ABSTRACT

Models for Rested Touchless Gestural Interaction

Darren Guinness, M.S.

Mentor: G. Michael Poor, Ph.D.

Touchless mid-air gestural interaction has gained mainstream attention with the emergence of off-the-shelf commodity devices such as the Leap Motion and the Xbox Kinect. One of the issues with this form of interaction is fatigue, a problem colloquially known as the "Gorilla Arm Syndrome." However, by allowing interaction from a rested position, whereby the elbow is rested on a surface, this problem can be limited in its effect. In this paper we evaluate 3 possible methods for performing touchless mid-air gestural interaction from a rested position: a basic rested interaction, a simple calibrated interaction which models palm positions onto a hyperplane, and a more complex calibration which models the arm's interaction space using the angles of the forearm as input. The results of this work found that the two modeled interactions conform to Fitts's law and also demonstrated that implementing a simple model can improve interaction by improving performance and accuracy.

Models for Rested Touchless Gestural Interaction

by

Darren Guinness, B.S.

A Thesis

Approved by the Department of Computer Science

Gregory D. Speegle, Ph.D., Interim 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

Greg Hamerly, Ph.D.

Peter Maurer, Ph.D.

A. Alexander Beaujean, Ph.D..

Accepted by the Graduate School
August 2015

J. Larry Lyon, Ph.D., Dean

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

TABLE OF CONTENTS

LIST OF FIGURES viii

LIST OF TABLES x

ACKNOWLEDGMENTS xi

DEDICATION xii

1 Introduction . . . . . . . . . . 1

1.1 Gestures . . . . . . . . . . 2

1.2 Classification . . . . . . . . . . 3

1.3 Current State . . . . . . . . . . 3

1.4 Fatigue . . . . . . . . . . 4

1.5 Our Work . . . . . . . . . . 5

2 Literature Review . . . . . . . . . . 6

2.1 Gestural Interaction . . . . . . . . . . 6

2.1.1 Gestural Fatigue . . . . . . . . . . 7

2.1.2 Gestural Pointing . . . . . . . . . . 9

2.1.3 Gestural Selection . . . . . . . . . . 10

2.2 Learning Effects . . . . . . . . . . 10

2.3 Pointing Device Evaluation . . . . . . . . . . 11

2.3.1 Standard Tasks . . . . . . . . . . 13

2.3.2 Fitts' Law . . . . . . . . . . 13

2.4 Accuracy Measures . . . . . . . . . . . . . . . . . . . . . . . . . . 15
2.4.1 Discrete Measures . . . . . . . . . . . . . . . . . . . . . . . . . 15
2.4.2 Continuous Measures . . . . . . . . . . . . . . . . . . . . . . . 17

3 Preliminary Investigations . . . . . . . . . . . . . . . . . . . . . . . . 19
3.1 Pilot 1 : Foundations . . . . . . . . . . . . . . . . . . . . . . . . . 19
3.1.1 Pointing Interactions . . . . . . . . . . . . . . . . . . . . . . . 19
3.1.2 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
3.1.3 Overview . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
3.2 Pilot 2: New Setup . . . . . . . . . . . . . . . . . . . . . . . . . . 24
3.3 Pilot summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27

4 Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
4.1 Interaction Design . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
4.1.1 Unmodeled Gestural Interaction . . . . . . . . . . . . . . . . . 29
4.1.2 Rested & Calibrated Interaction – Hyperplanar Model . . . . 30
4.1.3 Rested & Calibrated Interaction – Spherical Model . . . . . . 31
4.2 Experimental Design . . . . . . . . . . . . . . . . . . . . . . . . . . 33
4.2.1 Participants . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
4.2.2 Apparatus . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
4.2.3 Task . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
4.2.4 Design . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
4.2.5 Procedure . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34

5 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
5.1 Daily Improvement . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
5.2 Fitts' Regression . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36

5.3 Performance . . . . . . . . . . 37
5.3.1 Movement Time . . . . . . . . . . 37
5.3.2 Bivariate Throughput . . . . . . . . . . 38
5.4 Accuracy Measures . . . . . . . . . . 38
5.4.1 Target Entries . . . . . . . . . . 38
5.4.2 Task Axis Crosses . . . . . . . . . . 39
5.4.3 Movement Direction Change . . . . . . . . . . 39
5.4.4 Orthogonal Direction Change . . . . . . . . . . 39
5.4.5 Movement Variability . . . . . . . . . . 40
5.4.6 Movement Error . . . . . . . . . . 40
5.4.7 Movement Offset . . . . . . . . . . 40
5.5 Subjective User Feedback . . . . . . . . . . 41
5.5.1 Preference . . . . . . . . . . 41
5.5.2 Usability . . . . . . . . . . 41
5.5.3 Fatigue . . . . . . . . . . 42
5.5.4 Borg Scale . . . . . . . . . . 42
6 Analysis . . . . . . . . . . 44
7 Discussion . . . . . . . . . . 47
7.1 Discussion and Future Works . . . . . . . . . . 47
7.2 Conclusion . . . . . . . . . . 49
APPENDICES 50
APPENDIX A Tables, Figures and Transformations . . . . . . . . . . 51
A.1 Tables . . . . . . . . . . 51

A.2 Figures . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52

A.3 Transformation . . . . . . . . . . . . . . . . . . . . . . . . . . . 54

APPENDIX B Original Hyperplanar Description . . . . . . . . . . . . . . . . . 56

B.1 Solution Design . . . . . . . . . . . . . . . . . . . . . . . . . . . 56

B.1.1 Problem . . . . . . . . . . . . . . . . . . . . . . . . . . . 56

B.1.2 Solution . . . . . . . . . . . . . . . . . . . . . . . . . . . 57

BIBLIOGRAPHY . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 60

LIST OF FIGURES

2.1 User calibrating their space taken from (Jude et al., 2014b) with permission. This method of calibration was used for both the *Hyperplanar* and *Spherical* models. . . . . . 8

2.2 Multi-Directional task implementation from (Soukoreff and MacKenzie, 2004) used with permission . . . . . 14

2.3 An example of a Target Re-entry as seen in (MacKenzie et al., 2001) used with permission . . . . . 16

2.4 An example of a Target Axis Crossing as seen in (MacKenzie et al., 2001) used with permission . . . . . 16

2.5 An example of three Movement Direction Changes as seen in (MacKenzie et al., 2001) used with permission . . . . . 16

2.6 An example of two Orthogonal Direction Changes as seen in (MacKenzie et al., 2001) used with permission . . . . . 17

3.1 Gestural interaction models . . . . . 20

3.2 Throughput means and standard deviation from days 1-5 grouped by interaction . . . . . 22

3.3 User's forearm is being measured in the pronated position with the palm down . . . . . 25

3.4 Ergonomic Rest used to reduce strain on the elbow and forearm . . . 26

3.5 The stand used in the experiment (left), and the one used in the pilot (right) . . . . . 27

4.1 A front view of the Unmodeled (blue plane) and Hyperplanar (grey plane) interaction space. Both axis denotes interaction space measured in millimeters. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
4.2 A top view of the Unmodeled (red plane) and Hyperplanar (grey hyperplane) with tracked hand movement (black points). Both axes denote a dimension of the interaction space measured in millimeters. . . . . . 30
4.3 L-R: (1) Front view of *Spherical model*, (2) side view of the *Spherical* interaction, (3) plotting the angles $(\theta, \phi)$ which is essentially a quadrangle. 32

5.1 Mean movement time of all participants from day 3 as a function of effective Index of Difficulty (IDe), with computed Fitts' regression lines 37

A.1 Mean movement time of all participants from day 1 as a function of effective Index of Difficulty (IDe), with computed Fitts's regression lines 52
A.2 Mean movement time of all participants from day 2 as a function of effective Index of Difficulty (IDe), with computed Fitts's regression lines 53
A.3 Mean movement time of all participants from day 3 as a function of effective Index of Difficulty (IDe), with computed Fitts's regression lines 53

B.1 Commonly used gestural interaction method. . . . . . . . . . . . . . . 56
B.2 Comparison of the actual and calibrated space. . . . . . . . . . . . . 57

LIST OF TABLES


2.1 The differences between actual paths(blue), and the optimal path (grey), and how these are characterized by each of the continuous accuracy metrics. Used from (MacKenzie et al., 2001), with permission. . . . . 18

3.1 Performance improvement between days 1-5, reported in Cohen's *d*. . 22

3.2 Arm length approximation correctness (%) . . . . . . . . . . . . . . . 23

5.1 Mean ($\bar{x}$) and standard deviation ($\sigma$) of bivariate throughput for each device across days 1-3. The corresponding $d$ values indicate difference in performance from the previous day measured with Cohen's d. . . . 36

5.2 Metrics for Fitts' regression of each interaction. . . . . . . . . . . . . 36

5.3 Mean and standard deviation of movement time per trial for each interaction in milliseconds. Smaller values are better. . . . . . . . . . . 37

5.4 Mean and standard deviation of accuracy measures metrics taken on day 3 of the experiment. Bold text denotes a result of statistical significance . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 39

5.5 Mean reported usability & comfort metrics per interaction (1 = Most Negative, 5 = Most Positive). . . . . . . . . . . . . . . . . . . . . . 41

5.6 Mean reported fatigue per interaction (1 = Extreme, 5 = None) . . . 42

5.7 Mean reported effort per interaction (0 = Nothing at all, 0.5 = very very weak (just noticable), 1 = very weak, ..., 10 very, very strong) . 42

A.1 Means and standard deviation of throughput for each device per day 51

A.2 Means and standard deviation per trial of each accuracy metric on day 5 51

A.3 Intercept, slope, coefficient of determination ($R^2$) and correlation coefficient ($R$) of all 3 interactions over all 3 days. . . . . . . . . . . . 52

## ACKNOWLEDGMENTS

This work would not have been possible without the assistance of so many who have helped guide, and support me throughout this graduate program. I will first express my appreciation to my advisor Dr. Poor, for helping me get my start in the area of Human-Computer Interaction, guiding me through hectic times, and patiently working with me until I was able to fully contribute. To Alvin Jude, who's work ethic inspired me to push further even when everything seemed overwhelming, and for being a constant sounding board for new ideas. To Ashley Dover for taking precious time out of her days and weekends when we needed to meet deadlines.

To Dr. Greg Hamerly, who helped expand my perspective, and who's feedback helped me communicate more effectively. To Dr. Greg Speegle, who's experience helped me during the writing of this work. To Dr. Pablo Rivas-Perea for his assistance in recruiting of participants . To Dr. Peter Maurer for helping with the transformations.

To the Department of Computer Science at Baylor who educated and supported me throughout the program. Pat Hynan, and George Gonzalas who's responsiveness, and support allowed us to test and produce work much more efficiently. Sharon Humphrey for keeping track of the activity and conference events associated with the lab. To all participants, who helped drive the direction of the research via performance and feedback. Lastly, to my family for their continuous support and feedback.

*"If you wish to make an apple pie from scratch, you must first invent the universe."*

*-Carl Sagan*

# CHAPTER ONE

## Introduction

In the past decade a large amount of work has been conducted looking into new methods to interact with computers. This work has allowed touch interfaces such as tablets, mobile phones, and public displays to become ubiquitous through continuous improvements. Other newer methods such as input using Natural Language have also been improved to the point that commercial interfaces are starting to offer speech-based input services such as Siri, Google Now, and Amazon's Alexa. Reality-based Interfaces (RBIs) like the Xbox Kinect, Playstation Move, and Leap Motion which incorporate metaphors from the physical world are also starting to emerge. However, the traditional Desktop paradigm remains relatively unchanged, and is heavily relied upon in areas which require high performance and long durations, such as business and areas of gaming. This can be problematic, because not all users have the same needs, and thus better incorporation for other alternative input methods, can produce benefits for users who have difficulty with the standard paradigm.

The standard methods of interaction for the desktop computer has traditionally been limited to that of the Keyboard and Mouse. Although this desktop paradigm is still widely used and preferred by many, there are limitations which require good alternative input methods for the desktop computer. One such limitation is that the required motion of the keyboard and mouse can be a deterrent for persons who cannot physically use these methods due to disabilities or impairments. For instance, if a user had a hand impairment, or even if that user was missing a hand, they may have difficulties interacting with the computer using the traditional keyboard and mouse. This is partially addressed by commodity hardware that focuses on ergonomics such as ergonomic mice and keyboards, but the root of the issue lies in the interaction itself (Fagarasanu and Kumar, 2003). Subsequently, research done into the effects of mouse use has found that concentrated mouse use is associated with Carpal Tunnel (Keir et al., 1999). Other work has found strong evidence of a causal relationship between keyboard and mouse usage and Carpal Tunnel Symdrome (Fagarasanu and Kumar, 2003). This essentially means that there is a barrier to entry in standard desktop input, which even when overcome, the use of this particular input paradigm may result in impairment.

One type of interaction that can potentially cater to these issues is Gestural Interaction, or more commonly referred to as "Gestures." Gestural interfaces can be developed using different parts of the body as input, meaning that the user who had hand impairments, or a lack of a hand, would be able to interact using another body part such as their eyes, foot, or even forearm. This would seemingly allow users who had difficulties interacting with the traditional keyboard and mouse to interact with the desktop by leveraging the motions that they can perform or prefer.

## *1.1 Gestures*

Gestures in the context of Human-Computer Interaction (HCI) encompass a large group of techniques used to interact with a computer. The accepted definition of a gesture in HCI literature was defined by Kurtenbach and Hulteen to be "a motion of the body that contains information" (Kurtenbach and Hulteen, 1990). This broad definition lends itself to a wide research area which has incorporated expressions elicited from various areas of the body including the legs, face, eyes, fingers, and even hands (Han et al., 2011, Zelinsky and Heinzmann, 1996, Drewes and Schmidt, 2007, Grossman et al., 2004, Cockburn et al., 2011).

The fact that different body parts, and motions can be used makes it theoretically a powerful addition to the accessibility community. This was demonstrated last year, when gestural cursor navigation was evaluated in a case study by two persons who had impairments which made it difficult to use the mouse (Guinness et al., 2014). Although this finding is limited, there is little reason that interfaces such as the one mentioned above, cannot be extended to people without hands, or even arms by utilizing the motion in the forearms, or even the legs or feet to interact with the computer.

## *1.2 Classification*

The two main classifications of gestures are touch-based gestures, and *mid-air* gestures which is also known as touchless gestural interaction. Recently hybrid approaches, combining both touch and mid-air gestures, have also been studied (Chen et al., 2014). Touch-based gestures involve meaningful strokes executed on a touch-sensitive surface such as a touchscreen on a tablet, or cellular phone display (Poppinga et al., 2014), while *Mid-air* gestures are concerned with motion made from the body in mid air.

