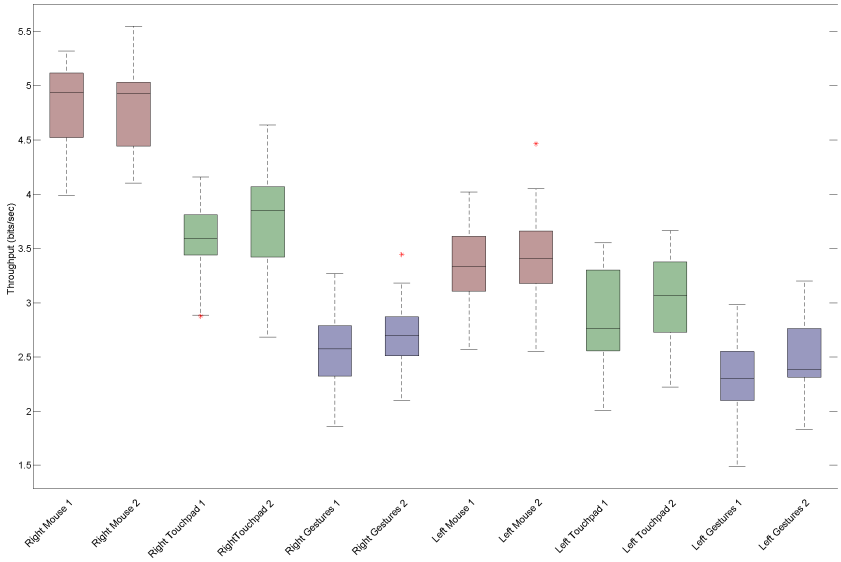

Figure B.2. Graphical representation of device and hand performance measured in throughput

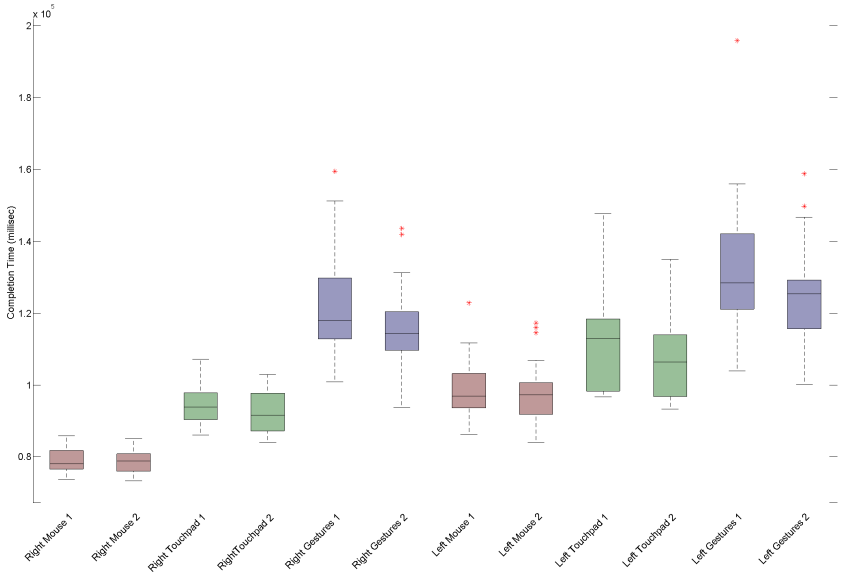

Figure B.3. Graphical representation of device and hand performance measured in throughput

unfairly penalizing the entire task in the event of one error. Whereas throughput or specifically the average task throughput considers the performance of each trial,

with the target size and distance as weights. In addition, the use of a log function in throughput gives a smoother graph with fewer outliers.

In addition, the use of completion time makes the results very dependent on the experiment used to assess it. Throughput on the other hand is more representative of the performance of the device and can be generalized across different experiments.

APPENDIX C

Exit Surveys

Experiments used in manuscript 2 and 3 which appear in chapters four and five respectively end with an exit survey. These exit surveys are included here, with headers indicating their source.

