Summer 2016

What role does effort play: the effect of effort for gesture interfaces and the effect of pointing on spatial working memory

Xiaoxing Liu
University of Iowa

WHAT ROLE DOES EFFORT PLAY: THE EFFECT OF EFFORT FOR GESTURE INTERFACES AND THE EFFECT OF POINTING ON SPATIAL WORKING MEMORY

by

Xiaoxing Liu

A thesis submitted in partial fulfillment
of the requirements for the Doctor of Philosophy
degree in Informatics (Information Science) in the
Graduate College of
The University of Iowa

August 2016

Thesis Supervisor: Professor Geb W. Thomas

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

____

PH.D. THESIS

____

This is to certify that the Ph.D. thesis of

Xiaoxing Liu

has been approved by the Examining Committee for
the thesis requirement for the Doctor of Philosophy degree
in Informatics (Information Science) at the August 2016 graduation.

Thesis Committee: ___________________________________
Geb W. Thomas, Thesis Supervisor

____
Juan Pablo Hourcade

____
Jon Winet

____
Nathan Fethke

____
Susan Wagner Cook

To my family, who live on the other side of this planet, but who have unconditionally supported me since I first saw this world.

ACKNOWLEDGEMENTS

The path toward this dissertation has been circuitous. It is complete thanks in a large part to the special people who challenged, supported, and stuck it out with me along the way. Without them, this dissertation would not have been possible.

First and foremost, I would like to express my sincere gratitude to my advisor and mentor, Dr. Geb Thomas, for the substantial influence his guidance and support has had on my research project. He has been extremely motivating, encouraging and enlightening when my steps faltered along the way. Dr. Thomas reached out to me in the beginning of the third year, the most difficult phase of my doctoral study. Since then he has guided me through my research project. He has also set an example of excellence as a researcher and provided immense help on academic career development. I feel extremely privileged to have been working with him.

I am very thankful to Dr. Haowei Hsieh, my advisor during my first two years of study, for all his help, advice and encouragement. Haowei introduced me to the study of human computer interaction (HCI) and led me to enter the HCI world. Thank you for sharing your knowledge and experience with me as an international researcher. The stories of your experiences will be forever cherished.

I also want to gratefully acknowledge the members of my Ph.D. committee for their time, guidance and valuable feedback. Dr. Susan Wagner Cook, Dr. Jon Winet, Dr. Juan Pablo Hourcade, and Dr. Nathan Fethke have provided professional ideas from different perspectives. Their diversity in research specialties is a valuable asset to my dissertation.

My time at the University of Iowa was made enjoyable in large part due to many friends and groups that became a part of my life. I must acknowledge my tremendous and deep gratitude to Xiyan Wang. Because of her patience and support during my good and bad times, I have been able to complete my degree. I am also grateful for time spent with friends (too many to list here but you know who you are!) for providing support and friendship that I needed, and for many other people and memories.

My deepest appreciation goes to my family, especially my mom, Xinxin Chang, for fostering a tenacious spirit in my personality and for her unconditional support of my Ph.D. study. Thank you for setting me on the path and for your love all my years in school.

ABSTRACT

Automatically recognizing gestures of the hand is a promising approach to communicating with computers, particularly when keyboard and mouse interactions are inconvenient, when only a brief interaction is necessary, or when a command involves a three-dimensional, spatial component. Which gestures are most convenient or preferred in various circumstances is unknown. This work explores the idea that perceived physical effort of a hand gesture influences users' preference for using it when communicating with a computer. First, the hypothesis that people prefer gestures with less effort is tested by measuring the perceived effort and appeal of simple gestures. The results demonstrate that gestures perceived as less effortful are more likely to be accepted and preferred. The second experiment tests similar hypothesis with three-dimensional selection tasks. Participants used the tapping gesture to select among 16 targets in two environments that differ primarily in the physical distance required to finish the task. Participants, again, favor the less effortful environment over the other. Together the experiments suggest that effort is an important factor in user preference for gestures. The effort-to-reliability tradeoff existing in the majority of current gesture interfaces is then studied in experiment 3. Participants are presented 10 different levels of effort-to-reliability tradeoff and decide which tradeoff they prefer. Extreme conditions were intentionally avoided. On average they rate their preferred condition 4.23 in a 10-point scale in terms of perceived effort, and can achieve a success rate of approximately 70%. Finally, the question of whether pointing to objects enhances recall of their visuospatial position in a three-dimensional virtual environment is explored. The results show that pointing actually decreases memory relative to passively viewing. All in all, this work suggests that effort is an

important factor, and there is an optimal balance for the effort-to-reliability tradeoff from a user's perspective. The understanding and careful consideration of this point can help make future gesture interfaces more usable.

PUBLIC ABSTRACT

This project aims to study the role physical effort plays in a computer interface utilizing hand gestures, and the role pointing plays on spatial memory. First we compare the physical effort, recognition accuracy, and user satisfaction of eight command gestures. Results suggest that gestures perceived as less effortful are more likely to be accepted and preferred in a gesture interface. Next we compare two pointing and selection conditions that differ only in the movement distance required to finish the task. In each condition participants select 16 targets using hand gestures. Effort plays a similar role in this test — participants favor the less effortful condition over the other. Together the results demonstrate that effort is an important factor for user satisfaction. In the third experiment we present 10 different conditions for the same selection task. The physical effort necessary to finish the task decreases along the 10 settings, while the selection reliability decreases as well. Thus we create 10 levels of effort-to-reliability tradeoff. Participants are asked to explore the conditions to decide which tradeoff they prefer. Extreme conditions were intentionally avoided by the participants. Results provide a user-selected success rate of 70%, and a perceived effort rating of 4.23 in a 10-point scale. Finally, we examine the influence of pointing in comparison to passively viewing an array of objects that participants are attempting to memorize in three-dimensional gesture interfaces. Pointing to an array decreases recall relative to passively viewing, but the effect seems unrelated to the 3D element of the task.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I:
INTRODUCTION

Gesturing is a natural part of human communication. When circumstances prevent effective verbal communication, people almost irrepressibly try to communicate with their hands (Goldin-Meadow & Morford, 1990). For example, deaf children, who have not yet been exposed to sign language, spontaneously exploit gestures to communicate (Fant, 1972).

This study emphasizes free-hand gestures, which include movements of the arms, hands, fingers, or some combination of these movements, but excludes gestures involving any kind of touch-based devices like a touchscreen or a touchpad. Thus, the definition of a free-hand gesture interface does not require the user to touch or handle an input device, although sometimes a controller or glove is used to measure the hand and finger positions.

It seems possible that gestures extend and facilitate thinking for speakers. Such a natural modality to convey information, however, has not yet been successful for communication between people and computers. Existing modes of human-computer interaction are mostly based on simple, mechanical devices like a keyboard and mouse. While people naturally use gestures for more effective communication with other people, current popular gesture interfaces for communication between people and machines are fatiguing (Stern, Wachs, & Edan, 2006) and are often considered slower than using a keyboard and mouse for traditional HCI tasks (Amft, Amstutz, Smailagic, Siewiorek, & Tröster, 2009; Cabral, Morimoto, & Zuffo, 2005; Kouroupetroglou et al., 2011; Pino, Tzemis, Ioannou, & Kouroupetroglou, 2013).

When designing user interface strategies, designers must often resolve tradeoffs between user needs/preferences and system reliability (Wright, Lin, O'Neill, Cosker, & Johnson, 2011), and the right balance has not yet been struck for gesture interfaces. Perhaps the reason the balance has not been struck is due to our lack of understanding of what constitutes an effective set of gestures from the user's perspective.