*Mid-air* gestural interaction implementations are further broken up into those that require the user to wear physical devices such as a glove, or ring, and those that require no physical contact which is known as the *Bare-handed interaction style*. In the latter, users interact with the computer without any device or wires attached. Von described this interaction style to be superior to implementations that required extra devices to be worn (Von Hardenberg and Bérard, 2001). While both approaches have their advantages, the *Bare-handed interaction style* is theorized to be a better form of input because it incorporates the "Come As You Are" design principle (Triesch and Von Der Malsburg, 1998). This principal states that users should not be required to wear a glove or specific markers to interact with the system (Wachs et al., 2011).

## *1.3 Current State*

Commodity gestural recognition hardware has become increasingly more precise, with devices in some cases demonstrated to be accurate up to a sub-millimeter level (Guna et al., 2014) with low latency (Brown et al., 2014). Due to these advances, researchers have started looking at gestures not simply for coarse actions, but for more fine use cases such as cursor navigation (Jude et al., 2014b, Brown et al., 2014, Pino et al., 2013, Sambrooks and Wilkinson, 2013), which require high precision, and accuracy.

It is important in these high precision and repetition use cases, that new methods are able to perform well in comparison to existing methods. This determined by providing comparisons between new and old methods, or retroactively by examining literature using the same tasks over similar parameters (Soukoreff and MacKenzie, 2004). Comparisons are performed using measures that describe the interface such as Performance, Accuracy, and consistency. When we looked at gestural cursor navigation in comparison to other input methods in a similar Index of Difficulty (ID) Range (1-4 bits) for those with Motor Impairments, we find that our implementation of gestural cursor navigation fits right above that of a Trackball, and below an ergonomic Mouse (Wobbrock and Gajos, 2008). This demonstrates promise as the gestural implementation performs similarly, but directs the motion to be elicited from the larger muscle groups such as the forearm, biceps, and shoulders rather than relying on wrist, and hand movements which has been previously associated with carpal tunnel.

### *1.4 Fatigue*

The most common implementations of gestural input involve the user holding their arms out in mid-air. This mode of interaction results in arm and shoulder fatigue (Teixeira, 2011, Wachs et al., 2011, Hincapié-Ramos et al., 2014) commonly referred to as "Gorilla Arm Syndrome" (Carmody, 2010). This will need to be addressed before gestures can be accepted as a ubiquitous mode of interaction. Brown et al identified a simple way to address this problem by allowing the user to rest the elbow on a surface (Brown et al., 2014). However, this simple solution may result in an interaction that is non-intuitive to the user, because it gives very little consideration to the actual mechanics of the human arm. In order to address this, Jude et al introduced a potential improvement called Personal Space (Jude et al., 2014b) which used a calibration step to first map the user's input space.

Both Brown et al, and Jude et al, addressed fatigue by performing gestural interaction from a rested position. However, each author makes a different claim: Brown et al (Brown et al., 2014) stated that fatigue is addressed simply by resting the elbow, while Jude et al (Jude et al., 2014b) claimed that fatigue was addressed by modeling the user's input space from a rested position.

### *1.5 Our Work*

In this work, we seek to investigate both approaches further, using standard evaluation methodologies provided by the ISO 9241-9 documentation. We aim to identify whether differences in performance exist between gestural interactions by using 3 different strategies: (1) a completely unmodeled approach, (2) the simple model introduced by Jude et al (Jude et al., 2014b) which models palm positions onto a hyperplane, and (3) a more complex model introduced here, which models the interaction space with no loss of information, using the angles of the forearm as input. We also aim to identify whether learning is present in these gestural interfaces through the use of a longitudinal design, as current works that use only one session have not found learning (Brown et al., 2014, Sambrooks and Wilkinson, 2013, Adhikarla et al., 2015, Adhikarla et al., 2015). For this study, the following two hypotheses were identified:

*H1* Users will learn gestural interaction over time, allowing for an improvement in performance.

*H2* An interaction with a model of the interaction space will perform better than an interaction which does not model the space.

The following metrics were collected to compare interactions: performance, accuracy, and subjective user feedback. Each of these metrics are explained within the following chapters.

# CHAPTER TWO

## Literature Review

### *2.1 Gestural Interaction*

Gestural Interaction is a technique that leverages gestures from the body to interact with a computer. This type of interaction technique has been studied for over 3 decades since Richard Bolt's first implementation in "Put-That-There" (Bolt, 1980). Gestural interaction implementations are typically divided into two different types: (1) those that require the user to wear gloves, devices or specific markers and (2) touchless gestural interaction. The latter leverages the "Come As You Are" design principle (Triesch and Von Der Malsburg, 1998), which states that users should not be required to wear devices or specific markers to interact with a system (Wachs et al., 2011).

Recent devices such as the Xbox Kinect, Leap Motion, and Myo Armband have gained popularity amongst researchers, with some capable of sub-millimeter accuracy in static situations (Guna et al., 2014). These devices have demonstrated the potential use of gestural interfaces in medical professions that require sterile environments (Wachs et al., 2011, Mentis et al., 2012, Bigdelou et al., 2012), as an accessibility device for those with impairments (Bailly et al., 2012, Guinness et al., 2014), and in mixed reality environments with head mounted displays (Ens et al., 2014). These applications demonstrate the usefulness of gestural interaction, but more work is still needed before its ubiquitous adoption. More specifically, though numerous implementations have been proposed for gestural cursor navigation, little work has been done to compare these methods against each other to discover improvements and which the methods work better, and why. In this work we will seek to close this gap by performing a larger investigation to uncover these answers.

### 2.1.1 Gestural Fatigue

A few years ago, fatigue was not given enough consideration in gestural interaction, and the standard method of controlling the computer via gestures involved users holding their arms up to the display for long periods of time. This method of interaction resulted in prolonged extension of the arm, which has since been known to cause a fatigue problem referred to as the "Gorilla Arm Syndrome" (Yoo et al., 2012, Carmody, 2010, Wachs et al., 2011). This fatigue problem has been described as a "known limitation" (Teixeira, 2011) of gestural interaction. Similarly, Segen and Kumar stated that fatigue is one of the main issues with gestures after prolonged interaction (Segen and Kumar, 2000).

After realizing the effects of fatigue on the interaction, Segen and Kumar conceptualized the idea that the interaction could be performed while the elbow was rested on a surface (Segen and Kumar, 2000). This simple solution was later incorporated in a Wizard-Of-Oz style experiment by Freeman et al, which allowed the user to appear as though they were interacting with the display while the elbow was rested (Freeman et al., 2012).

Brown et al implemented this method to perform cursor navigation in an experiment with 2 possible modes of input: the whole hand and finger pointing. Jude et al implemented a very similar approach to the previous in their Personal Space approach, whereby cursor navigation was likewise done with the whole hand, but the user's input space was first modeled during a calibration stage (Jude et al., 2014b) which can be seen in Figure 2.1. All these approaches make the same claim: allowing the users to perform gestural interaction from a rested position results in a more comfortable interaction and reduces fatigue inherent to gestural interaction. Correspondingly, gestural device manufacturers such as Leap Motion have taken notice of this research, and encouraged designers to use this rested elbow solution (Plemmons and Mandel, 2015).

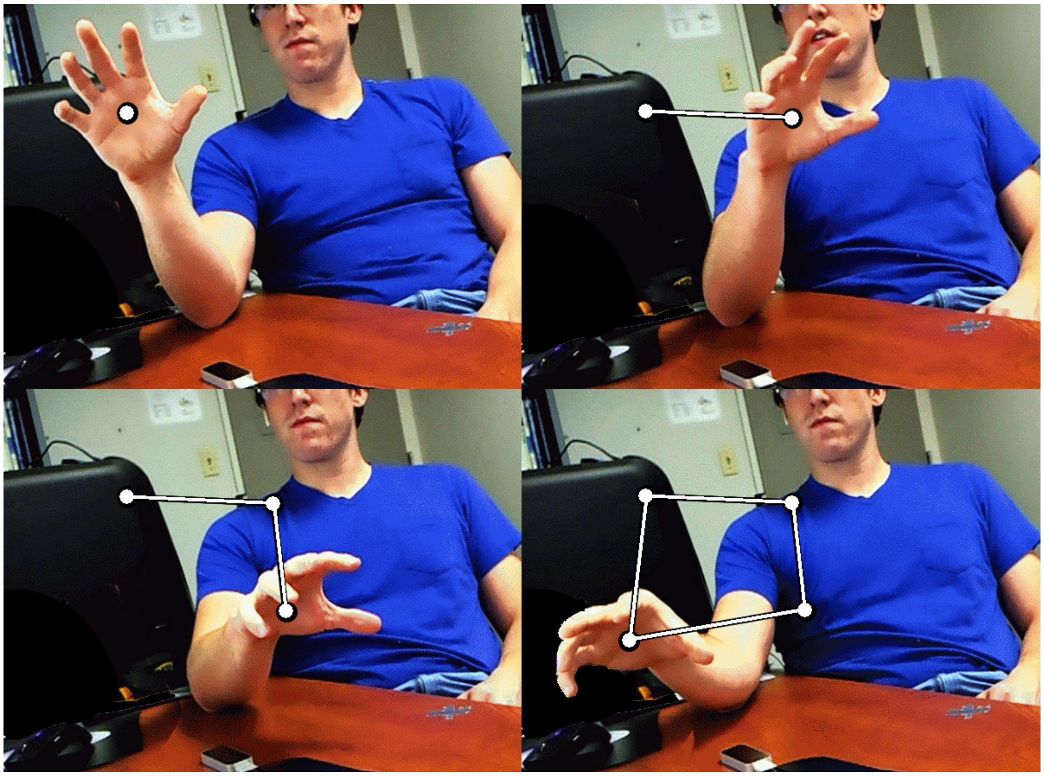


Figure 2.1. User calibrating their space taken from (Jude et al., 2014b) with permission. This method of calibration was used for both the *Hyperplanar* and *Spherical* models.

Conversely, 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 gestural interaction. This study also reported no improvements over the course of the experiment. We believe that this experiment found no performance improvements as it was counteracted by fatigue experienced by the participants.

In (Hincapié-Ramos et al., 2014), Ramos developed a novel metric *Consumed Endurance (CE)* to assess the level of fatigue felt during mid-air gestural interaction. In the study, the *CE* metric showed a strong correlation to the Borg CR10 scale and encouraged the use of this metric to evaluation fatigue. This is a very promising study, but currently has some limitations. The *CE* metric requires the use of a whole-body tracking system and that the arm is used in mid-air from a bent position. These limitations make it difficult to incorporate into our study as we are using the Leap Motion, which only tracks hand data, not the entire skeleton, and we allow the elbow to be rested, which is not expressed in the *CE* model.

We accept the premise of the research above and use an interaction which rests the elbow to reduce fatigue, and attempt to further the knowledge in the area by investigating the effects of modeled versus unmodeled interactions.

### *2.1.2 Gestural Pointing*

Recent gestural pointing implementations use one of two general pointing methods. “Ray pointing,” is a popular pointing method used by many designers when implementing gestural interaction (Brown et al., 2014, MacKenzie and Jusoh, 2001, Banerjee et al., 2011, Jota et al., 2010). Ray pointing uses ray casting to determine where the user is pointing (Jota et al., 2010). This is done by casting a ray from the pointing object and finding the intersection of the ray with the screen. This method has been demonstrated to be rapid but inaccurate (Cockburn et al., 2011).

The other popular gestural pointing method is ‘whole hand pointing’, which directly maps the 2D movements of the hand (by dropping 1 dimension usually depth) to the 2D cursor on screen (Brown et al., 2014). The user can then move their hand or finger within this navigation space to move the cursor on screen. ‘Whole hand pointing’ implemented by (Cockburn et al., 2011, Brown et al., 2014, Jude et al., 2014b), has been shown to be both rapid and accurate, even without visual feedback (Cockburn et al., 2011).

In this work we examine ‘whole hand pointing’ further using the *Unmodeled* and *Hyperplanar* approaches. The *Spherical* approach implements ’whole hand pointing’ without dropping the depth (Z) dimension.

### 2.1.3 Gestural Selection

Brown found that the finger tap gesture, which was considered the closest gesture to a mouse selection, performed inadequately for use in their experiment (Brown et al., 2014). To combat this, (Brown et al., 2014) implemented a bimanual selection method, where users used their other hand to press the space bar when the cursor was over the target. Other researchers (Hincapié-Ramos et al., 2014, Jude et al., 2014b, Sambrooks and Wilkinson, 2013) used a 'hover-select' or 'dwell' method which required the user to hover over the target for between 250 and 1500 milliseconds. This was the method we chose to implement in our own experiments.

## 2.2 Learning Effects

Schmidt and Lee indicated that we cannot directly observe learning. However, we can measure and report performance improvements, from which we can infer learning (Schmidt and Lee, 1988). New pointing devices are expected to demonstrate performance improvements over time, making a one day study less descriptive of the performance of the device. To account for these improvements, researchers have run longitudinal studies, performing analysis on the data when no performance improvements were found (MacKenzie et al., 2001). This approach was used in our experiment in line with H1 over 3 days, based on results from our pilot which showed no significant improvements after day 3.

A simple method to report performance improvements would be to calculate the difference between means of throughput between both rounds. Researchers have shown that a better way to measure performance increase is with effect size (Jude et al., 2014a) which is commonly measured using Cohen's $d$ (Cohen, 1992), a practice which has recently been encouraged for use in the HCI community (Kaptein and Robertson, 2012). This metric represents the difference of the mean between 2 groups over the pooled standard deviation, which can be seen in the following equation:

$$\mathrm{EffectSize} = \frac{\overline{\mathrm{x}_1} - \overline{\mathrm{x}_2}}{\mathrm{s}} \tag{2.1}$$

In our experiment, we measure the difference in performance between days to indicate performance gained.

Researchers have indicated that random-order practice, also known as Distributed Practice (e.g. A-B-C, B-C-A, C-A-B), generally benefits motor learning more than block order practice, known as massed practice (A-A-A, B-B-B, C-C-C) (Lin et al., 2007, Lee and Genovese, 1988). A 3x3x3 balanced Latin Cube design was used in our experiment, making the distributed practice approach trivial.

### *2.3 Pointing Device Evaluation*

In 1954 Paul Fitts held the belief that humans had a fixed information-transmission capacity, but warned that a person's motor system cannot be evaluated in isolation due to noise and the presence of other variables (Fitts, 1954). He however posited that if a subject had sufficient practice in a rapid motor condition to the point of overlearning, and the task's stimuli were held constant, the resulting performance would be limited by the demand from the condition on the person's channel capacity (Fitts, 1954).