C.1 Manuscript 2

Subject ID: _____ Age:_____

1. Do you own a personal computer (eg: desktop, laptop, netbook)?
   [ ] yes [ ] No

2. How much time do you spend on a computer each week?
   [ ] 0 to 1 hour [ ] 21 to 30 hours
   [ ] 1 to 5 hours [ ] 31 to 40 hours
   [ ] 6 to 10 hours [ ] 41 to 50 hours
   [ ] 11 to 20 hours [ ] More than 50 hours

3. Have you used gestural controllers before (Xbox Kinect, Leap Motion, or any other gesture device)? If so please indicate the type of device.
   _____

4. Do you have any physical impairment that makes it difficult to use a computer? If so please list the impairment
   [ ] Yes (please state) :_____
   [ ] No

5. Which is your dominant hand?
   [ ] Right hand [ ] Left hand [ ] Ambidextrous

6. Which hand do you normally use to control the mouse?
   [ ] Right [ ] Left

7. Which hand do you normally use to control the touchpad?
   [ ] Right [ ] Left

Please indicate if you agree or disagree with the following statements:

1. I experienced discomfort and fatigue while using gestures with my preferred hand in this test.
   [ ] strongly agree [ ] agree [ ] neutral [ ] disagree [ ] strongly disagree

2. I experienced discomfort and fatigue while using gestures with my non-preferred hand in this test.
   [ ] strongly agree [ ] agree [ ] neutral [ ] disagree [ ] strongly disagree

3. I experienced discomfort and fatigue while using the Touchpad with my preferred hand in this test.
   [ ] strongly agree [ ] agree [ ] neutral [ ] disagree [ ] strongly disagree

4. I experienced discomfort and fatigue while using the Touchpad with my non-preferred hand in this test.
   [ ] strongly agree [ ] agree [ ] neutral [ ] disagree [ ] strongly disagree

5. I liked using a touchpad in this test.
   [ ] strongly agree [ ] agree [ ] neutral [ ] disagree [ ] strongly disagree

6. I liked using the gestural input in this test.
   [ ] strongly agree [ ] agree [ ] neutral [ ] disagree [ ] strongly disagree

C.2 Manuscript 3

Subject ID: _____ Age:_____

1. Do you own a personal computer (eg: desktop, laptop, netbook)?
   [ ] yes [ ] No

2. How much personal time do you spend on a personal computer each week?
   [ ] 0 to 1 hour [ ] 21 to 30 hours
   [ ] 1 to 5 hours [ ] 31 to 40 hours
   [ ] 6 to 10 hours [ ] 41 to 50 hours
   [ ] 11 to 20 hours [ ] More than 50 hours

3. Do you have any physical impairment that makes it difficult to use a computer?
   [ ] yes [ ] No

4. Which is your dominant hand?
   [ ] Right hand [ ] Left hand [ ] Ambidextrous

5. Which hand do you normally use to control the mouse?
   [ ] Right [ ] Left

Please indicate if you agree or disagree with the following statements

1. I usually experience discomfort and fatigue while using a mouse.
   [ ] strongly agree [ ] agree [ ] neutral [ ] disagree [ ] strongly disagree

2. I experienced discomfort and fatigue while using gestures in this test.
   [ ] strongly agree [ ] agree [ ] neutral [ ] disagree [ ] strongly disagree

3. I experienced discomfort and fatigue while using the mouse in this test.
   [ ] strongly agree [ ] agree [ ] neutral [ ] disagree [ ] strongly disagree

4. For the tasks performed today, I prefer using the mouse over gestures.
   [ ] strongly agree [ ] agree [ ] neutral [ ] disagree [ ] strongly disagree

BIBLIOGRAPHY

Aigner, R., D. Wigdor, H. Benko, M. Haller, D. Lindlbauer, A. Ion, S. Zhao, J. T. K. V. Koh, and W. Redmond (2012). Understanding mid-air hand gestures: A study of human preferences in usage of gesture types for hci. *Microsoft Research TechReport MSR-TR-2012-111*.

Argyros, A. A. and M. I. Lourakis (2006). Vision-based interpretation of hand gestures for remote control of a computer mouse. In *Computer Vision in Human-Computer Interaction*, pp. 40–51. Springer.

Becker, M., E. Kefalea, E. Maël, C. Von Der Malsburg, M. Pagel, J. Triesch, J. C. Vorbrüggen, R. P. Würtz, and S. Zadel (1999, April). Gripsee: A gesture-controlled robot for object perception and manipulation. *Auton. Robots 6*(2), 203–221.

Benko, H., A. D. Wilson, and P. Baudisch (2006). Precise selection techniques for multi-touch screens. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1263–1272. ACM.