The primary objective of this study is to discover the relationship between the effort someone is willing to invest in a gesture and the value of the communication the gesture affords. Another interest of this study is how pointing gesture may influence one's spatial working memory. Specifically, we seek the answers to the following questions:

- Is effort a factor of user satisfaction for gestures and gesture interfaces? Would a gesture or an interface be more desirable because it is less effortful?
- How do effort and reliability, together, influence user preference?
- What would a user choose to be the optimal balance for the effort-to-reliability tradeoff?
- How does pointing to an object influence one's recall to that object's location, especially in a three-dimensional virtual space?

CHAPTER II:
BACKGROUND

Gesture Interfaces

Gesture interfaces enable people to communicate with digital devices in ways other than typing, clicking, or speech, usually through body movement or through facial expressions. Gesture interfaces are usually categorized as a Natural User Interface (NUI), because body movement and facial expressions are natural ways people interact with the world (Jain, Lund, & Wixon, 2011).

There are few examples of free-hand gesture interfaces in widespread use, like motion-activated sinks used in many public restrooms. Also several popular video game consoles, such as Microsoft Xbox®, Nintendo Wii®, and Sony PlayStation®, use free-hand gesture interfaces, albeit with different technologies. These free-hand gesture interfaces facilitate two categories of interaction: 1) menu system interaction, and 2) motion control in games.

With menu system interaction, pointing gestures control the cursor position and either a hovering gesture or a physical button click activates item selection. Together these gesture combinations provide functionality similar to the familiar point-and-click personal computer interface facilitated by a mouse. When using a mouse, people mentally map the region of the horizontal plane on which the mouse moves to the generally vertical area of display. With game consoles, people can point directly at the screen plane so the mapping between the action and the response is simplified.

Motion control in these games is often tied to gestures associated with real-world behaviors. For example, players in some car racing games (“Forza Motorsport (series),”

2014) players move their hands as if controlling a steering wheel. In sports-themed games ("Kinect Sports," 2014), players move their arms and hands as if they were swinging a tennis racket, bowling, or otherwise performing typical actions of the sport being simulated.

Why Do People Gesture?

People have an innate desire to gesture. Goldin-Meadow (1999) examines the role of gesture in communication and thinking, when gestures are used with or without speech, respectively. When gestures are required to carry the full burden of communication (as in conventional sign language), gestures mimic the lexicon and syntax of traditional oral language. When gestures accompany spoken language, however, they become more imagistic. This modality is the most common; 90% of gestures are found in the context of speech (D. McNeill, 1992).

A review of the research concerning how human communication is affected by the use of telecommunication media suggests that information is communicated nearly as effectively in the absence of gestures, on the telephone, or from behind a screen (Williams, 1977) as it is when communicated with the aid of gestures. This suggests that gestures are not essential to the interpretation of speech. However, in speaking situations with interfering noise, listeners rely on gestural cues. Roger (1978) asked 60 subjects to respond to multiple-choice items designed to test their comprehension of 12 videotaped spoken utterances. They respond to stimuli in one of three presentation conditions: audiovisual, audiovisual without lip and facial cues, and audio-alone, over four signal-to-noise ratios. The results indicate that gestures can, at times, significantly improve comprehension rates, even without lip and facial cues. Visual cues become increasingly useful as noise is introduced.

Iverson and Goldin-Meadow (1998) tested the hypothesis that “speakers gesture because they understand that gestures can convey useful information to the listener.” They found that people who are blind from birth gesture even when talking to a listener known to be blind. Their blind participants gesture at a rate not reliably different from that of sighted pairings. Thus, blind speakers do not seem to gesture solely to convey information to the listener. Another example of this pattern is the observation that people gesture while they are on the phone (Rimé, 1982). It seems possible that in addition to aiding communication with others, gestures may facilitate a speaker’s thinking.

With speech, gestures are linked to thinking, based on Cassell’s observation of “the extreme rarity of gestural errors” (Cassell, 1998). Spoken language is fragile. Speakers sometimes convey the wrong meaning by selecting unintended words or omitting important words. Gestures, however, virtually always portray the speaker’s communicative intent. Speakers may say “left” and mean “right,” but they reliably point toward the intended direction.

Gestures are integral to the process of communication. They help speakers produce communication and they help listeners perceive communicated intent. In his 1992 book *Hand and Mind*, McNeill (1992) concludes that gesture, language, and thought are different facets of a single mental/brain/action process. They are integrated on actionable, cognitive, and biological levels. He writes, “…language is inseparable from imagery. The imagery in question is gestures. Speech and gesture emerged in evolution together. Speech could not have evolved without gesture, and gesture could not have evolved without speech" (David McNeill, 2008).

Clearly, our natural evolution has led us to include gestures as a communication modality and gesture interfaces serve as the bridge between people and digital products. How gestures should be used to interact with digital devices is an active area of investigation.

Gesture Interface Technologies and Applications

Although gesture interfaces have not yet enjoyed widespread use, over the past three decades, researchers have been developing diverse technologies, applications, and algorithms for gesture recognition.

Technology

The two most common technological approaches for capturing finger and hand positions for gesture recognition are: a) wearable devices, especially gloves, and b) vision-based approaches.

The wearable devices use mechanical or optical sensors to detect body part movements and positions. The most common format is to attach the sensors to a glove (Baudel & Beaudouin-Lafon, 1993; Zimmerman, Lanier, Blanchard, Bryson, & Harvill, 1987). Other approaches include a wristband and pad (Rekimoto, 2001), a tiny projector and cameras coupled to a pendant (Mistry & Maes, 2009), and a wrist watch (Amft et al., 2009). Often a tracking system is integrated to measure the position and orientation of the hand(s). The tracking technologies used for this purpose include magnetic, (Bolt, 1980) inertial, and ultrasonic systems. The best data glove systems can provide accurate measurements of hand pose and movement. Unfortunately, they also require user-specific calibration, they restrict the naturalness and ease of interaction, and they are often expensive.

Forcing users to wear a device is an important limitation. Users resist wearing tracking devices and computer-bound gloves because it restricts the user's freedom-of-movement and it takes time to don and remove. Vision-based gesture recognition relies on video images, avoiding the need to physically restrict the user's freedom-of-movement.

Starner and Pentland (1995) demonstrated that one camera is sufficient to effectively and accurately recognize hand posture and gestures. However, in order to track hand movement in three dimensions, two or more cameras are needed. Although multiple cameras are more complicated to integrate than a single camera, the extra information is essential when three-dimensional information is required (Cabral et al., 2005; Crowley, Berard, & Coutaz, 1995; Lee et al., 2008; Nickel & Stiefelhagen, 2007; Quek, 1996; Wachs et al., 2008; Wah Ng & Ranganath, 2002).

There are situations where RGB images cannot provide sufficient information for effective gesture recognition. For example, images captured in dim or unpredictable lighting can be difficult to segment. Skin-colored objects in the scene can also interfere with the segmentation and recognition of hand gesture. Depth images provide an

Figure 1. RGB and Depth image in limited lighting condition.
Left-RGB image, Right-depth image

excellent alternative in these situations because they are less likely to be affected by lighting conditions or colors (Figure 1). Kinect, Microsoft's low-cost depth sensor, uses this approach and has enjoyed considerable attention within the gesture recognition community (Bellmore, Ptucha, & Savakis, 2011; Cerlinca & Pentiuc, 2012; Du & To, 2011; Li, 2012; Ren, Meng, Yuan, & Zhang, 2011; Riener, Rossbory, & Ferscha, 2011; Ronchetti & Avancini, n.d.; Van den Bergh & Van Gool, 2011; C. Yang, Jang, Beh, Han, & Ko, 2012).

Applications

Various researchers have proposed hand gestures as substitutes for traditional input methods (e.g. using a mouse as a pointing device). For example, Quek's (1996) FingerMouse® recognizes two-dimensional finger movements as input to the desktop and Crowley et al.'s FingerPaint (Crowley et al., 1995) enables a user to "draw" in a projected image.