To Human-Computer Interaction researchers, this meant that using this theory it would be possible to gauge a pointing interface's relative cognitive demand in comparison to other pointing devices, given a reasonable sample size. When one pointing condition obtained sufficiently better results than another, this meant that it presumably required less demand on the user's fixed information-transmission capacity in the motor task. Using this theory, Fitts designed an experiment and uncovered the relationship between movement time, and accuracy for persons in rapid motor tasks (Fitts, 1954).

Today this relationship is known as Fitts' law or Fitts's law, and is commonly used by HCI researchers to evaluate and identify improvements in pointing devices (Soukoreff and MacKenzie, 2004). If an pointing interaction conforms to Fitts' law, it can be used as a predictive model used in user interface design (Soukoreff and MacKenzie, 2004). In his experimentation Fitts proposed a new metric he denoted *Index of Performance*, which is now commonly referred to as *Throughput* which incorporated both speed and accuracy (Zhai, 2004, Soukoreff and MacKenzie, 2004). The definition of Throughput can be seen in Equation 2.2.

$$\text{Throughput} = \frac{\text{ID}}{\text{MT}} \tag{2.2}$$

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

$$\text{ID} = \log_2\left(\frac{2\text{D}}{\text{W}}\right) \tag{2.3}$$

Recent work has updated this definition to utilize the Shannon formulation of Index of Difficulty due to its higher correlation with movement time (Soukoreff and MacKenzie, 2004) which can be seen in Equation 2.4.

$$\text{ID} = \log_2\left(\frac{\text{D}}{\text{W}} + 1\right) \tag{2.4}$$

Because the cursor's final positions or end points, in relation to the target are not typically uniformly distributed, an *adjustment for accuracy* is performed by using *effective* *distance*, and *effective width* based on the post-hoc distributions of the observed distance and width parameters (Soukoreff and MacKenzie, 2004). The adjusted Index of Difficulty equation can be seen in Equation 2.5.

$$ID_e = log_2\left(\frac{D_e}{W_e} + 1\right) \tag{2.5}$$

We also adjust for accuracy using bivariate end point deviation ($SD_{x,y}$). We use this to compute bivariate throughput, which has been demonstrated to have higher explanatory power (Wobbrock et al., 2011). When the adjustments are made the definition of throughput that we use is defined by Equation 2.6.

$$\text{Throughput} = \frac{ID_e}{MT} \tag{2.6}$$

### *2.3.1 Standard Tasks*

In order to allow for researchers to compare their results with each other, and avoid confusion Soukoreff and MacKenzie encouraged the use of the ISO 9241-9 standard for evaluating pointing devices (Soukoreff and MacKenzie, 2004). After which Mackenzie, Wobbrock, and others have provided software suites which implement multiple tasks designed based of the ISO 9241-9 documentation (Wobbrock et al., 2008). In these suites, two main tasks are generally present, the first is the 1-dimensional vertical ribbons task design based upon Fitts's original Reciprocal Tapping experiment. The second is the "Ring-of-circles" task which allows for the evaluation in 2-dimensions and can be seen in Figure 2.2.

### *2.3.2 Fitts' Law*

To investigate whether an interface conforms to Fitts' Law, a linear regression is performed over movement times (MT) on the corresponding effective Index of

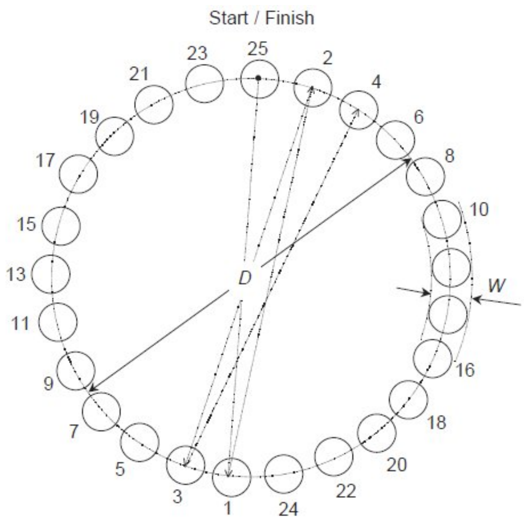



Figure 2.2. Multi-Directional task implementation from (Soukoreff and MacKenzie, 2004) used with permission

Difficulty conditions. The result is a regression equation which takes the form:

$$MT = a + b \times IDe \tag{2.7}$$

where $a$ is the intercept term, $b$ is the slope of the regression line and $IDe$ is the Index of Difficulty.

The intercept term represents the estimated time from the regression line, to navigate to a target when the ID = 0 (lowest difficulty). The intercept $a$ is desired to be near zero, and if positive is ideally below 400ms (Soukoreff and MacKenzie, 2004). Once the regression has been performed, the *coefficient of determination ($R^2$)*, or "goodness of fit", is typically reported as it signifies the strength of the association between the IDe and MT (Pino et al., 2013). $R^2$ is typically described as the degree to which the regression model explains the variation of the data, and the higher the $R^2$, the better the fit (MacKenzie and Teather, 2012). E.g., if $R^2 = .9$, this would mean that the regression line explains 90% of the variation.

The *linear correlation coefficient (R)* measures the strength and direction of the linear relationship between IDe and MT (Pino et al., 2013). When there is no linear correlation or a weak linear correlation, $R$ is close to 0. A correlation is described as strong if $R > 0.8$ and weak when $R < 0.5$ (Pino et al., 2013).

## *2.4 Accuracy Measures*

Apart from performance measured in throughput, we also measure the accuracy of each interaction. We use the accuracy measures introduced by Mackenzie et al which include Target Re-entry, Task Axis Crossing, Movement Direction Change, Orthogonal Direction Change, Movement Variability, Movement Error, and Movement Offset (MacKenzie et al., 2001). These measures were introduced to further characterize problems with pointing devices as previous literature would explain that there were differences in performance between two pointing devices, but previous metrics were not able to describe the behavior enough to identify why these differences occur (MacKenzie et al., 2001). A brief description of each metric and mathematical definitions from the original paper will be provided in this chapter, for further information see *Accuracy Measures for Evaluating Computer Pointing Devices* (MacKenzie et al., 2001).

### *2.4.1 Discrete Measures*

MacKenzie et al. introduced discrete accuracy measures to characterize discrete events such as a entry to a target, or a change in velocity. These new measures would be able to better describe performance differences between pointing devices. Lower values are considered better for each of these discrete measures, as larger values can signal issues with the pointing interface (MacKenzie et al., 2001).

*2.4.1.1 Target Re-entry.* A *Target Re-entry* is said to occur when the pointer enters the intended target, exits, and then re-enters the target (MacKenzie et al., 2001). This behavior is demonstrated in Figure 2.3. Mackenzie et al. further states that if this behavior is encountered 2 times within 10 trials, the metric is reported as 0.2 TRE per trial (MacKenzie et al., 2001).

Figure 2.3. An example of a Target Re-entry as seen in (MacKenzie et al., 2001) used with permission

*2.4.1.2 Task Axis Crossing*. In pointing tasks, the Task Axis which represents a straight line between the starting point of the cursor, and the center of the target (MacKenzie et al., 2001). This represents the optimal path of the cursor to which the actual path taken can be compared (Vatavu et al., 2013). When the cursor's position goes through the task axis during a trial a *Task Axis Crossing* is said to occur (MacKenzie et al., 2001). This behavior is described visually in Figure 2.4.

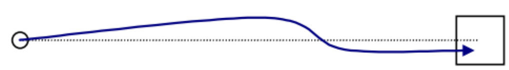


Figure 2.4. An example of a Target Axis Crossing as seen in (MacKenzie et al., 2001) used with permission

*2.4.1.3 Movement Direction Change*. A *Movement Direction Change* occurs when the cursor's path relative to the task axis changes (MacKenzie et al., 2001). This behavior is captured in Figure 2.5.

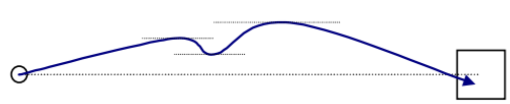


Figure 2.5. An example of three Movement Direction Changes as seen in (MacKenzie et al., 2001) used with permission

*2.4.1.4 Orthogonal Direction Change.* A *Orthogonal Direction Change* is said to occur when the cursor's direction changes along the axis perpendicular to the task axis (MacKenzie et al., 2001). If this measure is unusually high over a good amount of trials it can represent a control problem in the pointing interface (MacKenzie et al., 2001). An example of two Orthogonal Direction Changes is pictured in Figure 2.6.

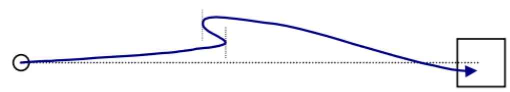


Figure 2.6. An example of two Orthogonal Direction Changes as seen in (MacKenzie et al., 2001) used with permission

### *2.4.2 Continuous Measures*

MacKenzie et al. also introduced continuous measures that could capture different aspects of the pointing behavior. For each of these continuous measures the ideal value is 0 (MacKenzie et al., 2001). Table 2.1 characterizes each continuous measure, and the relationships between them.

*2.4.2.1 Movement Variability (MV).* *Movement Variability (MV)* characterizes the extent to which the sampled screen coordinates from the pointing interaction lie on a straight line parallel to the task axis (MacKenzie et al., 2001). An ideal trial would obtain a result of $MV = 0$ (MacKenzie et al., 2001). Equation 2.8 depicts the mathematical definition of Movement Variability with the assumption that the task axis is drawn at $y = 0$.

$$MV = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} \tag{2.8}$$

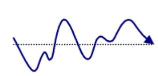
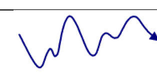
*2.4.2.2 Movement Error (ME).* *Movement Error* represents the mean deviation of the cursor's sampled path from the task axis, regardless of whether the points lie above or below the task axis (MacKenzie et al., 2001). Equation represents the definition of Movement Error when the task axis is lies on y=0.

$$ME = \frac{\sum |y_i|}{n} \tag{2.9}$$

*2.4.2.3 Movement Offset (MO).* *Movement Offset* represents the average deviation of the sample points to the task axis (MacKenzie et al., 2001). Equation 2.10 depicts the definition of MO when the task axis is y=0.

$$MO = \bar{y} \tag{2.10}$$

Table 2.1. The differences between actual paths(blue), and the optimal path (grey), and how these are characterized by each of the continuous accuracy metrics. Used from (MacKenzie et al., 2001), with permission.

| Metric |  |  |  |  |
|---|---|---|---|---|
| Movement Variability | Low | Low | High | High |
| Movement Error | Low | Very High | High | Very High |
| Movement offset | Low | High | Low | High |

# CHAPTER THREE

## Preliminary Investigations

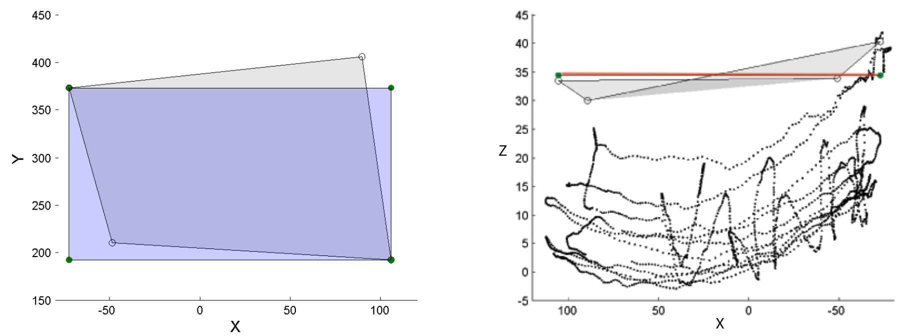
Due to the novelty of the area, it was identified early that extensive piloting was necessary to identify parameters of the experiment, and issues with each interface. We investigated learning effects, optimal target size and distance, fatigue, and performance through multiple pilots starting in July and ending in February. A brief summary of each pilot found to be important is provided in this chapter.

### *3.1 Pilot 1 : Foundations*

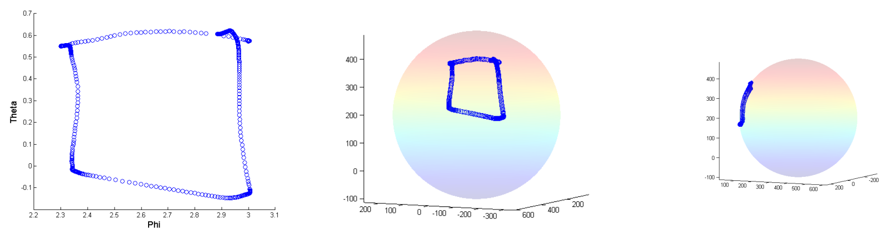
The first pilot can be considered a precursor to the full experiment. The pilot consisted of 5 persons who each participated over 5 consecutive days using 5 different interactions forming a Latin Cube design (5x5x5). The pilot was conducted to provide initial insight into learning effects, performance, and subjective user feedback using the each of the interactions. As naming conventions changed throughout the year, we provide all names for each interaction, but the decided names will be bolded.

#### *3.1.1 Pointing Interactions*

The study incorporated 5 different pointing interactions which are considered to be conditions in the study. The first interaction used was the *planar* or which is later referred to as the ***hyperplanar*** interaction, introduced by Jude et al. (Jude et al., 2014b) which allows the user to calibrate all 4 corners of the screen, and models the input space using a hyperplane containing the four points. An image of this interaction model is visualized in Figure 3.1a. The second interaction is known as the ***Spherical*** interaction, which also calibrates all 4 corners, but models the input space using a portion of a sphere who's' radius is defined by the length of the forearm. This interaction model can be seen in Figure 3.1b. In this pilot the *Spherical* interaction learned the approximate elbow position, and forearm's radius in the calibration step using a Least Squares sphereFit. The next interaction used was the standard whole hand pointing method used by (Brown et al., 2014, Cockburn et al., 2011) which was denoted the *de facto* interaction and was split up into rested (*de facto rested*) and unrested (*de facto unrested*) conditions. The *de facto* interaction were later renamed the ***unmodeled*** interaction, and the unrested condition was not examined in later work. A front and side view of this interaction model can be found in Figure 3.1a.




(a) Unmodeled(left = blue, right=red) and hyperplanar (grey) front and side views




(b) Raw calibration data superimposed onto a sphere

Figure 3.1. Gestural interaction models

### *3.1.2 Results*

Analysis of our results were done to analyze 4 main aspects of each interaction: (1) Daily Improvement, (2) Performance, (3) Accuracy, and (4) Subjective User Feedback. Metrics from (1), (2) and (3) were measured per trial across all participants per day. With 70 trials and 5 participants, we obtained 350 points per day per interaction which were used in our analysis. Significance was determined when $p < 0.05$.