Billinghurst, M. (1998). Put that where? voice and gesture at the graphics interface. *ACM SIGGRAPH Computer Graphics 32*(4), 60–63.

Biswas, K. and S. K. Basu (2011). Gesture recognition using microsoft kinect®. In *Automation, Robotics and Applications (ICARA), 2011 5th International Conference on*, pp. 100–103. IEEE.

Bødker, S. (2006). When second wave hci meets third wave challenges. In *Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles*, NordiCHI '06, New York, NY, USA, pp. 1–8. ACM.

Bolt, R. A. (1980). *Put-that-there: Voice and gesture at the graphics interface*, Volume 14. ACM.

Carmody, T. (2010). Why gorilla arm syndromerules out multitouch notebook displays. *Wired, Oct 10*.

Coe, R. (2002). It's the effect size, stupid: What effect size is and why it is important. *Paper presented at the British Educational Research Association annual conference, Exeter, UK*.

Cohen, J. (1992). A power primer. *Psychological bulletin 112*(1), 155.

Dahlbäck, N., A. Jönsson, and L. Ahrenberg (1993). Wizard of oz studieswhy and how. *Knowledge-based systems 6*(4), 258–266.

Davis, J. and M. Shah (1994, Apr). Visual gesture recognition. *Vision, Image and Signal Processing, IEE Proceedings - 141*(2), 101–106.

E-Speaking LLC. e-speaking. `http://www.e-speaking.com`. Accessed: 2014-01-30.

Ester, M., H.-P. Kriegel, J. Sander, and X. Xu (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Volume 96.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology 47*(6), 381.

Foster, D. (2002). *The Global Etiquette Guide to Africa and the Middle East: everything you need to know for business and travel success*. John Wiley & Sons.

Freeman, D., R. Vennelakanti, and S. Madhvanath (2012). Freehand pose-based gestural interaction: Studies and implications for interface design. In *Intelligent Human Computer Interaction (IHCI), 2012 4th International Conference on*, pp. 1–6. IEEE.

Goble, J. C., K. Hinckley, R. Pausch, J. W. Snell, and N. F. Kassell (1995). Two-handed spatial interface tools for neurosurgical planning. *IEEE Computer 28*(7), 20–26.

Grafton, S. T., E. Hazeltine, and R. B. Ivry (2002). Motor sequence learning with the nondominant left hand. *Experimental Brain Research 146*(3), 369–378.

Grandhi, S. A., G. Joue, and I. Mittelberg (2011). Understanding naturalness and intuitiveness in gesture production: Insights for touchless gestural interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, New York, NY, USA, pp. 821–824. ACM.

Grossman, T., D. Wigdor, and R. Balakrishnan (2004). Multi-finger gestural interaction with 3d volumetric displays. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, New York, NY, USA, pp. 61–70. ACM.

Hare, J., S. Gill, G. Loudon, and A. Lewis (2013). The effect of physicality on low fidelity interactive prototyping for design practice. In *Human-Computer Interaction–INTERACT 2013*, pp. 495–510. Springer.

Hauptmann, A. G. and P. McAvinney (1993). Gestures with speech for graphic manipulation. *International Journal of Man-Machine Studies 38*(2), 231–249.

Ishii, H. (2008a). Tangible bits: beyond pixels. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*, pp. xv–xxv. ACM.

Ishii, H. (2008b, June). The tangible user interface and its evolution. *Commun. ACM 51*(6), 32–36.

Jacob, R. J., A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum (2008). Reality-based interaction: a framework for post-wimp interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 201–210. ACM.

Johnson, R., K. O'Hara, A. Sellen, C. Cousins, and A. Criminisi (2011). Exploring the potential for touchless interaction in image-guided interventional radiology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, New York, NY, USA, pp. 3323–3332. ACM.

Jude, A., G. M. Poor, and D. Guinness (2014). Personal space: User defined gesture space for gui interaction. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, New York, NY, USA, pp. 1615–1620. ACM.

Kabbash, P., I. S. MacKenzie, and W. Buxton (1993). Human performance using computer input devices in the preferred and non-preferred hands. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, pp. 474–481. ACM.

Karam, M. et al. (2005). A taxonomy of gestures in human computer interactions.