Other researchers have applied gesture interfaces to specific applications or fields. Wachs et al. (Wachs et al., 2008) describe a hand-gesture-tracking device called Gestix that allows surgeons to browse MRI images in a sterile environment within the operating room. Nickel and Stiefelhagen (2007) test a real-time system for human-robot interaction to detect pointing gestures and to estimate pointing directions. Trigueiros, Ribeiro, & Lopes (2012) use a Kinect® to remotely navigate a robot. Riener et al. (2011) use gestures to read, browse, search, or delete emails while driving. Wilson (2010) turns tabletops into touchscreen devices with a downward-looking Kinect® sensor. Cabral et al. (2005) build a virtual reality environment in which gestures are used for three-dimensional navigation or visualization.

Suarez and Murphy (2012) survey 37 papers that involve hand gesture recognition with depth images. The variety of applications and interfaces is still limited. Twenty-four of thirty-seven papers include real-world applications but are mainly for testing their hand localization and gesture classification algorithms.

Current Research Issues in Gesture Interfaces

Despite its long history and the many gesture-related research projects reported in the literature, gesture interfaces still pose a wide variety of interesting research questions and fundamental puzzles that remain relatively unexplored.

Questions Concerning Performance

Communicating with gestures seems to be slower than using a keyboard and mouse for traditional HCI tasks (Amft et al., 2009).

Fitts' Law (Fitts, 1954) is widely used to model the time required to complete a pointing task based on a function of distance to the target, *D*, and the width of the target, *W*. Fitts' Law is typically expressed mathematically as:

$$T = a + b \cdot \log_2(1 + \frac{D}{W})$$

where *T* is the average time taken to complete the movement and *a* and *b* are parameters that depend on the user and the details of the task. The expression allows two pointing techniques to be compared, accounting for the natural speed-to-accuracy tradeoff that is typically evident in a pointing task. The second component of the formula, $\log_2(1 + \frac{D}{W})$, is often referred to as the Index of Difficulty. It increases logarithmically as the travel distance increases or the target gets smaller.

Kouroupetroglou et al. (2011) use Fitts' Law to compare the performance of gesture-based, two-dimensional and three-dimensional pointing tasks with both a Wiimote controller and a standard mouse. For two-dimensional tasks a mouse is superior to the Wiimote; the throughput (mean Index of Difficulty / mean Movement Time) is 41.2% lower for the Wiimote than the mouse, and the missed clicks count is three times higher for the Wiimote than the mouse.

Cabral et al. (2005) describe a two-dimensional, computer vision-based gesture recognition system for interaction in VR environments. They tested the interface in several scenarios, including using gestures as a point-and-click device, as a three-dimensional navigation tool, and as a three-dimensional visualization tool. They found that the average time to complete a point-and-click experiment was 7.3 seconds using a mouse and 26.1 seconds using gestures. Their Fitts' Law analysis also suggests that the performance with a mouse is much better than the performance with gestures, particularly when the Index of Difficulty increases.

Due to the relatively low performance of gestures in traditional point-and-click tasks, some authors conclude that gesture interfaces should be relegated to a complimentary or alternative interface to the existing desktop paradigm, or that gestures should be developed as specialized interfaces for specific applications, where they might be more efficient for limited tasks (Nielsen, Störring, Moeslund, & Granum, 2004). However, most evaluations reported in the literature are mainly based on traditional desktop tasks. Their ability to sufficiently evaluate the performance of a gesture interface is arguable. Although gestures may be less effective in two-dimensional pointing tasks, they may be more effective for other applications. For example, although Cabral et al.

(2005) found poor performance with the point-and-click tasks, they found advantages for gestures in the two other scenarios tested. They found that 50% of the users reported their gestures as excellent for the three-dimensional visualization experiment. Also, the overall sensation of immersion in the virtual reality environment was improved by using a gesture interface in the three-dimensional navigation experiment.

As a follow-up experiment to Kouroupetroglou et al. (2011) (which found superior Fitts' law performance for the mouse in two-dimensional pointing tasks), Pino et al. (2013) replaced the Wiimote with a Kinect interface and disabled the "Enhance Pointer Precision" option for the mouse. For two-dimensional tasks, their findings were consistent with Kouroupetroglou et al.: the throughput (dividing Index of Difficulty by the mean Movement Time) using Kinect was almost 39% lower than using the mouse, and missed clicks (clicks off target) were 50% higher. However, for three-dimensional pointing tasks, the situation was reversed. The Kinect interface performed better than the mouse interface. The mouse had a 9% lower throughput than the Kinect, and missed clicks were almost identical.

The effectiveness of gesture interfaces depends heavily on the application in which they are used. For example, Gestix (Wachs et al., 2008) is a gesture interface for use in surgical operating rooms. The questionnaire (Lewis, 1991) used to test its usability evaluates the task experience, ease, completion time, and overall satisfaction. The authors report that, "The system is easy to use, with fast response and quick training times. Learning tests for 8 gestures required only 10 trials to converge with an average task performance time of 22 seconds." In this case, the utility of the system in overcoming the sterilization issues associated with a keyboard and mouse interface makes up for a slightly decreased performance. Overall, the users seem to be quite satisfied with the tradeoff.

Lack of Evaluation Method

It seems that no consensus currently exists on how to evaluate gesture interfaces. It is difficult to answer the question “What is a good gesture interface?” particularly from a user’s viewpoint. This lack of understanding of how to evaluate gesture interfaces has limited opportunities to improve their application.

Gesture recognition accuracy has been a popular preoccupation among researchers; many papers focus on developing algorithms to achieve higher accuracy. Tools investigated for hand gesture recognition include: hidden Markov models (HMMs), particle filtering and condensation, finite-state machines (FSM), and neural networks (Mitra & Acharya, 2007). For example, Starner and Pentland (Starner & Pentland, 1995) demonstrate a real-time HMM-based system for recognizing sentence-level American Sign Language (ASL) without explicitly modeling the fingers. They report 99% accuracy for hand recognition with colored gloves and 92% accuracy without gloves. Yang and Ahuja (M.-H. Yang & Ahuja, 2001) use a time-delayed neural network to recognize 40 hand gestures of ASL. Their experiment yields a recognition rate of 99.02% and 96.21% for training set and testing set, respectively.

For computer vision-based gesture recognition, the quality of the acquired images has a large impact on recognition rate, depending on algorithms used. Nguyen, Pham, & Jeon (2009) use gray scale morphology and geometric calculations to find fingertip locations. They detect hands based on skin colors and report 90% to 95% recognition accuracy for open finger gestures, but only 10 to 20% when the fingers are closed.

To improve gesture recognition accuracy for computer vision-based approaches and decrease sensitivity to poor imaging conditions, many researchers have explored the use of depth images. The advantage of depth images over RGB images for gesture recognition is the ease of performing hand segmentation. In applications in which users face the camera and gesture with their hands forward toward the camera, a simple depth threshold is often sufficient to isolate the hands. In some cases, systems can more reliably predict hand depth based on the location of other body parts, like the head. For example, Ren et al. develop a system that uses the Microsoft Kinect sensor for gesture recognition (Ren et al., 2011). Their lexicon consists of 14 gestures representing the numbers 0 – 9, and the operators +, -, *, and /. They apply a shape distance metric called Finger-Earth Mover's Distance for hand gesture recognition and report a mean accuracy of 90.6%. Yang et al. (2012) implement an HMM-based gesture recognition system with depth images captured with the Kinect sensor. For the 8 basic commands they use (up, down, left right, circle counter-clockwise, circle clockwise, push to device, pull from device) they report an overall accuracy of 96.66%.

Gesture recognition accuracy describes how reliably an algorithm or system is able to identify a gesture from a defined gesture set. It is likely that designers select specific gestures that may be easily distinguished by the technology being employed in the recognition. These decisions may sacrifice the utility, memorability, comfort and naturalness of the gestures from the user's viewpoint in favor of ensuring that the system will work reliably as it has been defined.