Additionally, the analysis of the interaction's performance and accuracy were only done based on the last day of interaction. This is done in-line with existing literature where performance analysis is only performed when there is no longer any statistically significant learning effects (MacKenzie et al., 2001). A 5-way repeated measures ANOVA was constructed for each pointing condition $\times$ metric to determine if there was improvement in the pointing condition over time. Throughput was found to be significantly different in each gestural model over the 5 days. Post-hoc tests revealed that throughput on day 1 was significantly different than performance on days 3-5 in each gestural condition. Significant differences in accuracy were found until the very last day of the experiment, days 1-4 were therefore treated as practice. However, all five days worth of qualitative data was used to evaluate subjective user feedback.

*3.1.2.1 Performance.* We measure performance by throughput as per Equation 2.6 and illustrated in Figure 3.2 and Table A.1. A one-way ANOVA showed a statistically significant difference between interactions ($F(4, 1749) = 21.18, p < .001$). As the ANOVA showed statistical significance, a Least Significant Difference (LSD) was used post-hoc to identify which interactions differ. This test showed that the touchpad's throughput was significantly higher than all gestural models ($p <$*.001*). No significant differences were found between each gestural models.

*3.1.2.2 Improvement.* We measure Improvement as the increase in throughput between day 1 and day 5, and reported with the effect size metric Cohen's *d*.

In general, an effect size of 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). From Table 3.1 we can see that most gestural interaction have a medium

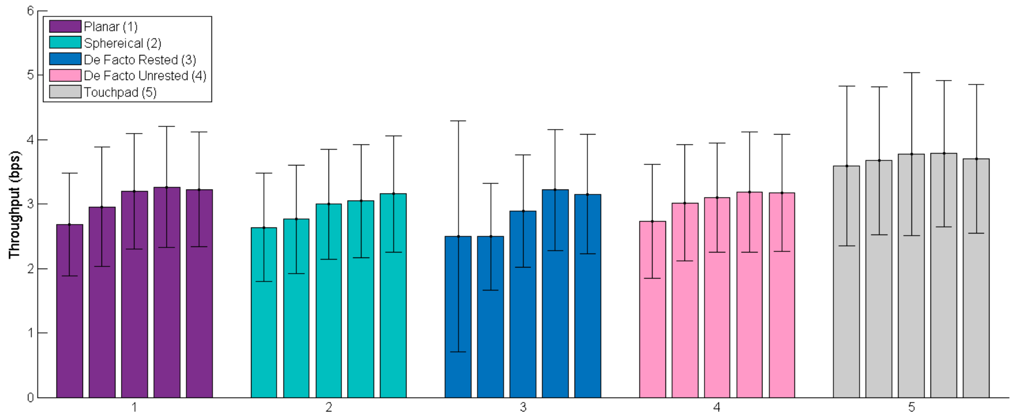



Figure 3.2. Throughput means and standard deviation from days 1-5 grouped by interaction

effect in throughput, which indicates overall performance improvement, from which which we infer that learning has occurred. The touchpad on the other hand had no practically significant increase in performance.

Table 3.1. Performance improvement between days 1-5, reported in Cohen's *d*.

| Source | Cohen's *d* |
|---|---|
| Planar | 0.64 |
| Spherical | 0.60 |
| De Facto Rested | 0.46 |
| De Facto Unrested | 0.50 |
| Touchpad | 0.09 |

*3.1.2.3 Accuracy.* Accuracy metrics were collected which included the previously described metrics: Movement Error (ME), Movement Offset (MO), Movment Variability (MV), and Target Reentry (TRE). Table A.2 depicts the accuracy results between each interaction.

Overall the touchpad was found to be significantly more accurate than all other interactions in terms of Movement Error. No significant differences were present in the Movement Offset results. The touchpad was found to have significantly better

Movement Variability than the *de facto rested* interaction. Once again the touchpad was also found to have significantly better results in terms of Target Reentry than the *de facto rested*, and *de facto unrested* interactions. The *Planar* interaction was found to have significantly better Target Reentry performance than the *de facto rested* and *de facto unrested* interactions.

*3.1.2.4 Arm-length approximation.* Since the *Spherical* interaction required the measurement of the forearm, and elbow position, we wanted to examine how close the approximation from the SphereFit was to the actual measurements. We compared the correctness of our arm-length approximation over each of the 5 consecutive days to the actual length of the arm measured on the first day. The results are measured in percentage of overall difference and is shown in Table 3.2. We see that the values ranged from 75% to 91%, with a trend that shows an overall increase in accuracy over time.

Table 3.2. Arm length approximation correctness (%)

| Metric | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 |
|---|---|---|---|---|---|
| Approx % | 75.40 | 86.80 | 79.71 | 89.24 | 91.51 |

### *3.1.3 Overview*

In this small study we found that none of the gestural models on day 5, obtained significantly better throughput, meaning that resting the elbow results in no significant effect on throughput. However, modeling the input space using a hyperplane space improves the single accuracy metric Target Reentry over the Unmodeled (*de facto*) approaches. We also found that performance improvement as measured by throughput is not significantly different after day 3. This finding was incorporated into later studies.

A more expanded version of the study was submitted to the 2015 SIGCHI conference. Although the study was not accepted, the reviewers provided critical feedback which was directly used to construct the full experiment. The most prominent considerations were to increase the number of participants, use the ISO 9241-9 task which is considered the standard task for pointing device evaluation, evaluate only gestural devices without a benchmark, and to include a published work that had been presented that year.

We took these critiques into consideration and began to design a more focused experiment. We concluded that in addition to the reviewers concerns, our evaluation could remove other nuisance variables such as the differing calibration regions between each gestural interaction, and reduce the duration of the calibration by removing the learning of the forearm and elbow position, and rather measure the user's arm radius and elbow's position.

### *3.2 Pilot 2: New Setup*

To reduce nuisance variables, a one shot calibration was designed and implemented which would save the original source points captured from the Leap Motion Controller into a matrix. With these source points we could then toggle the different gestural interaction models over-top which would use either two points to construct the space in the *unmodeled* interaction, or use all four points, which was used in the *Hyperplanar* and *Spherical* interactions. The toggle switch of the gestural interactions was enabled through a key binding (CTRL-ENTR) so users were not able to explicitly identify that they were using different interactions.

To allow for this one shot calibration, we needed to measure the user's forearm length, and the relative position of the elbow to the Leap Motion Controller. This was taken to be the sphere's radius, and center point respectively in the interaction. These measurements were collected using a metric tape measure in millimeters to reflect the Leap Motion Controller's setup. The arm was measured as seen in Figure 3.3 in

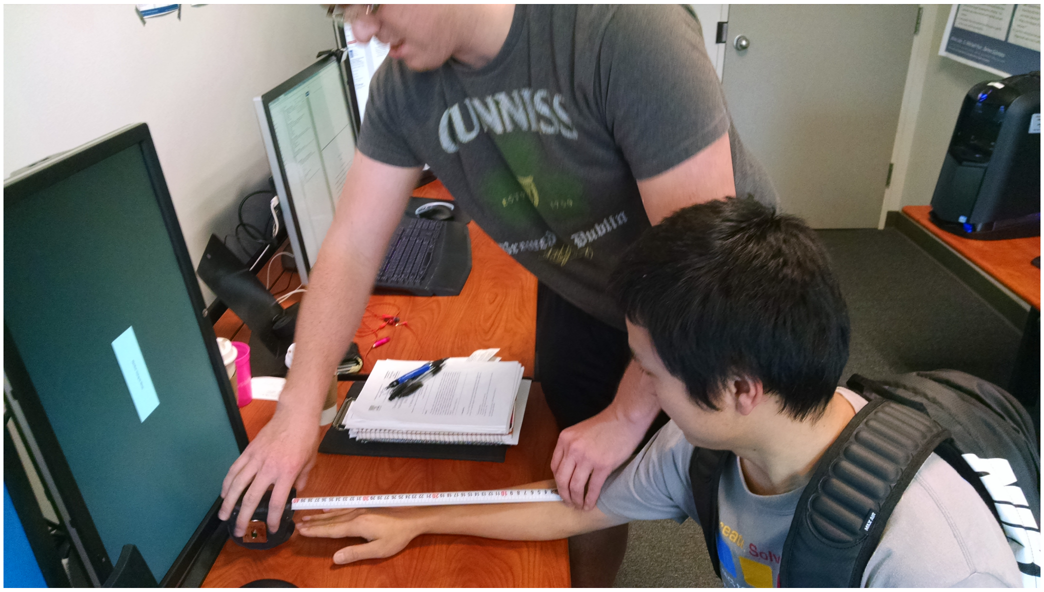


Figure 3.3. User's forearm is being measured in the pronated position with the palm down

a *pronated palm down* position after consulting a medical expert about biological tissues such as fats, and muscles impacting the measurements.

In our previous studies, and pilots we also discovered that simply resting the elbow on the table posed some strain on the user during interaction (Guinness et al., 2014). Due to this finding we also chose to incorporate an ergonomic elbow rest which can be seen in Figure 3.4.

Additionally, we switched from our moving squares task used in our previous studies (Jude et al., 2014b, Jude et al., 2014a, Guinness et al., 2014) to the ISO 9241-9 ring of circles task, which is the current standard for pointing device evaluations and is described further in the previous chapter. In order to use the new task a selection method needed to be chosen. To stay in line with out previous work we decided to enact a hover/dwell selection, but reduced the dwell time to 250 milliseconds as our reviewers from the previous pilot submission thought that the previous 500 millisecond duration was too long.

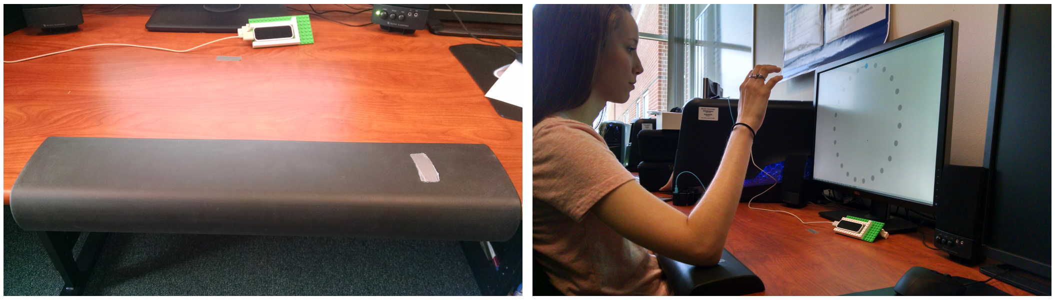


Figure 3.4. Ergonomic Rest used to reduce strain on the elbow and forearm

Due to the above changes we needed to run another preliminary investigation to identify suitable experiment parameters, and identify issues in our new setup. So we decided to focus on user feedback and finding suitable Amplitude/Width parameters which were needed for the new ISO 9241-9 task. At this point version 2.2.5 of the Leap Motion SDK was released which provided positional data about the wrist. We experimented with this tracking because the wrist offers a more stable reference point in our interaction than the hand, as the hand can bend, and roll which we do not wish to encourage due to ergonomic concerns.

Using 5 participants we demonstrated that the new wrist tracking did not improve performance over hand tracking. Upon further investigation using the provided Diagnostic Visualizer, the wrist tracking did not appear to actually be tracking the wrist, but building a wrist model from the hand with pre-programmed values. This was discovered because in instances where the hand was not able to be seen, or had become hidden, the wrist model would deform in the visualizer. The model would then only become stable once the hand was seen, even though the wrist was clearly positioned for the tracking.

In addition, we identified suitable parameters for the ISO task to use target amplitudes {256, 512, 1024, 1408} and target widths {64, 96, 128}. Using these values we were able to provide 10 unique Index of Difficulty (ID) values ranging from 1.52-4.58 bits which is in line with current literature's encouragements (Soukoreff

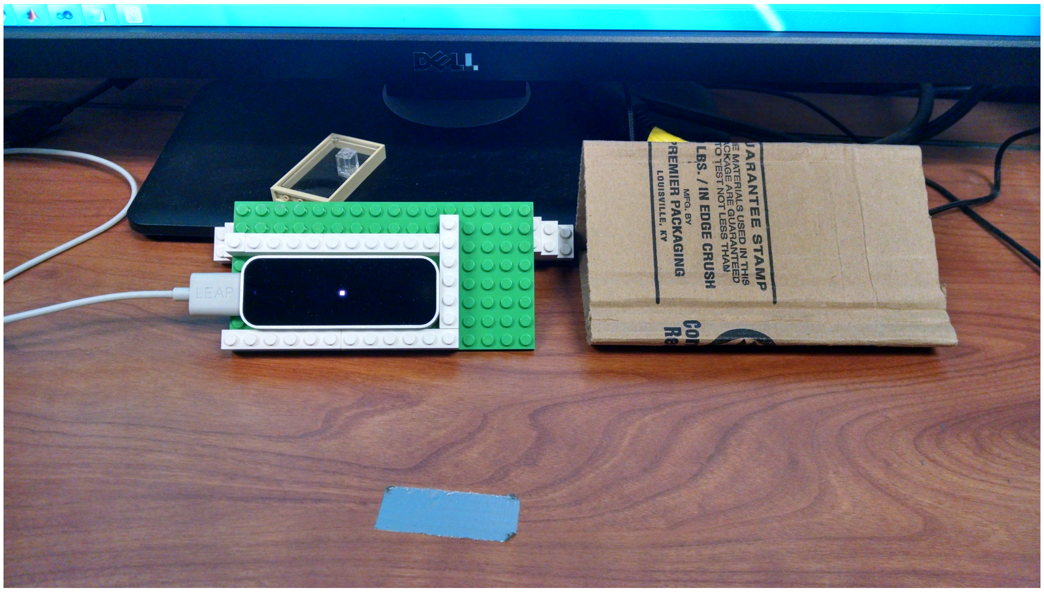



Figure 3.5. The stand used in the experiment (left), and the one used in the pilot (right)

and MacKenzie, 2004). During this pilot we also found that many of the participants accidentally knocked the Leap Motion Controller off of it's angled stand. We decided to construct 3 new angled stands; one of which can be seen in Figure 3.5 using Lego blocks to keep the Leap Motion Controller tightly in place. We also wanted to make sure that our performance improvements over sessions were present in our new setup. Using one participant and 5 consecutive sessions, we saw that performance did improve significantly between days, but that the majority occurred within the first 3 days.