Keir, P. J., J. M. Bach, and D. Rempel (1999). Effects of computer mouse design and task on carpal tunnel pressure. *Ergonomics 42*(10), 1350–1360.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kim, J.-S., K.-H. Park, J.-B. Kim, J.-H. Do, K.-J. Song, and Z. Bien (2000). Study on intelligent autonomous navigation of avatar using hand gesture recognition. In *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, Volume 2, pp. 846–851. IEEE.

Kopp, S. (2013). Giving interaction a hand: Deep models of co-speech gesture in multimodal systems. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ICMI '13, New York, NY, USA, pp. 245–246. ACM.

Lai, K., J. Konrad, and P. Ishwar (2012). A gesture-driven computer interface using kinect. In *Image Analysis and Interpretation (SSIAI), 2012 IEEE Southwest Symposium on*, pp. 185–188. IEEE.

Lewis, C. (1982). *Using the" thinking-aloud" method in cognitive interface design*. IBM TJ Watson Research Center.

Lin, C.-H. J., K. J. Sullivan, A. D. Wu, S. Kantak, and C. J. Winstein (2007). Effect of task practice order on motor skill learning in adults with parkinson disease: a pilot study. *Physical therapy 87*(9), 1120–1131.

Loclair, C., S. Gustafson, and P. Baudisch (2010). Pinchwatch: a wearable device for one-handed microinteractions. In *Proc. MobileHCI*.

Lyons, M. J. (2004). Facial gesture interfaces for expression and communication. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, Volume 1, pp. 598–603. IEEE.

MacKenzie, I. S. et al. (2013). A note on the validity of the shannon formulation for fitts index of difficulty. *Open Journal of Applied Sciences 3*(06), 360.

MacKenzie, I. S., T. Kauppinen, and M. Silfverberg (2001). Accuracy measures for evaluating computer pointing devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 9–16. ACM.

Mai, N. T. T., T. T. T. Hai, and N. V. Son (2011). Wizard of oz for designing hand gesture vocabulary in human-robot interaction. In *Knowledge and Systems Engineering (KSE), 2011 Third International Conference on*, pp. 232–238. IEEE.

Mitra, S. and T. Acharya (2007). Gesture recognition: A survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 37*(3), 311–324.

Motion, L. (2012). Leap. *URL: https://www. leapmotion. com/[last accessed 2014-02-04]*.

Nielsen, J. (1993, April). Noncommand user interfaces. *Commun. ACM 36*(4), 83–99.

Nielsen, M., M. Störring, T. B. Moeslund, and E. Granum (2004). A procedure for developing intuitive and ergonomic gesture interfaces for hci. In *Gesture-Based Communication in Human-Computer Interaction*, pp. 409–420. Springer.

Norman, D. A. (2010, May). Natural user interfaces are not natural. *interactions 17*(3), 6–10.

Norman, D. A. and J. Nielsen (2010). Gestural interfaces: a step backward in usability. *Interactions 17*(5), 46–49.

Olsen Jr, D. R. (2007). Evaluating user interface systems research. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, pp. 251–258. ACM.

Oviatt, S. (1996). User-centered modeling for spoken language and multimodal interfaces. *IEEE multimedia 3*(4), 26–35.

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

Pino, A., E. Tzemis, N. Ioannou, and G. Kouroupetroglou (2013). Using kinect for 2d and 3d pointing tasks: performance evaluation. In *Human-Computer Interaction. Interaction Modalities and Techniques*, pp. 358–367. Springer.

Pollard, N. S. and V. B. Zordan (2005). Physically based grasping control from example. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '05, New York, NY, USA, pp. 311–318. ACM.

Poor, G., B. J. Tomlinson, D. Guinness, S. D. Jaffee, L. M. Leventhal, G. Zimmerman, and D. S. Klopfer (2013). Tangible or gestural: Comparing tangible vs. kinect interactions with an object manipulation task. In *7th International Conference on Tangible, Embedded and Embodied Interaction*, TEI '13, New York, NY, USA. ACM.

Quek, F., D. McNeill, R. Bryll, S. Duncan, X.-F. Ma, C. Kirbas, K. E. McCullough, and R. Ansari (2002). Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction (TOCHI) 9*(3), 171–193.

Rogers, Y., H. Sharp, and J. Preece (2011). *Interaction Design: Beyond Human-Computer Interaction*. John Wiley & Sons.