A fairer evaluation of system performance should better balance the needs of the user with the need to make the system reliable. Perhaps the reason this balance has not been struck is a lack of understanding of what constitutes an effective set of gestures from the users' viewpoint.

What Are the Right Gestures?

User-Defined Gestures

Several researchers have attempted to derive gesture sets from natural user behavior. In the attempt to define a standard gesture set to be used in further studies to compare gesture recognition systems, Wright et al. (2011) conducted an empirical study in which participants are asked to perform gestures for given tasks. Tasks ranged from basic computer tasks such as "Select …" to more abstract tasks, such as "Show me a …." Participants could interact with any device or service in their surroundings. Gestures made by each participant were recorded. The researchers then extracted the most common gestures for different tasks. Both the system performance and the user performance and preference were evaluated for the derived gesture set. Their results indicate a clear tradeoff between the needs of the system and the desires of the user for specific gestures. For example, the second most preferred gesture for the "pick up" task was the "grasp and pick up" gesture; but the same gesture was also the second worst in terms of system error. The researchers summarized that if the system was designed strictly based on the desires of users, it would not provide a reliable gesture-recognition rate. Their result showed only 61% gesture-recognition accuracy. In this case, there is an evident tradeoff between user preference and recognition reliability.

While a user-defined gesture set is a promising way to evaluate user preference, the only criterion used in the above research is a subjective questionnaire on "how well the gesture matches the action." This approach evaluates only the intuitiveness of the selected gesture set, ignoring other important factors such as the effort required to make a gesture.

Nielsen et al. (2004) propose a human-based procedure for developing intuitive and ergonomic gesture sets. The steps include: a) find the functions; b) collect gestures from user domain; c) extract gesture vocabulary; and d) benchmark the chosen gesture vocabulary. They consider several usability concerns including: the learning rate, ergonomics, and intuition. Particularly for step c) the authors emphasize ergonomics:

- Evaluate internal force caused by posture
    - Deviation from neutral position
    - Outer limits
    - Forces from inter-joint relations
- Evaluate frequency and duration of that gesture
- Consider the effects on the wrist based on wrist and finger posture

They run three tests to evaluate a gesture's semantics, memorability, and physical stress. A gesture's semantics relates to a gesture's logical connection to its semantic interpretation, a feature that is similar to Wright's evaluation. The three-element evaluation is more comprehensive from a user's viewpoint. However, although the authors mention that it should be possible to unambiguously recognize a gesture set, their procedure does not explore challenges related to recognizing a specific feature set or the compromises that these challenges may force in design decisions.

Using Effort as a Measurement

One aspect of gestures that has not been well explored in the literature is how fatiguing a gesture is and the amount of effort people are willing to invest in order to communicate with a system. Several researchers include fatigue as a gesture usability issue. For example, in Cabral et al. (2005) the authors report "40% of the users complained about fatigue of the arms." In Nielsen et al. (2004) "Half of the testees found the gesturing to be more tiring than expected." Developing a less effortful gesture interface, however, appears to be beyond the scope of most published research.

Both user needs and technical requirements need to be considered to design an effective gesture interface. Often, measuring system performance is relatively easy with widely implemented usability testing methods. Measuring intuitiveness and stress (or effort) is more challenging because there are few widely accepted measurement approaches and these factors are subjective and must be obtained empirically. Also, the causes of stresses and their impact on a user's behavior while using a gesture interface is generally unexplored. This knowledge gap seems to be the bottleneck in the design of optimal gesture vocabularies.

To some extent, people's limb trajectories follow predictable patterns. Gestures appear to minimize some "cost," either to minimize metabolic energy costs, movement time, distance, peak velocity, peak acceleration, or jerk (rate of change of acceleration) (Alexander, 1997; Engelbrecht, 2001; Flash & Hogan, 1985; Nelson, 1983). For example, Flash and Hogan (1985) propose a minimum-jerk theory, which postulates that hand trajectories are chosen such that the time integral of the squared magnitude of the hand jerk is minimal. To evaluate the model, simulated trajectories were compared to measured hand trajectories. Results showed that observed human planar two-joint arm movement matched the model. As they summarized, "…unconstrained point-to-point motions are approximated straight with bell-shaped tangential velocity profiles; curved motions have portions of low curvature joined by portions of high curvature; at points of high curvature, the tangential velocity is reduced.” Uno, Kawato, & Suzuki (1989) experimented with a number of minimum theories such as minimum energy, minimum torque, and minimum time before they decided to postulate the minimum-torque-change theory. They developed an experimental apparatus similar to that used by Flash and Hogan. Their model successfully predicts observed trajectories under various conditions such as planar free movement, via-point movement, and constrained movement under the external force.

Stern et al. (2006) proposed a gesture vocabulary design approach that evaluated both intuitiveness and comfort. Using techniques similar to the user-defined gesture approaches mentioned above, they presented the result of a command and allowed the user to physically compose the gesture most associated with the command. To evaluate effort, participants held each hand pose for 25 seconds, followed by a 15 second rest. Then the participant rates the perceived level of stress on a scale of 0 – 10. This study supports the notion that participants favor easy gestures and inadvertently filter out stressful gestures when selecting gestures.

Effect of Pointing on Spatial Working Memory

Abundant research has shown not only that gesture facilitates thinking, it also aids memory. Gesturing when trying to recall information may facilitate speakers’ access to that information in both verbal and spatial memory. Speakers gesture more when trying to remember infrequent words than when trying to remember frequent words (Krauss & Hadar, 1999), and when describing visual objects (like a picture) from memory than when the objects are visually accessible (De Ruiter, 1998; Wesp, Hesse, Keutmann, &

Wheaton, 2001). Studies also indicate that speakers gesture more when describing objects that tax spatial working memory, such as drawings that are difficult to remember and encode verbally, than when describing objects that can be readily labeled, such as a house, or a flower (Morsella & Krauss, 2004).

It is evident, though, that the relationship between memory system, gesture, and the intervening cognitive processes are quite complex, even within the field of spatial working memory. On one hand, perception underlies action, providing information to the various action systems (Hannus, Cornelissen, Lindemann, & Bekkering, 2005). On the other hand, action can influence perception as well. For example, it is suggested that a specific action intention, such as grasping, can enhance visual processing of action-relevant features, such as orientation (Bekkering & Neggers, 2002). This leads to the larger question: how does action influence spatial working memory? One specific question is when trying to memorize the location of an object, what is the effect of pointing to that object as it appears relative to passively viewing that object, especially when this happens in a three-dimensional gesture interface? While it's reasonable to assume that pointing to an object would enhance memory for that object's locations, as pointing would activate both visual and motor encodings, it is also possible that the pointing might have a negative effect on memory because the requirement of pointing to an object may leave fewer resources available for memorization of the location.

The scientific literature on the influence of action on memory is mixed. In the classic model of visuospatial working memory (VSWM) (Logie, 1995), the author distinguished between visual and spatial components of working memory by proposing a static visual component known as the visual cache dealing with storage of visual information, and an active rehearsal mechanism known as the inner scribe, which codes in terms of spatial movements. In this context, secondary tasks involving a motor component have been found to have disruptive effects on the concurrent encoding and rehearsal of spatial arrays of items (Kemps, 2001; Quinn, 1994). For example, Vandierendonck et al. (2004) found that concurrent execution of a matrix-tapping task during stimulus presentation impaired performance of the immediately following recall in the Corsi blocks tasks, and this adverse effect on performance was reliable in sequences of all lengths both in backward and in forward-recall.