### *3.3 Pilot* Summary

From the first pilot we found that learning was present in gestural interfaces in the form of performance improvements between sessions. This finding is contradicted by current literature that has only been evaluated in one session (Brown et al., 2014, Sambrooks and Wilkinson, 2013, Adhikarla et al., 2015). Given the current literature's state we consider that these learning effects should be reported in our next study. This meant that we would have to incorporate multiple consecutive experiments using a longitudinal design in the full study. Since we did not find many improvements after day 3, we chose to use 3 consecutive sessions to be able to identify learning effects, but also minimize the cost of the experiment. We also identified that the target amplitude parameters {256, 512, 1024, 1408} and target widths {64, 96, 128} were suitable for user's to obtain and for our Fitts' regression models to achieve a reasonable ID range to be used in the regression.

# CHAPTER FOUR

## Methods

### 4.1 Interaction Design

The interactions in this experiment were used specifically to test the hypotheses of this research. Therefore, we used an unmodeled interaction described by Brown et al (Brown et al., 2014), a simple model introduced by Jude et al (Jude et al., 2014b), and a more complex model which we introduced for this experiment. We also built 3 angled stands $(30, 36, 44°)$ with Legos, to hold the Leap on a tilted incline as recommended by the previous authors (Jude et al., 2014b).

#### 4.1.1 Unmodeled Gestural Interaction

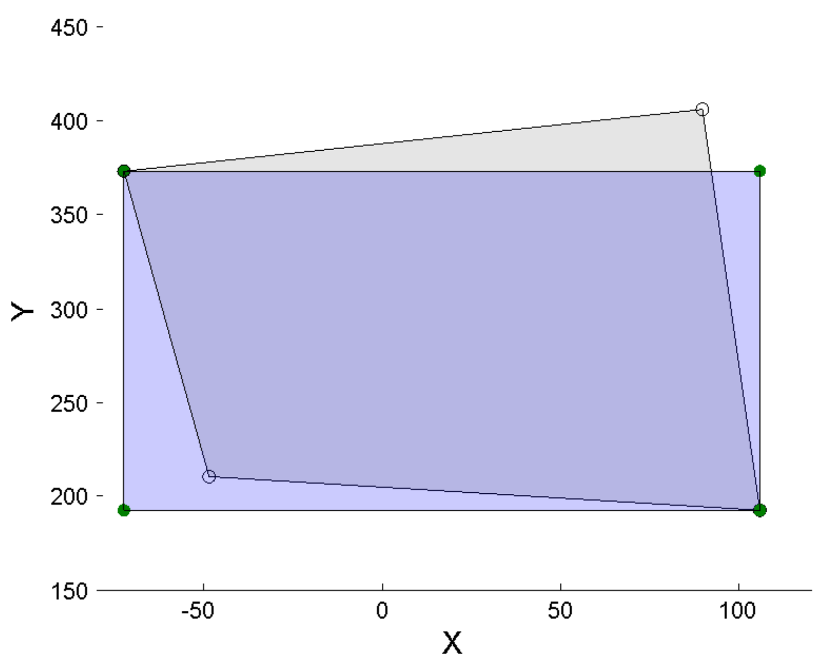



Figure 4.1. A front view of the Unmodeled (blue plane) and Hyperplanar (grey plane) interaction space. Both axis denotes interaction space measured in millimeters.

In this interaction, the space is constructed using two diagonal points from the calibration as shown in Figure 4.1. This model was replicated from (Brown et al., 2014), which drops the Z dimension and only uses X and Y for input into the source matrix. The points were chosen such that each screen boundary could be obtained without lifting the elbows from the table. We identify this approach as the *Unmodeled* interaction.

### 4.1.2 Rested & Calibrated Interaction – Hyperplanar Model

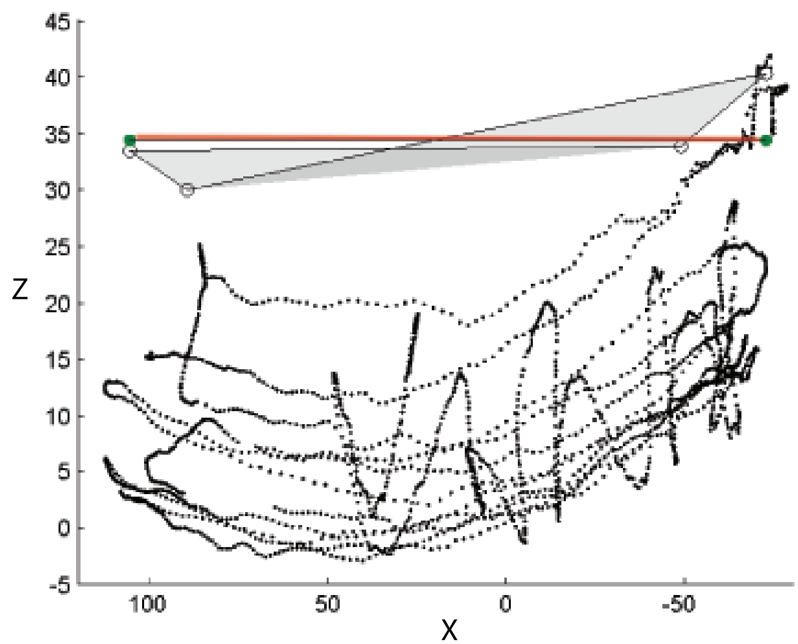



Figure 4.2. A top view of the Unmodeled (red plane) and Hyperplanar (grey hyperplane) with tracked hand movement (black points). Both axes denote a dimension of the interaction space measured in millimeters.

The intuition behind this interaction is that the unmodeled interaction creates a space that does not map well to the interaction space. The example in figure 4.1 shows how it could be difficult for the user to hit the bottom left of the interaction space. By first modeling the interaction space, this interaction allows the users to be able to easily reach all corners of the screen from a rested position.

This interaction first constructed by Jude et al. and was referred to as a "planar" interaction (Jude et al., 2014b). However, we believe the term "hyperplanar" is more descriptive as as there is no guarantee that the 4 calibrated points in 3D lie on the same plane. This is better illustrated in figure 4.2.

We also note that despite the *Hyperplanar* approach being closer to the interaction space, it is not as accurate as it could be. The black dots in figure 4.2 show the actual interaction, which indicates some information loss despite all 3 dimensions (X, Y, Z) in cartessian space used to build the model. The original author's description of this model is given in Appendix B.

#### *4.1.3 Rested & Calibrated Interaction – Spherical Model*

The *Spherical* model introduced here is based on two intuitions. The first is that the movement of the hand from a rested position forms a part of a sphere, as shown in Figure 4.3. And second, that controlling an inherently 2-dimensional interface such as a monitor will be easier if the input itself is based on 2 dimensions. We built the *Spherical* model on both of these intuitions. The interaction itself uses a spherical coordinate system with the azimuth ($\theta$), mapped to medial and lateral shoulder rotation, and zenith ($\phi$), mapped to elbow flexion and extension (McLester and Pierre, 2007), of the forearm as input, making it 2-dimensional. This translation equates to a feature reduction from 3 features (X, Y, Z) to 2 ($\theta, \phi$) and a constant radius ($r$) with no loss of information. In contrast, other interactions that perform a reduction in dimensions generally do so by eliminating one dimension, generally Z or depth.

The feature reduction from 3 free features to 2 effectively removes one linearly independent column from the source and destination matrix representations. Losing this column causes a loss of precision as each matrix is now further away from full

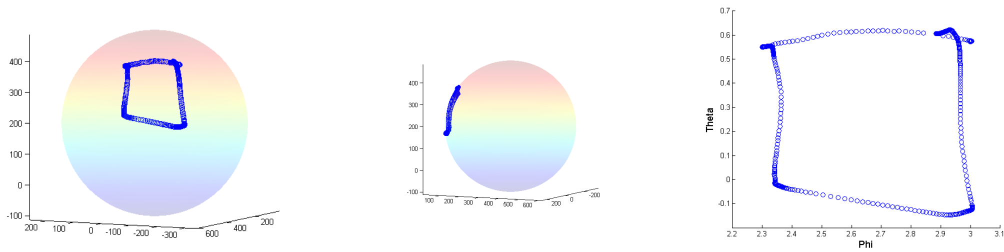



Figure 4.3. L-R: (1) Front view of *Spherical model*, (2) side view of the *Spherical* interaction, (3) plotting the angles $(\theta, \phi)$ which is essentially a quadrangle.

rank (Nagy et al., 1996) when compared to the *Hyperplanar* implementation. To account for this we incorporate a plane-to-plane homography, otherwise known as a projective transformation with homogeneous estimation, which projects a 3rd dimension into a 2D image. We use this 3rd dimension to preserve rank during the transformation. The transformation also provides both determined and overly-determined solutions with a bounded error (Criminisi et al., 1999).

Although the input required is the azimuth $(\theta)$ and zenith $(\phi)$ of the forearm, these values were not directly obtainable from our input device, the Leap Motion controller. We did not consider using a different device that did have these values as it would be an unfair experiment, where there would be interaction with the input device. We therefore used the provided input by the Leap Motion controller, which is the X, Y, and Z positions of the palm in Cartesian coordinates, and transformed it to corresponding $\theta$, $\phi$ and a static $r$. We measured the length of participants' forearms as radius $r$. The center of the interaction sphere is fixed and marked on the table surface, and the users are expected to position their elbow on this exact point throughout the experiment. Given these inputs, we were able to translate the coordinates in X, Y and Z to $\theta$ and $\phi$. We then used $\theta$ and $\phi$ as input and the screen coordinates $x$ and $y$ as the intended output. We observed that using this coordinate system produces a model that closer represents the user's input, including being able to account for the curvature in the input, which was not achievable in the *Hyperplanar* model.

## *4.2 Experimental Design*

### *4.2.1 Participants*

15 participants (M=9, F=6) between 19-27 years of age (mean = 20.6) took part in the experiment. All participants were students from a local university. Two participants self-reported as ambidextrous, but all elected to use their right hand for all interactions. All but 3 participants had used gestural interfaces before. Previously used gestural interfaces were limited to the Wii, Kinect, and/or PlayStation Move. Participants were compensated for their time.

### *4.2.2 Apparatus*

All testing was conducted in a lab setting on a 30-inch Dell monitor set to 2560 x 1600 resolution. The Leap Motion Controller was used to recognize the hand position for the gestural navigation. The computer used an Intel i7-3820 CPU with 8 cores clocked at 3.6 GHz, with 32 GB RAM and ran Windows 7. The 36 degree stand was used in all cases except for 2 participants with longer forearms, for which the 30 degree stand was chosen.

### *4.2.3 Task*

The ISO 9241-9 ring-of-circles task implementation from (Wobbrock et al., 2011) was used to evaluate the performance of each interaction. This software was modified such that a 250 millisecond hover is used for selection. The task utilized 4 amplitudes {256, 512, 1024, 1408} and 3 target widths {64, 96, 128 } for 10 unique IDs ranging from 1.52-4.58 bits. Target amplitude and widths were identified from current literature (Brown et al., 2014) and extensive piloting. As we were using a large display, piloting revealed that a target width of 64px was the smallest target that was able to be selected by participants. The first 3 trials of each condition were taken as practice since this was the default in the software used.

#### *4.2.4 Design*

We used a 3x3x3 balanced Latin cube design with 3 interaction styles over 3 days and 3 orders. Each day included 1 session which lasted roughly 1 hour and was split into 2 rounds. In each round, participants would use all of the three interactions based on the Latin cube ordering. Participants were not told that they were using different interaction models.

#### *4.2.5 Procedure*

Participants were required to watch a video detailing the interaction and calibration method before they began trials on the first day. After the video, the calibration stage would begin. Once calibrated, participants were asked to test the interaction. A recalibration was allowed until they were pleased with the interaction. After which, participants were asked to watch a video detailing the ring-of-circles task. They then performed the task using the three gestural models. Participants were encouraged to take notes on the interaction they just used after each task, for ranking purposes. This was repeated until all 6 tasks (3 interactions $\times$ 2 rounds) were evaluated. Participants were then asked to rank the interactions from best to worst. These steps were repeated exactly every day of the experiment, with the videos only shown on day 1.

Only 1 calibration (see Figure 2.1) was performed each day at the beginning of the experiment to control for differing calibrations. Each model was then dynamically computed from the original source input points. All gestural interactions were performed with an off-the-shelf Leap Motion controller.

# CHAPTER FIVE

## Results

Analysis of our results was done to investigate 5 main aspects of each interaction: (1) Daily Improvement, (2) Conformance to the Fitts' Model, (3) Performance, (4) Accuracy, and (5) Subjective User Feedback. Metrics from (2), (3), and (4) were measured per trial across all participants per day. In each ring-of-circles task there were 23 trials (3 practice) in each of the 4 amplitude $\times$ 3 width conditions, which meant that there was a total of $20 \times 4 \times 3 = 240$ trials per task per participant. We incorporated 2 rounds with 15 participants, for a total of $240 \times 2$ rounds $\times 15$ participants $= 7200$ trials per interaction per day. In each analysis, the assumption of Sphericity was violated, thus a Greenhouse-Geisser ($p_{GG}$) epsilon correction was used to determine significance. Post-hoc tests were administered if the epsilon corrected p-value was less than 0.05 ($p_{GG} < .05$) using the MATLAB 'Bonferroni Method', which uses critical values from the t-distribution after an adjustment for multiple comparisons is made. We report the effect size of the Repeated Measures ANOVA as $\eta_p^2$ which is interpreted (0.01 = small, 0.06 = medium, 0.14 = large) as determined by Cohen (Cohen, 1988, Richardson, 2011). In this design $\eta_p^2$ is equivalent to $\eta_G^2$ since each regression only considers one manipulated factor and the participants (Bakeman, 2005). We report pairwise effect size, as measured by Cohen's $d$ and fall back on Cohen's own guidelines for practical significance (0.2 = small, 0.5 = medium, 0.8 = large), as there are no domain-specific guidelines for pointing device evaluation.

### *5.1 Daily Improvement*

We measure daily improvement as a difference in bivariate throughput between days in order to check for learning effects. These values are shown in Table 5.1, and all interactions were found to have statistical difference between days.

Table 5.1. Mean ($\bar{x}$) and standard deviation ($\sigma$) of bivariate throughput for each device across days 1-3. The corresponding $d$ values indicate difference in performance from the previous day measured with Cohen's d.

| | Unmodeled | | | Hyperplanar | | | Spherical | | |
|---|---|---|---|---|---|---|---|---|---|
| Day | $\bar{x}$ | $\sigma$ | $d$ | $\bar{x}$ | $\sigma$ | $d$ | $\bar{x}$ | $\sigma$ | $d$ |
| 1 | 2.35 | .38 | - | 2.44 | .43 | - | 2.32 | .40 | - |
| 2 | 2.69 | .48 | .77 | 2.74 | .40 | .74 | 2.61 | .39 | .74 |
| 3 | 2.79 | .44 | .23 | 2.82 | .41 | .18 | 2.67 | .38 | .15 |

The *Unmodeled* interaction showed a significant difference in performance between days ($F(2, 14370) = 3573$, $p_{GG} < .001$, $\eta_p^2 = 0.33$), as did the *Hyperplanar* interaction ($F(2, 14370) = 2786$, $p_{GG} < .001$, $\eta_p^2 = 0.28$) and the *Spherical* interaction ($F(2, 14370) = 2437$, $p_{GG} < .001$, $\eta_p^2 = 0.25$). The post-hoc test showed all interactions demonstrated significance between all days ($p < .001$).