Sambrooks, L. and B. Wilkinson (2013). Comparison of gestural, touch, and mouse interaction with fitts' law. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration*, pp. 119–122. ACM.

Schmidt, R. A. and T. Lee (1988). *Motor Control and Learning, 5E*. Human kinetics.

Sears, A., C. Plaisant, and B. Shneiderman (1991). A new era for touchscreen applications: High precision, dragging icons, and refined feedback. *Advances in human-computer interaction 3*.

Segen, J. and S. Kumar (2000, July). Look ma, no mouse! *Commun. ACM 43*(7), 102–109.

Seidler, R. D. (2004). Multiple motor learning experiences enhance motor adaptability. *Journal of cognitive neuroscience 16*(1), 65–73.

Shneiderman, B. (2000, September). The limits of speech recognition. *Commun. ACM 43*(9), 63–65.

Soukoreff, R. W. and I. S. MacKenzie (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of fitts law research in hci. *International Journal of Human-Computer Studies 61*(6), 751–789.

Sullivan, K. J., B. J. Knowlton, and B. H. Dobkin (2002). Step training with body weight support: effect of treadmill speed and practice paradigms on poststroke locomotor recovery. *Archives of physical medicine and rehabilitation 83*(5), 683–691.

Teixeira, V. (2011). *Improving elderly access to audiovisual and social media, using a multimodal human-computer interface*. Ph. D. thesis, Faculdade de Engenharia, Universidade do Porto.

Tennant, H. R. (1980). Evaluation of natural language processors. Technical report, DTIC Document.

Triesch, J. and C. Von Der Malsburg (1998). Robotic gesture recognition by cue combination. In *Informatik98*, pp. 223–232. Springer.

van Dam, A. (1997, February). Post-wimp user interfaces. *Commun. ACM 40*(2), 63–67.

Vatavu, R.-D. (2012). User-defined gestures for free-hand tv control. In *Proceedings of the 10th European Conference on Interactive Tv and Video*, EuroiTV '12, New York, NY, USA, pp. 45–48. ACM.

Vogel, D. and R. Balakrishnan (2005a). Distant freehand pointing and clicking on very large, high resolution displays. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pp. 33–42. ACM.

Vogel, D. and R. Balakrishnan (2005b). Distant freehand pointing and clicking on very large, high resolution displays. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology*, UIST '05, New York, NY, USA, pp. 33–42. ACM.

Von Hardenberg, C. and F. Bérard (2001). Bare-hand human-computer interaction. In *Proceedings of the 2001 workshop on Perceptive user interfaces*, pp. 1–8. ACM.

Wachs, J., H. Stern, Y. Edan, M. Gillam, C. Feied, M. Smith, and J. Handler (2006). A real-time hand gesture interface for medical visualization applications. In *Applications of Soft Computing*, pp. 153–162. Springer.

Wachs, J. P., M. Kölsch, H. Stern, and Y. Edan (2011, February). Vision-based hand-gesture applications. *Commun. ACM 54*(2), 60–71.

Wachs, J. P., H. I. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied, and M. Smith (2008). A gesture-based tool for sterile browsing of radiology images. *Journal of the American Medical Informatics Association 15*(3), 321–323.

Weichert, F., D. Bachmann, B. Rudak, and D. Fisseler (2013). Analysis of the accuracy and robustness of the leap motion controller. *Sensors (Basel, Switzerland) 13*(5), 6380.

Wigdor, D. and D. Wixon (2011). *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier.

Wu, Y. (2000). What else should an hci pattern language include. *Procs. of Pattern Languages for Interaction Design: Building Momentum@ CHI*.

Yoo, J., S. Lee, and C. Ahn (2012). Air hook: Data preloading user interface. In *ICT Convergence (ICTC), 2012 International Conference on*, pp. 163–167. IEEE.

Zhai, S. (2004). Characterizing computer input with fitts law parametersthe information and non-information aspects of pointing. *International Journal of Human-Computer Studies 61*(6), 791–809.

Zhang, X. and I. S. MacKenzie (2007). Evaluating eye tracking with iso 9241-part 9. In *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*, pp. 779–788. Springer.

Zigelbaum, J., A. Browning, D. Leithinger, O. Bau, and H. Ishii (2010). G-stalt: A chirocentric, spatiotemporal, and telekinetic gestural interface. In *Proceedings of the Fourth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '10, New York, NY, USA, pp. 261–264. ACM.