More recently, however, Chum et al. (2007) showed that in specific conditions, pointing movements may also facilitate spatial working memory. In that study, the authors develop a task that requires concurrent maintenance of two spatial arrays, one of which was encoded only through passive visual observation (the *no-move* array), whereas the other was encoded by visual observation accompanied with limb pointing movements (the *move* array). During encoding, the sample arrays were presented one item at a time on a computer touch-screen display. Upon testing, only one array (move or no-move) was presented, which may have been identical to the sample array, or may have been one item shifted in position. Recognition performance was measured by a same-different judgment. Results showed that performance was significantly higher in the move than in the no-move condition. The authors suggested that pointing might increase spatial-based perception, or form a stronger egocentric spatial encoding. At the same time, a series of subsequent studies (Dodd & Shumborski, 2009; Rossi-Arnaud, Spataro, & Longobardi, 2011; Spataro, Marques, Longobardi, & Rossi-Arnaud, 2015) demonstrated that the enhanced recognition memory by pointing to spatial arrays occurred only when concurrent maintenance of two spatial arrays (move and no-move) were required and instructions were given within trials. Dodd & Shumborski (2009) showed that when pointing instruction was blocked (e.g., participants pointed to or passively viewed all items in an array as opposed to pointing to some while passively viewing others), pointing to an array of objects actually decreased memory relative to passively viewing that array. Rossi- Arnaud et al. (2011) revealed a second limit that pointing movements do not improve visuospatial memory when the design is mixed but instructions are manipulated between separate trials, rather than within the same trial. Spataro et al. (2015) examined the relationship between pointing movements and serial recall of spatial positions. They found that both item and order memory were significantly worse for pointed-to arrays rather than those that were passively viewed. Furthermore, pointing impaired the recall of the initial and middle positions more than the recall of the final positions.

Conclusion and Research Questions

Despite the fact that gestures are a natural part of human communication, and despite many years of research, gesture interfaces have been widely adopted for only a few niche application areas. Much of the existing research has emphasized gesture recognition technologies while relatively few projects have emphasized the selection of gestures for particular applications. There are clear design tradeoffs between the needs of the gesture recognition system requirements and the comfort and difficulty of the gesture for the user. How people might behave when faced with such a tradeoff has not been adequately studied. This lack of understanding regarding the role of user effort in performing gestures has limited the improvement of gesture interfaces. Relevant research suggests that user effort is reasonably predictable and generalizable: the effort of performing a gesture affects the attractiveness of the gestures from the user's viewpoint. Therefore this essential relationship, between the effort someone is willing to invest in performing a gesture and the value the gesture affords, needs to be explored because the relationship has an excellent chance of advancing the success of gesture interfaces.

Another factor that is often overlooked in evaluating gesture interfaces is how the use of gestures in a computer interface may affect a user's perception and memory of specific information. Previous studies examining the relationship between pointing gestures and spatial working memory have involved only pointing to a touchable object (although the objects in computer-based testing are virtual objects, pointing to a touch-screen was required) in a 2D space. Examining the same relationship in a gesture interface may build a better understanding on the cognitive effect gestures have and how gesture interfaces can benefit from it. One specific aim of this study is to further examine the influence of pointing on spatial working memory, by manipulating a three-dimensional factor in a virtual environment, in which participants perform in-air pointing gestures to a gesture interface without touching any real world physical object.

The following questions are of specific interest of this study.

Is Effort a Factor of User Satisfaction for Gestures and Gesture Interfaces?

The literature suggests that there is a relationship between the physical effort a gesture requires and a user's satisfaction with that gesture. For convenience, I focus on two common tasks in gesture interfaces: 1) command tasks, such as swiping left to see next picture in a gallery application, and 2) pointing and navigating tasks, such as using hand movements to click and move a target in a graphical user interface. My hypothesis is that for both these tasks, people prefer gestures that require less effort. I test this hypothesis by comparing user preferences for pairs of similar gestures that differ primarily in the effort they require. Experiment 1 emphasizes command tasks; Experiment 2 emphasizes pointing and navigation tasks.

How Do Effort and Reliability Together Influence User Preference?

In this study, reliability refers to the frequency that the system correctly recognizes the user's intended gesture. Often designers must compromise preferred gestures in favor of gestures that are more accurately or reliably recognized by the sensor system. My hypothesis is that people prefer gestures that require less effort and that also communicate accurately and reliably. The lack of system reliability may negatively affect user preference. Experiment 1 tests this hypothesis by comparing user preference for the same gesture while varying the recognition rate for the gesture between 100% and the rate found with a current commercial. Further, it should show evidence that effort is an effective measurement to evaluate a gesture interface in addition to the commonly used measurement of recognition reliability.

Where Does the Tradeoff Reach a Tipping Point from a User's Perspective?

There is a tradeoff between system reliability and user preference, and often increasing system reliability means increasing the user's effort. I hypothesize that users intuitively determine how much additional effort should be expended in order to gain better reliability. Measuring this tradeoff is important in designing gesture interfaces. Experiment 3 explores this tradeoff to measure how much effort a user is willing to sacrifice to gain greater reliability.

How Do Pointing Movements Influence Spatial Working Memory in a 3D Gesture Interface?

It is well established that gesture facilitates memory and makes it last. Of interest here is whether pointing has an effect on a user's spatial working memory. My hypothesis is that simply pointing to an object does not enhance memory for that object and may actually impair memory as the mental effort required to execute a motor movement may leave fewer resources available to memorize object location. Experiment 4 verifies this hypothesis by comparing participants' recall in two, move and no-move conditions. In the move condition, participants are asked to point to, and memorize a series of locations. In the no-move condition, participants are asked to passively view, and memorize the spatial array.

Together, these specific aims reveal a clearer understanding of the relationship between system reliability and user effort. This understanding may help researchers and practitioners to create better gesture interfaces and develop a more productive and efficient human-computer interaction.

CHPATER III:
ESTIMATING THE EXPERIMENTAL SAMPLE SIZE

This chapter describes the theoretical statistical framework used to determine the sample sizes needed for the first three experiments.

The larger the sample size, the more precise a measurement. However, larger sample sizes require more participants or longer testing sessions, which can be costly and wasteful (Kraemer & Thiemann, 1987). The estimate of a reasonable sample size starts by calculating the critical value, t: (Jeff Sauro, 2012)

$$t = \frac{d}{sem}$$

Where d is the minimum size of the difference to be measured in the experiment, and sem is the standard error of the mean. The standard error of the mean is calculated by dividing the standard deviation of the measurement, s, by the square root of the sample size, n, or

$$sem = \frac{s}{\sqrt{n}}$$

Combining these and rearranging yields:

$$n = \frac{t^2 s^2}{d^2}$$

Thus, the sample size, n, depends on the sample variance, s^2, the critical value, t, and critical difference, d.

I obtained s^2 from a pilot study. In the two pilot experiments, participants provided subjective responses on 10-point scales. The standard deviations of their responses in pilot experiment 1 and 2 are presented in Table 1 and Table 2. The largest standard deviation, *s*=2.54, was used to estimate the necessary sample sizes for the main experiments.

Difference	**Q1 Std Dev**	**Q3 Std Dev**
Pinch - Grab	2.03	2.50
Key Tap S - L	1.90	2.12
Swipe Left - Up	1.40	2.54
Circle S - L	1.50	2.12

Table 1. Standard Deviation from pilot experiment 1

Difference	**Std Dev**
Q1 A-B	1.96
Q3 A-B	1.60

Table 2. Standard Deviation from pilot experiment 2

There are no mathematical criteria to determine the appropriate value of *d*, the smallest difference between the obtained and true values the experiment needs to detect. The value of *d* should be the smallest interval of practical interest to the experimenter. I arbitrarily chose a value of 10%, or 1 point in the 10-point subjective scales used in the experiments.