Due to these differences, we perform the full analysis of the interactions for performance, accuracy, and subjective feedback on data from day 3 only, while days 1 and 2 are considered practice. Therefore, all metrics reported in this chapter are from day 3 of the experiment, unless stated otherwise.

## *5.2 Fitts' Regression*

Table 5.2. Metrics for Fitts' regression of each interaction.

| Metric | Unmodeled | Hyperplanar | Spherical |
|---|---|---|---|
| Intercept | 585.9 | 79.9 | 16.3 |
| Slope | 155.5 | 295.7 | 331.5 |
| $R^2$ | 0.018 | 0.738 | 0.758 |
| $R$ | 0.418 | 0.859 | 0.870 |

A Fitts' law model for each interaction was built by regressing the mean movement times (MT) on the corresponding effective Index of Difficulty conditions. The result is a regression equation in the form of Equation 2.7.

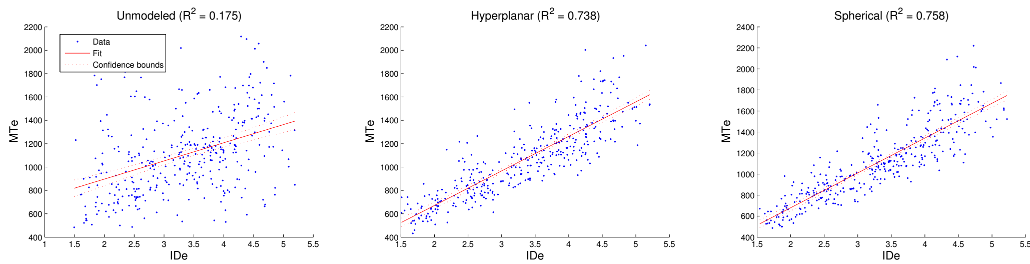



Figure 5.1. Mean movement time of all participants from day 3 as a function of effective Index of Difficulty (IDe), with computed Fitts' regression lines

The results for each day and interaction are shown in Table 5.2, and in Figures A.1, A.2, A.3, while a visualization of the data from day 3 can be seen in figure 5.1. The plots show that the modeled approaches are better explained by Fitts' law than the unmodeled approach. Additionally, Table A.3 shows that the two modeled approaches consistently improve their fit between days, while the unmodeled approach does not.

## *5.3 Performance*

The speed (specifically the Movement Time) of each interaction is considered a naive metric (Soukoreff and MacKenzie, 2004), but is reported nonetheless as we consider it a good description of each interaction. A better metric to use is throughput, and specifically bivariate throughput (Wobbrock et al., 2011).

### *5.3.1 Movement Time*

Table 5.3. Mean and standard deviation of movement time per trial for each interaction in milliseconds. Smaller values are better.

| | Unmodeled | | Hyperplanar | | Spherical | |
|---|---|---|---|---|---|---|
| Day | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| 3 | **1073.6** | 495.39 | **1063.4** | 481.89 | 1110.4 | 538.24 |

A repeated measures ANOVA showed a significant difference in movement time, which was adjusted for dwell time, between interactions ($F(2, 14370) = 18.619$,

$p_{GG} < .001$, $\eta_p^2 = 0.003$). The post-hoc test found that the *Unmodeled* interaction had significantly lower movement time than the *Spherical* interaction ($p <.001$, $d = 0.07$). The post-hoc test also showed that *Hyperplanar* interaction had significantly lower movement time than the *Spherical* ($p <.001$ $d = 0.09$).

### 5.3.2 Bivariate Throughput

A repeated measures ANOVA showed a significant difference in performance between interactions measured with bivariate throughput ($F(2, 14370) = 369.81$, $p_{GG} < .001$, $\eta_p^2 = 0.05$). A post-hoc test showed the *Hyperplanar* interaction had higher bivariate throughput than both the *Unmodeled* ($p <.005$, $d = .05$) and *Spherical* ($p <.001$, $d = 0.35$) interactions. The post-hoc test also showed that the *Unmodeled* interaction had higher bivariate throughput than the *Spherical* interaction ($p <.001$, $d = 0.29$).

## 5.4 Accuracy Measures

We report accuracy measures based on the metrics introduced by Mackenzie (MacKenzie et al., 2001) consisting of Target Re-entry (TRE), Target Axis Crossing (TAC), Orthogonal Direction Change (ODC), Movement Variability (MV), Movement Error (ME) and Movement Offset (MO) seen in Table 5.4. These measures were taken directly from the *FittsStudy* software (Wobbrock et al., 2011), except for TRE, as the software reports target entries (TE) instead, where TRE=TE-1. A lower number is better for all metrics except MO, where a closer distance to 0 is better.

### 5.4.1 Target Entries

The *Unmodeled* interaction had the best TE score, followed by the *Hyperplanar* interaction. A repeated measures ANOVA showed a significant difference in the number of Target Entries between interactions ($F(2, 14370) = 27.20$, $p_{GG} < .001$, $\eta_p^2 = 0.004$). The post-hoc test identified a statistical significance between the *Unmodeled* and *Spherical* interactions ($p <.001$, $d = 0.11$), and between the *Hyperplanar* and *Spherical* interactions ($p <.001$, $d = 0.10$).

Table 5.4. Mean and standard deviation of accuracy measures metrics taken on day 3 of the experiment. Bold text denotes a result of statistical significance

| | Unmodeled | | Hyperplanar | | Spherical | |
|---|---|---|---|---|---|---|
| Metric | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| TE | **1.27** | 0.68 | **1.28** | 0.70 | 1.35 | 0.78 |
| TAC | **1.45** | 1.20 | **1.49** | 1.17 | 1.55 | 1.25 |
| MDC | **2.77** | 1.69 | **2.78** | 1.66 | 2.87 | 1.79 |
| ODC | **1.08** | 1.30 | **1.05** | 1.25 | 1.19 | 1.38 |
| MV | 23.50 | 20.73 | **22.94** | 18.85 | 23.83 | 20.06 |
| ME | 26.70 | 20.42 | **25.72** | 18.67 | 26.42 | 19.26 |
| MO | -7.81 | 28.32 | 4.51 | 26.57 | **2.04** | 27.25 |

### *5.4.2 Task Axis Crosses*

The *Unmodeled* interaction had the best TAC score, followed by the *Hyperplanar* interaction. A repeated measures ANOVA showed this to be statistically significant ($F(2, 14370) = 11.054$, $p_{GG} < .001$, $\eta_p^2 = 0.002$). The post-hoc test identified a significant difference between the *Unmodeled* and *Spherical* interactions ($p <.001$, $d$ =0.08). The *Hyperplanar* interaction was also found to be significantly different from the *Spherical* interaction ($p <.05$, $d$ = 0.04).

### *5.4.3 Movement Direction Change*

The *Unmodeled* interaction had the best MDC score, followed by the *Hyperplanar* interaction. A repeated measures ANOVA showed this to be statistically significant ($F(2, 14370) = 8.262$, $p_{GG} < .001$, $\eta_p^2 = 0.001$). Post-hoc tests showed a significant difference between the *Unmodeled* and *Spherical* interactions ($p <.005$, $d$ = 0.06). The post-hoc test also identified a significant difference between the *Hyperplanar* and *Spherical* interactions ($p <.005$, $d$ =0.06).

### *5.4.4 Orthogonal Direction Change*

The *Hyperplanar* interaction had the best ODC score, followed by the *Unmodeled* interaction. A repeated measures ANOVA showed this to be statistically significant ($F(2, 14370) = 23.786$, $p_{GG} < .001$, $\eta_p^2 = 0.003$). A post-hoc test showed a significant difference between the *Unmodeled* and *Spherical* interactions ($p <$*.001*, $d$ = 0.08), and between the *Hyperplanar* and *Spherical* interactions ($p <$*.001*, $d$ =0.10).

### 5.4.5 Movement Variability

The *Hyperplanar* interaction had the best MV score, followed by the *Unmodeled* interaction. A repeated measures ANOVA showed a significant difference between interactions ($F(2, 14370) = 3.918$, $p_{GG} < .05$, $\eta_p^2 = 0.0005$). A post-hoc tests showed a significant difference between the *Hyperplanar* and *Spherical* interactions ($p <$*.05*, $d$ = 0.05).

### 5.4.6 Movement Error

The *Hyperplanar* interaction had the best ME score, followed by the *Spherical* interaction. A repeated measures ANOVA showed this to be statistically significant ($F(2, 14370) = 5.161$, $p_{GG} < .01$, $\eta_p^2 = 0.0007$). The post-hoc test identified a significant difference between the *Hyperplanar* and *Unmodeled* interactions ($p <$*.005*, $d$ = 0.05).

### 5.4.7 Movement Offset

The *Spherical* interaction had the best MO score, followed by the *Hyperplanar* interaction. A repeated measures ANOVA showed this to be statistically significant ($F(2, 14370) = 424.97$, $p_{GG} < .001$, $\eta_p^2 = 0.06$). The post-hoc test identified a significant difference between the *Spherical* and *Hyperplanar* interactions ($p <$*.001*, $d$ =0.09), and between the *Spherical* and *Unmodeled* interactions ($p <$*.001*, $d$= 0.36). The post-hoc test also showed that the *Hyperplanar* interaction was significantly different than the *Unmodeled* interaction ($p <$*.001*, $d$ =0.45).

### 5.5.1 Preference

Participants were asked to rank the interactions by preference on a scale of 1-6, with 1 being the best interaction and 6 being the worst. We combined the rounds that used the same interaction. From these rankings we can order the interactions by most preferred to least as such: 1) *Hyperplanar*, 2) *Unmodeled*, 3) *Spherical*, with their respective scores of 2.57, 3.60, 4.33. A Friedman test showed that there was a statistically significant difference in preference rank between the interactions ($\chi^2(2) = 13.505, p = 0.0012$). Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied. The only statistically significance found was that the *Hyperplanar* interaction was significantly preferred over the *Spherical* ($Z = -3.524, p < .001$, $r = .23$).

### 5.5.2 Usability

Table 5.5. Mean reported usability & comfort metrics per interaction (1 = Most Negative, 5 = Most Positive).

| Metric | Unmodeled | Hyperplanar | Spherical |
|---|---|---|---|
| Operation smoothness | 4.13 | 4.27 | 3.97 |
| Operational effort | 4.20 | 4.30 | 4.03 |
| Accuracy | 3.82 | 4.27 | 3.80 |
| Operation speed | 4.27 | 4.33 | 4.23 |
| General comfort | 4.27 | 4.40 | 4.07 |
| Overall operation | 4.17 | 4.33 | 4.03 |

Participants were asked to fill out 5-point independent rating Likert scale questions from Annex C. of the ISO 9241-9 which evaluated usability and comfort of the interaction immediately after using the interaction. Table 5.5 depicts the questions and the mean reported ratings about the usability the interaction. The reported ratings show reasonably high usability in all interactions. Friedman tests showed no statistical significance between interactions.

### 5.5.3 *Fatigue*

Table 5.6. Mean reported fatigue per interaction (1 = Extreme, 5 = None)

| Metric | Unmodeled | Hyperplanar | Spherical |
|---|---|---|---|
| Finger fatigue | 4.53 | 4.50 | 4.50 |
| Wrist fatigue | 4.37 | 4.37 | 4.17 |
| Arm fatigue | 4.33 | 4.33 | 4.33 |
| Shoulder fatigue | 4.57 | 4.53 | 4.40 |
| Neck fatigue | 4.87 | 4.87 | 4.87 |

Participants filled out a 5-point independent rating Likert scale questions from ISO 9241-9 Annex C to collect data in regards to fatigue. Table 5.6 depicts the questions and the mean reported ratings per interaction. The reported ratings demonstrate a minor presence of fatigue in each interaction. Friedman tests showed no statistical significance between interactions.

### 5.5.4 *Borg Scale*

Table 5.7. Mean reported effort per interaction (0 = Nothing at all, 0.5 = very very weak (just noticable), 1 = very weak, ..., 10 very, very strong)

| Metric | Unmodeled | Hyperplanar | Spherical |
|---|---|---|---|
| Arm Effort | 1.20 | 1.37 | 1.27 |
| Shoulder Effort | 0.95 | 1.02 | 1.05 |
| Neck Effort | 0.15 | 0.20 | 0.27 |

Previous work in gestures encouraged the use of the Borg Scale for arm, shoulder, and neck effort (Hincapié-Ramos et al., 2014), and were therefore included in our subjective assessment. We used the Borg scale from ISO 9241-9 Annex C. Table 5.7 shows the questions and the mean reported ratings for each interaction. The table demonstrates that minimal effort was required when using the interactions. Friedman tests showed no statistical significance between interactions.

# CHAPTER SIX

## Analysis

The goal of this study was to evaluate the hypotheses using the metrics from the previous section. We report our findings of the hypotheses by splitting hypothesis H2 to better illustrate the results based on each of the interactions.

*H1* Users will learn gestural interaction over time, allowing for an improvement in performance.

We accept *H1* as the performance improvement analysis showed a significant improvement in all models between each day, which is indicative of learning. Performance improvement as measured by Cohen's $d$ also reinforces, and describes this finding by showing a medium to large improvement from day 1 to day 2, and a small improvement from day 2 to day 3.

*H2a* An interaction with a simple model of the interaction space will perform better than an interaction which does not model the space.

We accept *H2a*. This decision was based on a comparison of performance, accuracy and Fitts' law conformance between the *Hyperplanar* and *Unmodeled* interaction.

In terms of performance, The *Hyperplanar* interaction obtained statistically better bivariate throughput than the *Unmodeled* interaction. However, this improvement was minor and not practically significant based on Cohen's interpretation.

With respect to accuracy, the *Hyperplanar* interaction obtained statistically better Movement Offset than the *Unmodeled* interaction, which was found to be a small-medium effect as defined by Cohen. It also had statistically better Movement Error, although the practical significance is minor.