The critical value, *t*, determines the desired level of statistical confidence. I chose to set *t* to provide a confidence interval of 95%. A complication in using *t* is that the value of *t* depends on its degrees of freedom, which depends on the sample size that we want to estimate. Diamond (2001) describes how to use iteration to circumvent this difficulty by suggesting the use of a *z*-score for *t*. Once a sample size is estimated with this initial estimate, that sample size can be used to determine *t* in the next iteration. Then iterate until the sample size converges.

In this case:

- The variance from the pilot scores was 6.47 (s=2.54).
- The critical difference (*d*) is 10% (or 1 point for a 10-point scale).
- The desired level of confidence is 95% (initial value of *z* equals 1.96, for a two-tailed test).

The estimation is presented in Table 3, leading to a sample size of 28.

	Initial	**Round 2**	**Round 3**
t	1.96	2.06	2.05
t2	3.84	4.26	4.21
s	2.54	2.54	2.54
s2	6.47	6.45	6.45
d	1.00	1.00	1.00
d2	1.00	1.00	1.00
df	24.00	27.00	27.00
Calculation	24.86	27.48	27.17
Round Up	**25.00**	**28.00**	**28.00**

Table 3. Estimate of sample size

CHAPTER IV:
EXPERIMENT 1 – EFFORT AND USER PREFERENCE OF EIGHT COMMAND GESTURES

Basic hand gesture elements are the foundational blocks of a gesture interface. The overall effort of interacting with a gesture interface is directly related to the effort of the most frequently used gesture elements. Experiment 1 measures the effort and user preference for gesture elements independent of any specific task. It explores whether users prefer less effortful gestures by asking participants to repeat each gesture multiple times, then rate how tiring that gesture was and whether or not they liked the gesture.

A pilot test indicated that system recognition reliability strongly influenced user ratings of effort and satisfaction. When the system repeatedly failed to recognize a user's attempt to perform a gesture, the user had to perform the same gesture many times. As a result, the users became frustrated and increased their subjective rating of the gesture effort and lowered their subjective rating of preference. Thus the main experiment was redesigned to eliminate the confounding effect of gesture recognition reliability. First users performed each gesture for the experimenter independent of the Leap system, simulating a 100% reliability rate. This is similar to the Wizard of Oz experiment approach (Kelley, 1984), in which participants interacted with a computer that participants believed to be intelligent, but was actually being operated by a human being. After the first scenario, participants repeated the experiment with the Leap system, experiencing the effects of occasionally unrecognized performance attempts. This allowed me to measure the effect of reliability on the gesture attempts.

The Gestures

The experiment considered eight gestures. Each of these can be used in a gesture interface as commands themselves or as components of more complex gestures. The gestures are grouped in pairs; each pair consists of an easy gesture and a relatively effortful one (Table 4).

Pair	Easy Gesture	Difficult Gesture
1	Pinch	Grab
2	Key Tap – Finger	Key Tap - Palm
3	Swipe Left	Swipe Up
4	Circle Small	Circle Large

Table 4. Eight gestures tested in experiment 1

Pinch and Grab

To perform a pinch gesture, squeeze together the thumb and index finger. The two fingers must touch and form a closed circle. To perform a grab gesture, hold the hand in a neutral position, then form a fist. All five fingers must be adducted and each finger joint flexed.

Key Taps

A key tap is performed by flexing the curved index finger at the first joint or holding the index finger extended and adducting the wrist, or some combination of the two, as if tapping a keyboard key. The gesture is difficult for some people because it involves movement in two directions – the fingertip must move downward, then upward (Figure 2). An easy key tap (tap small) is defined as tapping with only the finger while the elbow is at rest. Meanwhile, a difficult key tap (tap large) involves maintaining the forearm in the air while using the whole palm or forearm to tap, like bouncing a ball.

Figure 2. Key Tap Gesture

Swipe

Swiping is a long, linear movement of the hand and fingers (Figure 3). Swiping in the horizontal plane is considered less effortful because it can be accomplished with a pronation and supination of the wrist with only a small movement at the elbow. To swipe upward, a user generally needs to move their arm movement, which requires more effort. The experiment compared swiping left with swiping upward.

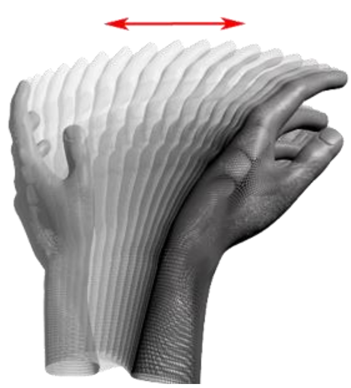

Figure 3. Swipe Gesture

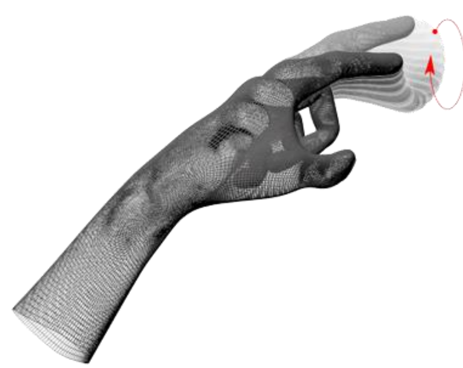

Figure 4. Circle Gesture

Circle

A circle gesture is performed by using a finger to draw a circle in the air (Figure 4). The two gestures in this group differ in the size of the circle. An easy, small circle has a radius smaller than 6 cm, while the difficult, large circle has a radius greater than 8 cm.

The pair of gestures in each group requires similar movements, but one requires greater effort than the other. Participants performed all eight gestures, and then gave their ratings on a survey. The hypothesis is that the difficult gesture will cost more effort. This will be checked with user perception of effort.

Method

Participants

Sixteen female and fourteen male adults, with a combined mean age of 25.6 years, were recruited from the University of Iowa community and participated in this experiment. Twenty-one participants had experience with gesture interfaces including in-game motion. The participants were divided into two cohorts balanced by gender and gesture interface experience. The cohorts completed the gestures in reverse sequence. This control was intended to offset any effect introduced by the order the procedures, including the effect of fatigue.

Materials

Leap Motion Controller (Leap) (“Leap Motion,”) was used to capture hand gestures. Leap is a small sensor that sits on the desktop and monitors any movement in the hemispheric space above it. It integrates two IR cameras and three infrared LEDs. Very few details about the controller have been released by the company. Its effective working space is approximately a 50cm x 50cm x 50cm area. Unlike Microsoft Kinect, which is designed to acquire and recognize full body and arm gestures, Leap monitors only a small area and is more suitable for hand gestures. The controller itself is accessed and programmed through APIs with support for various platforms and languages. This experiment used its JavaScript API and all test tasks were programmed as HTML webpages. The participants used the graphical interface displayed in Google's Chrome browser while moving their hands above the Leap controller placed on a desk in front of them to finish the tasks in this experiment.

Procedure

Before beginning, the experimenter demonstrated each gesture to be performed. Participants learned and practiced all eight gestures in the warm-up session until they could reliably reproduce each of the intended motions from memory.

The first cohort performed the gesture pairs in the following order: 1) pinch, 2) grab, 3) tap small, 4) tap large, 5) swipe left, 6) swipe up, 7) circle small, and 8) circle large. The second cohort performed them in reverse order.

For the first round of testing, each participant performed each gestures 20 times for the experimenter, simulating a completely reliable gesture recognition system. After completing each gesture pair, each participant rated the perceived effort and how well they liked the gestures on a 10-point scale for both the easy and difficult gestures. The survey questions were:

Q1. Please rate how physically tiring this gesture was on a scale from 0 representing for not at all tiring to 10 for very fatiguing.

Q2. How appealing is this gesture for a future computer application? Please rate your feeling on a scale from 0 for "I hate it" to 10 for "I like it."

In the second round of testing, the participants repeated each gesture until the system successfully recognized it 20 times, with feedback given for each successful completion of each individual attempt presented in the graphical user interface. In the second round, participants were paced at the speed of one gesture attempt per second.

After completing each pair, participants responded to the following three survey questions:

Q3. Please rate how physically tiring this gesture was, to use in this system, on a scale from 0 representing for not at all tiring to 10 for very fatiguing.

Q4. Please rate your perception of how well the system recognized this gesture correctly on a scale from 0 for poor recognition to 10 for perfect recognition.

Q5. How appealing is this gestures for a future computer application? Please rate your feeling on a scale from 0 for “I hate it” to 10 for “I like it.”

Experiment 1 tested the following specific hypotheses:

1. Participants find the easy gesture in each pair to be less tiring than the difficult gesture. Specifically, a Wilcoxon signed ranks test will reveal a significant difference in Q1 responses for hard and easy gestures in each gesture pair. The null hypothesis would indicate that there is no difference in the perceived effort.

2. Participants rate the easy gesture in each pair more favorably than the difficult gesture. Specifically, a Wilcoxon signed ranks test will reveal a significant difference in Q2 responses for hard and easy gestures in each pair. The null hypothesis is that there is no difference in user preference.

3. Grouping across all gestures and participants, less tiring gestures tend to receive higher user appeal rating. Specifically, after pooling all 8 gestures and all participants, there will be a significant correlation between Q1 and Q2. In this correlation, a negative r value would represent a negative correlation between appeal and effort, supporting this hypothesis.

4. User satisfaction is negatively affected when the system does not consistently and reliably recognize user input. Specifically, there will be a significant correlation between system reliability (Q4) and the change in appeal for each gesture from the simulated 100% gesture recognition to the actual recognition (Q2 – Q5). A significant negative correlation would confirm this hypothesis.

5. Finally, both effort and system reliability will be important factors in overall user satisfaction with a gesture. A multivariable regression with appeal rating (Q5) as a dependent variable and both perceived effort (Q1) and reliability (Q4) ratings as independent variables will reveal that each of these factors significantly correlate with participants' preferences.

Results

Perceived Effort

Figure 5 shows the mean perceived effort rating of the eight gestures (participant response to Q1). The difference between easy and difficult gestures in each pair was analyzed by four separate Wilcoxon signed ranks tests. The results presented in Table 5 show that the differences between three gesture pairs (pinch / grab, $Z = -0.36$, $p = 0.02$; tap pairs, $Z = -2.98$, $p = 0.003$; and circle pairs, $Z = -3.18$, $p = 0.001$) were significant.

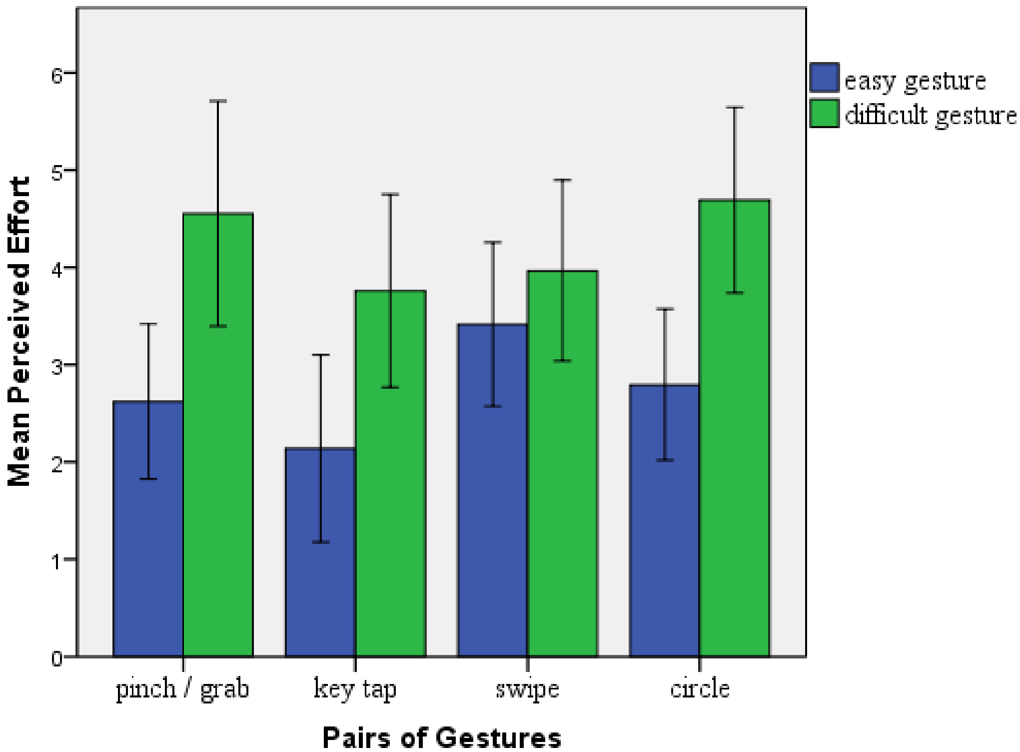

Figure 5. Experiment 1: Mean perceived effort for eight gestures from 0 for not tiring at all to 10 for very fatiguing. Error bars indicate the 95% confidence interval of the mean.

	Pinch/Grab	Key Tap	Swipe	Circle
Z	-3.057	-2.982	-1.955	-3.178
Asymp. Sig. (2-tailed)	.002	.003	.051	.001

Table 5. Wilcoxon signed ranks test to Q1 of each pair of gestures.

User Satisfaction

Figure 6 shows the mean participant appeal rating (Q2) for all eight gestures. The difference in appeal rating between easy and difficult gestures in each pair was analyzed by four separate Wilcoxon signed ranks tests. The results presented in Table 6 show that the differences between two tap gestures ($Z = -2.22$, $p = 0.027$) and two circle gestures ($Z = -2.44$, $p = 0.015$) were significantly different. However, the differences in appeal rating for the pinch and swipe pairs were not significant.

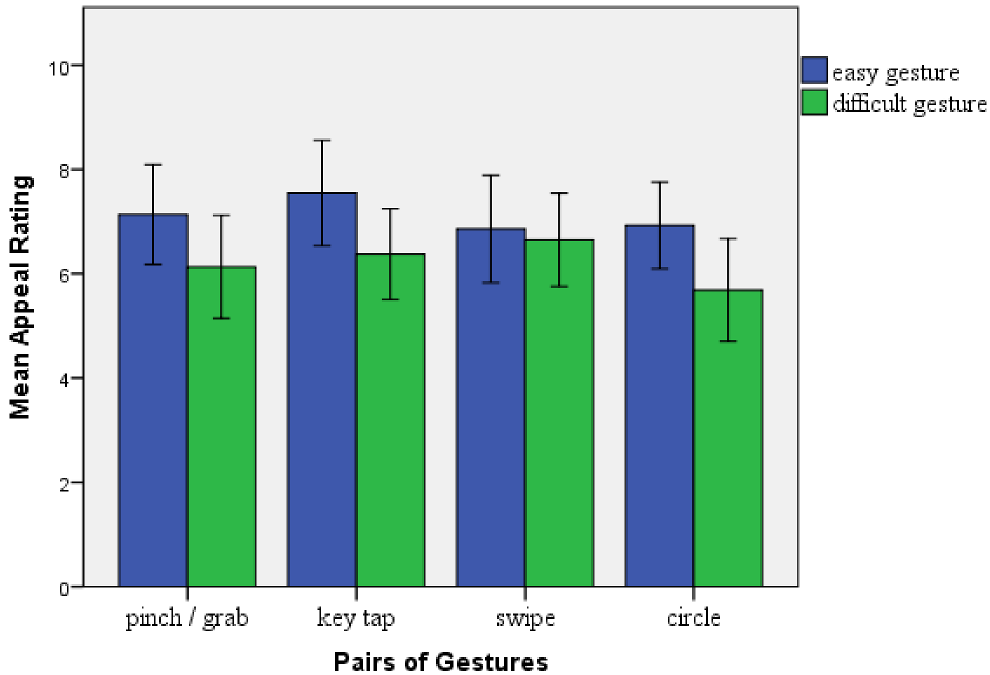

Figure 6. Mean appeal rating for eight gestures from 0 for not appealing
(I hate it) to 10 for very appealing (I like it). Error bars: 95% confidence interval of the mean.