In terms of Fitts' conformance, the *Hyperplanar* interaction demonstrated a strong linear correlation ($R$) between Movement Time and IDe, had an intercept that was within the ideal range, and a higher goodness of fit ($R^2$). The *Unmodeled* interaction, on the other hand, demonstrated a weak linear correlation ($R$), an intercept that was not within the ideal range, and a lower $R^2$. It can also be seen in Table A.3 that the *Hyperplanar* interaction demonstrated an improvement over time with regards to Fitts' law, while the *Unmodeled* interaction did not.

Since the *Hyperplanar* interaction was found to be as good, or better, in the individual metrics of performance, accuracy, and predictability as measured by Fitts' conformance, we consider the *Hyperplanar* interaction as a whole to obtain better results than the *Unmodeled* interaction.

*H2b* An interaction which models the interaction space using a sphere will perform better than an interaction which uses a hyperplane.

We fail to reject the null hypothesis in the case of *H2b*. To evaluate this hypothesis we compared the *Hyperplanar* and *Spherical* interactions using each of the collected metrics: performance, accuracy, and Fitts' conformance.

In terms of performance, the *Hyperplanar* interaction obtained statistically better bivariate throughput than the *Spherical* interaction, which was found to be a small-medium effect as defined by Cohen.

In terms of accuracy, *Hyperplanar* interaction obtained statistically better accuracy than the *Spherical* interaction in 5 of the 7 accuracy metrics. None are considered to be practically significant. The *Spherical* interaction obtained significantly better Movement Offset than the *Hyperplanar* interaction, but was not practically significant.

Fitts' regressions demonstrated a strong linear correlation, an intercept within the ideal range, and a relatively high 'goodness of fit' in both the *Hyperplanar*, and *Spherical* interactions. However the *Spherical* interaction was found to have a higher correlation, goodness of fit, and an intercept that was closer to the ideal value when compared to the *Hyperplanar* interaction.

In addition to the aforementioned quantitative metrics, users also showed preference towards the *Hyperplanar* interaction which was shown to be significant. Taking all these into consideration, we cannot conclude that the *Spherical* interaction is better than the *Hyperplanar*, thus we fail to reject the null hypothesis.

# CHAPTER SEVEN

# Discussion

## *7.1 Discussion and Future Works*

Over the course of this work, we found performance improvements between all 3 days of the study which we believe is caused by participants learning the interaction. While this may seem trivially true, current literature has in some cases failed to identify learning in these gestural interfaces (Brown et al., 2014, Sambrooks and Wilkinson, 2013, Adhikarla et al., 2015). We believe this difference is due to our experimental design, which used a longitudinal study and random-order practice. With the inclusion of these experimental design parameters performance improvements were found in both the pilot and main study.

We found that fatigue and effort were reported to be minimal from the rested position, as demonstrated in Table 5.6 and Table 5.7. This finding further reinforces current literature that has stated that resting the elbow during interaction reduces fatigue (Freeman et al., 2012, Segen and Kumar, 2000, Brown et al., 2014, Jude et al., 2014b). This finding suggests that fatigue is not as large of a problem in the current state of gestural interaction, as previous research has stated that fatigue was one of the main problems with gestural interaction (Segen and Kumar, 2000).

Additionally, modeling the interaction space resulted in an interaction which conformed to Fitts' law. Conversely, we found that an interaction which does not model the input space was not well explained by Fitts' law, nor was it within the ideal intercept range in each of the 3 days. This means that the relationship between movement time and Index of Difficulty (IDe) is not consistently predictable in the unmodeled approach, and offers less value to those designing interfaces (Soukoreff and MacKenzie, 2004). We believe this poor conformance may be attributed to the little incorporation of the actual biomechanical motion in the unmodeled approach. The approach used to detect these results was in line with the current standard's encouragement to regress the participant's mean movement time (MT) over their respective effective IDe, as opposed to regressing over mean of means per IDe (Soukoreff and MacKenzie, 2004).

We found that the interaction which used a simple model was significantly better than the unmodeled interaction in terms of both performance and accuracy. Only the latter was found to be practically significant as interpreted by Cohen. However, the interaction whose model was more complex in terms of the input space performed significantly worse than the other two models despite its mapping to the biomechanics of the body. This may be caused by the strict assumptions of the model. The two angles, azimuthal ($\theta$) and elevation ($\phi$), were inferred from the previously measured elbow position and the palm position provided by the Leap Motion. Therefore, this model requires a higher tracking precision in order to perform optimally.

Upon review of the notes taken during the experiment, which included participants' feedback, we identified a few issues with the *Spherical* interaction. A recurring issue was with the static elbow placement. Participants had difficulty finding their calibrated elbow position even when a marker was present. We also found that during the pilot, participants mistakenly knocked the tracker off of its stand during their note-taking between rounds. This did not seem to significantly penalize the more simple models but caused large cursor jitters during the *Spherical* interaction. We attempted to solve this issue by building a new stand using Lego blocks which allowed the tracker to tightly fit into place. While this increased stability, the stand itself was still movable when hit. We posit that a better interaction can be built using a device which directly tracks the arm in terms of relative angles. The Myo Armband is promising example of a new device that tracks these angles directly, and we plan to use in future work. Another option would be to allow for dynamic elbow tracking in which both the hand and the elbow are dynamically tracked so that the elbow does not need to be in a fixed position. We believe that the better fit offered by the *Spherical* approach could yield positive results if these considerations are addressed.

While we provided overall effect size using $\eta_p^2$, and pairwise effect sizes for performance and accuracy metrics as measured by Cohen's d, interpreting these effect sizes however proved to be difficult. These interpretations are meant to be domain-specific (Morris and Fritz, 2013), but no such guidelines exist within the domain of pointing device evaluation. We had to therefore fallback on the interpretations provided by Cohen, which are meant to be used as a last resort (Lakens, 2013). By providing the effect size of our study, we aim to provide better context for future research, and to contribute towards establishing guidelines for interpreting effect size within this domain.

## *7.2 Conclusion*

In this study we used a longitudinal design to evaluate the two hypotheses. We learned that modeling the interaction space results in an interaction which can be explained by Fitts' law. Conversely, we learned that an unmodeled approach conforms weakly to Fitts' law. We also learned that a simple model of the user's interaction space resulted an interaction that was as fast and more accurate than an interaction which did not model the user's interaction space. Furthermore, we introduced a more complex model of the interaction space which maps the arm movement from a rested position to the 2D screen with no loss of information using the forearm angles as input. This more complex model did not exhibit better performance nor accuracy than the simpler model. We posit that this is due to the interaction having too many constraints and being unsuitable for use with an input device which uses the hand position as input. Finally, we showed that gestural interaction demonstrated performance improvements over multiple sessions, from which we infer learning.

APPENDICES

# APPENDIX A

## Tables, Figures and Transformations

### *A.1 Tables*

Table A.1. Means and standard deviation of throughput for each device per day

| Source | Day 1 | | Day 2 | | Day 3 | | Day 4 | | Day 5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | Stdev | Mean | Stdev | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| Planar | 2.68 | 0.80 | 2.96 | 0.92 | 3.20 | 0.89 | 3.26 | 0.94 | 3.22 | 0.89 |
| Spherical | 2.63 | 0.84 | 2.76 | 0.84 | 3.00 | 0.85 | 3.04 | 0.88 | 3.15 | 0.89 |
| De Facto Rested | 2.50 | 1.79 | 2.49 | 0.83 | 2.89 | 0.87 | 3.22 | 0.94 | 3.15 | 0.93 |
| De Facto Unrested | 2.73 | 0.88 | 3.01 | 0.90 | 3.10 | 0.85 | 3.18 | 0.94 | 3.17 | 0.90 |
| Touchpad | 3.59 | 1.24 | 3.67 | 1.15 | 3.77 | 1.27 | 3.79 | 1.13 | 3.70 | 1.15 |

Table A.2. Means and standard deviation per trial of each accuracy metric on day 5

| Source | Movement Error | | Movment Offset | | Movement Variability | | Target Reentry | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Stdev | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| Planar | 66.06 | 48.94 | -3.81 | 75.93 | 51.12 | 40.59 | 0.214 | 0.457 |
| Spherical | 61.50 | 43.14 | 1.02 | 69.23 | 48.48 | 34.45 | 0.263 | 0.540 |
| De Facto Rested | 68.47 | 50.88 | 1.82 | 76.94 | 54.65 | 43.35 | 0.311 | 0.516 |
| De Facto Unrested | 63.32 | 48.32 | 1.51 | 74.44 | 48.99 | 37.58 | 0.303 | 0.556 |
| Touchpad | 53.96 | 57.36 | 6.44 | 70.66 | 44.94 | 52.73 | 0.177 | 0.404 |

Table A.3. Intercept, slope, coefficient of determination ($R^2$) and correlation coefficient ($R$) of all 3 interactions over all 3 days.

| | Unmodeled | | | | Hyperplanar | | | | Spherical | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Intercept | Slope | $R^2$ | $R$ | Intercept | Slope | $R^2$ | $R$ | Intercept | Slope | $R^2$ | $R$ |
| 1 | 453.05 | 248.18 | 0.31 | 0.56 | 113.29 | 334.08 | 0.63 | 0.79 | 18.97 | 381.91 | 0.65 | 0.81 |
| 2 | 785.04 | 108.17 | 0.07 | 0.26 | 59.47 | 308.46 | 0.73 | 0.86 | 38.62 | 332.73 | 0.71 | 0.84 |
| 3 | 585.86 | 155.53 | 0.18 | 0.42 | 79.86 | 295.67 | 0.74 | 0.86 | 16.30 | 331.50 | 0.76 | 0.87 |

## *A.2 Figures*

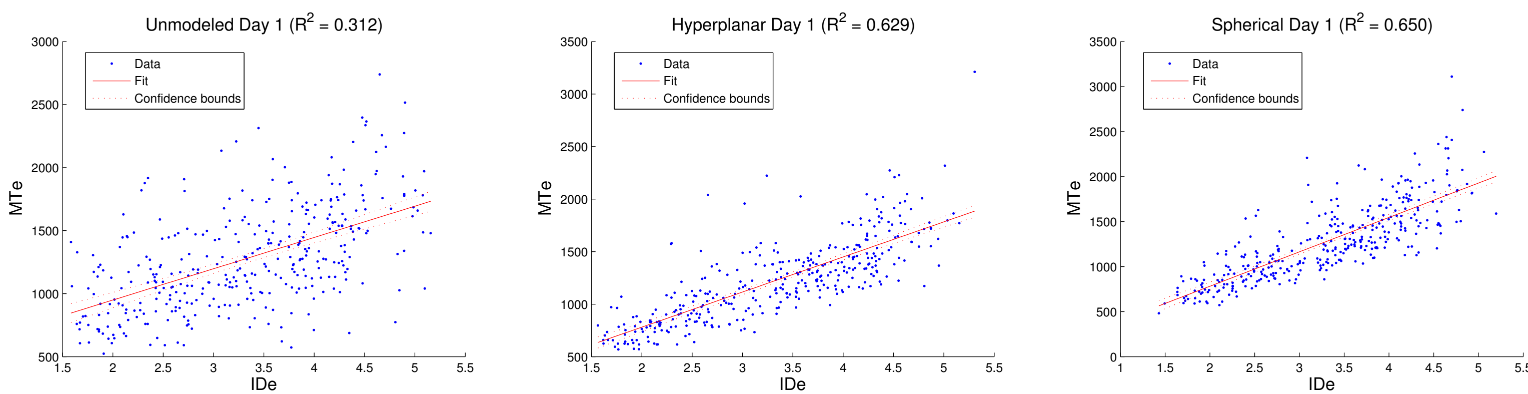



Figure A.1. Mean movement time of all participants from day 1 as a function of effective Index of Difficulty (IDe), with computed Fitts's regression lines

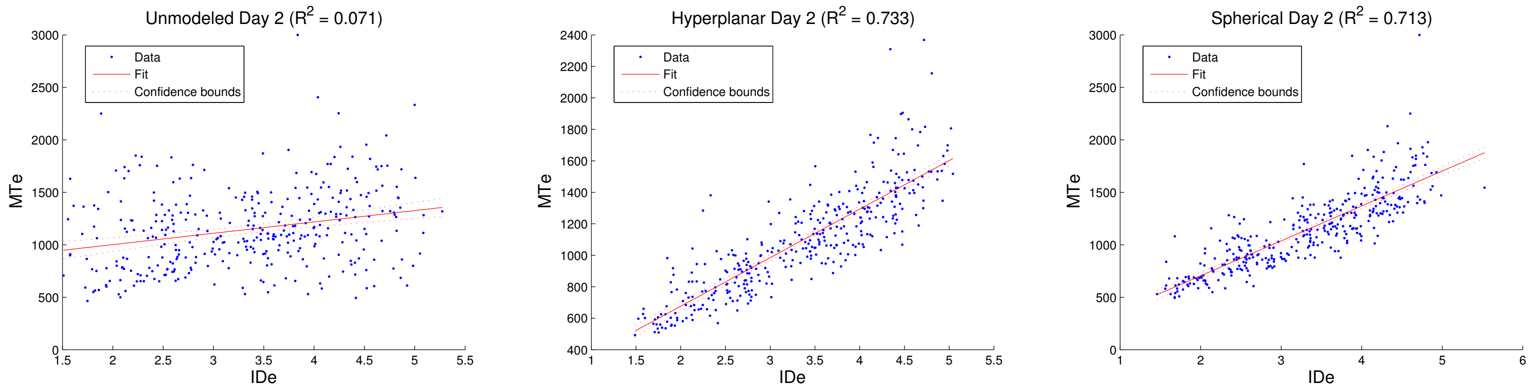



Figure A.2. Mean movement time of all participants from day 2 as a function of effective Index of Difficulty (IDe), with computed Fitts's regression lines

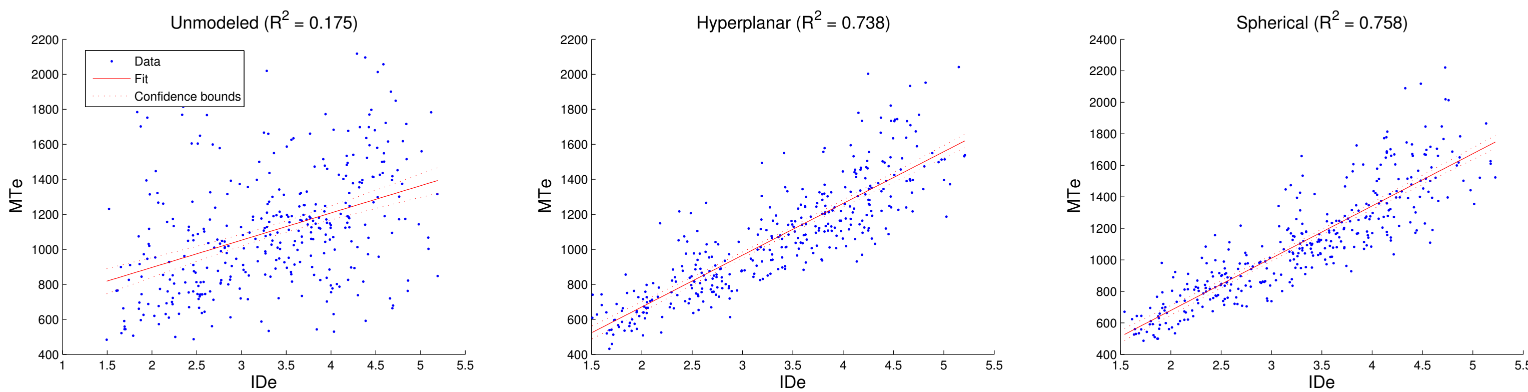



Figure A.3. Mean movement time of all participants from day 3 as a function of effective Index of Difficulty (IDe), with computed Fitts's regression lines

### *A.3 Transformation*

convertToSphere($\hat{h_x}, \hat{h_y}, \hat{h_z}, R$ )

$\hat{h} = [\hat{h_x}, \hat{h_y}, \hat{h_z}]$, estimated hand positions from the leap controller

$e = [e_x, e_y, e_z]$, measured elbow position from the leap controller

$radius = R$, where R is equal to the length of the arm from elbow to palm.

$V = \hat{h} - e$

$\theta = \arccos{(V_y/R)}$

$\phi = \arctan 2(V_z, V_x) \approx \arctan{(\frac{V_z}{V_x})}$ but with the correct quadrant computed

end

generateProjectiveTransformation()

$V_\theta = [\theta_1, \theta_2, \theta_3, \theta_4]$ , $V_\phi = [\phi_1, \phi_2, \phi_3, \phi_4]$ $X_{dest} = [x_1, x_2, x_3, x_4]$, $Y_{dest} = [y_1, y_2, y_3, y_4]$

$$B = \begin{bmatrix} \theta_1 & \phi_1 & 1 & 0 & 0 & 0 & -x_1 * \theta_1 & -x_1 * \phi_1 \\ 0 & 0 & 0 & \theta_1 & \phi_1 & 1 & -y_1 * \theta_1 & -y_1 * \phi_1 \\ \theta_2 & \phi_2 & 1 & 0 & 0 & 0 & -x_2 * \theta_2 & -x_2 * \phi_2 \\ 0 & 0 & 0 & \theta_2 & \phi_2 & 1 & -y_2 * \theta_2 & -y_2 * \phi_2 \\ \theta_3 & \phi_3 & 1 & 0 & 0 & 0 & -x_3 * \theta_3 & -x_3 * \phi_3 \\ 0 & 0 & 0 & \theta_3 & \phi_3 & 1 & -y_3 * \theta_3 & -y_3 * \phi_3 \\ \theta_4 & \phi_4 & 1 & 0 & 0 & 0 & -x_4 * \theta_4 & -x_4 * \phi_4 \\ 0 & 0 & 0 & \theta_4 & \phi_4 & 1 & -y_4 * \theta_4 & -y_4 * \phi_4 \end{bmatrix}$$

$$A = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ x_2 \\ x_3 \\ y_3 \\ x_4 \\ y_4 \end{bmatrix}$$

$A\lambda = B$, solved using SVD

$A = [\lambda(1:6), 0, 0, 1]$

$C = [\lambda(7:8), 1]$

end

computePoint()

$[\theta, \phi]$ =convertToSphere($\hat{h_x}, \hat{h_y}, \hat{h_z}, R$)

$V = [\theta, \phi, 1]$

$left = A * V'$ , A is a global matrix solved above

$rightScalar = C * V'$, C is a global matrix also solved for above

$projected = left/rightScalar$

end

# APPENDIX B

## Original Hyperplanar Description

For additional information, I will include the original description of the *Hyperplanar (Planar)* interaction model. The description that follows was taken directly from Alvin Jude's thesis titled *Giving the Users a Hand: Towards Touchless Hand Gestures for the Desktop* with permission (Jude, 2014).

### *B.1 Solution Design*

#### *B.1.1 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 B.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.

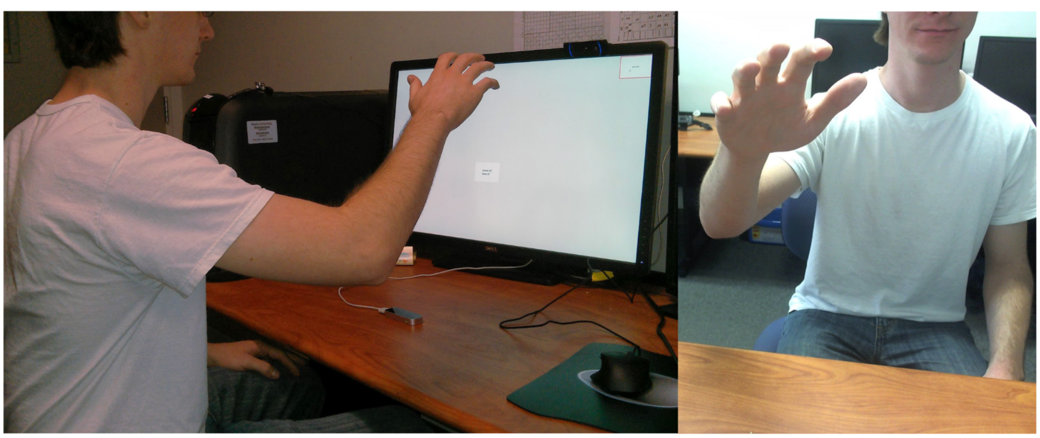


Figure B.1. Commonly used gestural interaction method.

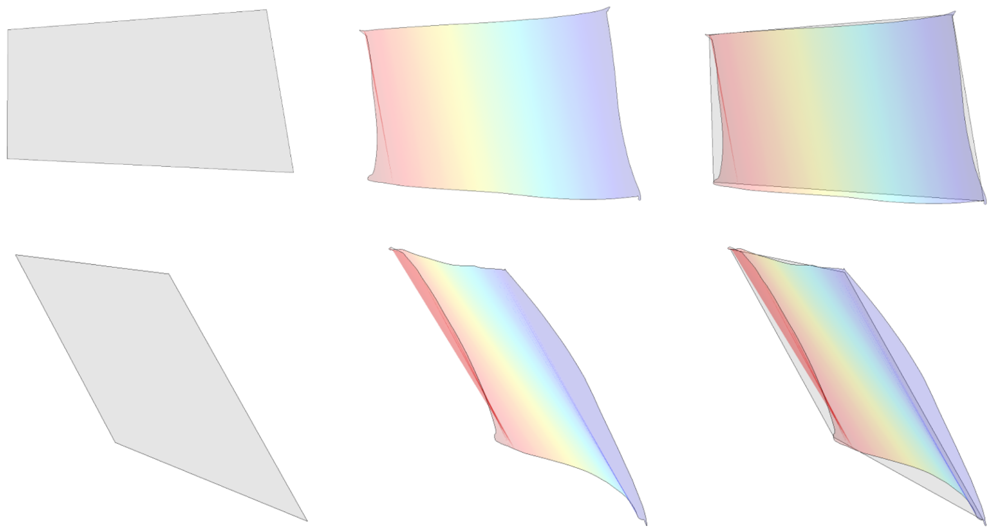


Figure B.2. Comparison of the actual and calibrated space.

### *B.1.2 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 B.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

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 × Transformation = Destination, and the 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.

# BIBLIOGRAPHY

Adhikarla, V. K., Sodnik, J., Szolgay, P., and Jakus, G. (2015). Exploring direct 3d interaction for full horizontal parallax light field displays using leap motion controller. *Sensors*, 15(4):8642–8663.

Bailly, G., Müller, J., Rohs, M., Wigdor, D., and Kratz, S. (2012). Shoesense: A new perspective on gestural interaction and wearable applications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1239–1248, New York, NY, USA. ACM.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods*, 37(3):379–384.

Banerjee, A., Burstyn, J., Girouard, A., and Vertegaal, R. (2011). Pointable: an in-air pointing technique to manipulate out-of-reach targets on tabletops. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 11–20. ACM.

Bigdelou, A., Schwarz, L., and Navab, N. (2012). An adaptive solution for intra-operative gesture-based human-machine interaction. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces*, IUI '12, pages 75–84, New York, NY, USA. ACM.

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

Brown, M. A., Stuerzlinger, W., and Filho, E. J. M. (2014). The performance of un-instrumented in-air pointing. In *Proceedings of the 2014 Graphics Interface Conference*, GI '14, pages 59–66, Toronto, Ont., Canada, Canada. Canadian Information Processing Society.

Carmody, T. (2010). Why gorilla arm syndrome'rules out multitouch notebook displays. *Wired, Oct*, 10.

Chen, X., Schwarz, J., Harrison, C., Mankoff, J., and Hudson, S. E. (2014). Air+ touch: interweaving touch & in-air gestures. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 519–525. ACM.

Cockburn, A., Quinn, P., Gutwin, C., Ramos, G., and Looser, J. (2011). Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback. *International Journal of Human-Computer Studies*, 69(6):401–414.

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. (1988). Statistical power analysis for the behavioral sciences.

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

Criminisi, A., Reid, I., and Zisserman, A. (1999). A plane measuring device. *Image and Vision Computing*, 17(8):625–634.

Drewes, H. and Schmidt, A. (2007). Interacting with the computer using gaze gestures. In *Human-Computer Interaction–INTERACT 2007*, pages 475–488. Springer.

Ens, B., Hincapié-Ramos, J. D., and Irani, P. (2014). Ethereal planes: A design framework for 2d information space in 3d mixed reality environments. In *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction*, SUI '14, pages 2–12, New York, NY, USA. ACM.

Fagarasanu, M. and Kumar, S. (2003). Carpal tunnel syndrome due to keyboarding and mouse tasks: a review. *International Journal of Industrial Ergonomics*, 31(2):119 – 136.

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.

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

Grossman, T., Wigdor, D., and Balakrishnan, R. (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, pages 61–70, New York, NY, USA. ACM.

Guinness, D., Poor, G. M., and Jude, A. (2014). Gestures with speech for hand-impaired persons. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*, ASSETS '14, pages 259–260, New York, NY, USA. ACM.

Guna, J., Jakus, G., Pogačnik, M., Tomažič, S., and Sodnik, J. (2014). An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. *Sensors*, 14(2):3702–3720.

Han, T., Alexander, J., Karnik, A., Irani, P., and Subramanian, S. (2011). Kick: investigating the use of kick gestures for mobile interactions. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 29–32. ACM.

Hincapié-Ramos, J. D., Guo, X., Moghadasian, P., and Irani, P. (2014). Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 1063–1072, New York, NY, USA. ACM.

Jota, R., Nacenta, M. A., Jorge, J. A., Carpendale, S., and Greenberg, S. (2010). A comparison of ray pointing techniques for very large displays. In *Proceedings of Graphics Interface 2010*, GI '10, pages 269–276, Toronto, Ont., Canada, Canada. Canadian Information Processing Society.

Jude, A. (2014). Giving the Users a Hand: Towards Touchless Hand Gestures for the Desktop. Master's thesis, Baylor University, TX, USA.

Jude, A., Poor, G. M., and Guinness, D. (2014a). An evaluation of touchless hand gestural interaction for pointing tasks with preferred and non-preferred hands. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, pages 668–676. ACM.

Jude, A., Poor, G. M., and Guinness, D. (2014b). 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. ACM.

Kaptein, M. and Robertson, J. (2012). Rethinking statistical analysis methods for chi. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1105–1114, New York, NY, USA. ACM.

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

Kurtenbach, G. and Hulteen, E. A. (1990). Gestures in human-computer communication. In *The Art of Human Computer Interface Design*, pages 309–317. Addison-Wesley.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. *Frontiers in psychology*, 4.

Lee, T. D. and Genovese, E. D. (1988). Distribution of practice in motor skill acquisition: Learning and performance effects reconsidered. *Research Quarterly for Exercise and Sport*, 59(4):277–287.

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

MacKenzie, I. S. and Jusoh, S. (2001). An evaluation of two input devices for remote pointing. In *Engineering for human-computer interaction*, pages 235–250. Springer.

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

MacKenzie, I. S. and Teather, R. J. (2012). Fittstilt: The application of fitts' law to tilt-based interaction. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pages 568–577. ACM.

McLester, J. and Pierre, P. (2007). *Applied Biomechanics: Concepts and Connections*. Cengage Learning.

Mentis, H. M., O'Hara, K., Sellen, A., and Trivedi, R. (2012). Interaction proxemics and image use in neurosurgery. In *Proceedings of the SIGCHI Conference on human factors in computing systems*, pages 927–936. ACM.

Morris, P. E. and Fritz, C. O. (2013). Effect sizes in memory research. *Memory*, 21(7):832–842.

Nagy, J. G., Plemmons, R. J., and Torgersen, T. C. (1996). Iterative image restoration using approximate inverse preconditioning. *Image Processing, IEEE Transactions on*, 5(7):1151–1162.

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

Plemmons, D. and Mandel, P. (2015). Leap motion developers.

Poppinga, B., Sahami Shirazi, A., Henze, N., Heuten, W., and Boll, S. (2014). Understanding shortcut gestures on mobile touch devices. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pages 173–182. ACM.

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, 6(2):135–147.

Sambrooks, L. and Wilkinson, B. (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*, pages 119–122. ACM.

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

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

Soukoreff, R. W. and MacKenzie, I. S. (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.

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

Triesch, J. and Von Der Malsburg, C. (1998). Robotic gesture recognition by cue combination. In *Informatik'98*, pages 223–232. Springer.

Vatavu, R.-D., Anthony, L., and Wobbrock, J. O. (2013). Relative accuracy measures for stroke gestures. In *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ICMI '13, pages 279–286, New York, NY, USA. ACM.

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

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

Wobbrock, J. O., Cutrell, E., Harada, S., and MacKenzie, I. S. (2008). An error model for pointing based on fitts' law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1613–1622, New York, NY, USA. ACM.

Wobbrock, J. O. and Gajos, K. Z. (2008). Goal crossing with mice and trackballs for people with motor impairments: Performance, submovements, and design directions. *ACM Trans. Access. Comput.*, 1(1):4:1–4:37.

Wobbrock, J. O., Shinohara, K., and Jansen, A. (2011). The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1639–1648, New York, NY, USA. ACM.

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

Zelinsky, A. and Heinzmann, J. (1996). Real-time visual recognition of facial gestures for human-computer interaction. In *Automatic Face and Gesture Recognition, 1996., Proceedings of the Second International Conference on*, pages 351–356. 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.