	Pinch/Grab	Key Tap	Swipe	Circle
Z	-1.839	-2.218	-1.147	-2.436
Asymp. Sig. (2-tailed)	.066	.027	.251	.015

Table 6. Wilcoxon signed ranks test to Q2 of each pair of gestures.

Correlation between Effort and User Satisfaction

With the responses for all eight gestures grouped together, a Spearman's rank-order correlation was run to determine the relationship between effort and user preference (Figure 7 shows the scatterplot of a gesture's appeal rating as a function of the perceived effort rating). There was a strong, negative correlation between appeal and effort rating, which was statistically significant, $r_s(230) = -0.54$, $p < 0.001$.

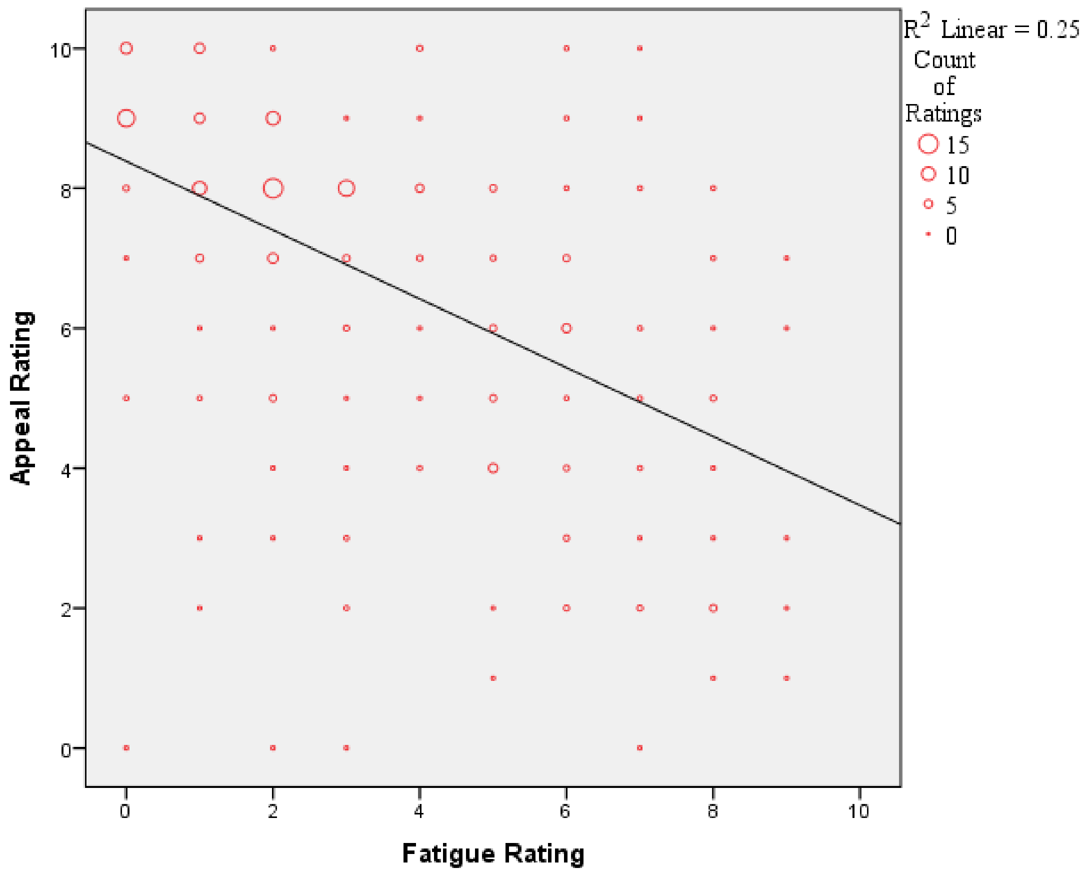

Figure 7. The eight gestures' appeal rating as a function of the fatigue rating.
Stacked plots are represented by the circle size.

Perceived Reliability

Figure 8 presents the perceived reliability, the participants' response to Q4, for all eight gestures. Among all gestures, "swipe left" received the lowest perceived reliability rating (mean = 6.66, standard deviation = 0.80). "Grab" received the highest perceived reliability (mean = 7.83, standard deviation = 0.81).

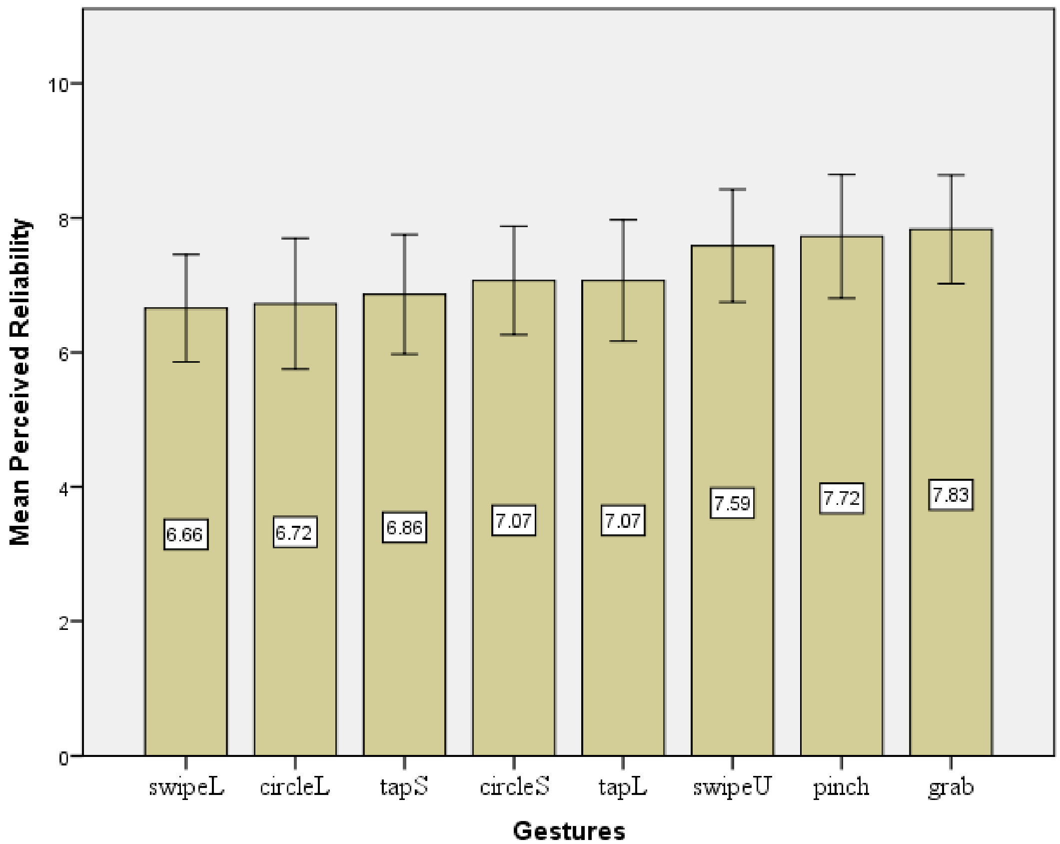

Figure 8. Mean Perceived Reliability for eight gestures. Error bar: 95% confidence interval.

Impact of System Recognition Reliability on User Satisfaction

The Spearman's rank-order correlation between participants' perceived reliability (Q4) and the difference in appeal based on a projected change in reliability (Q2 – Q5) was measured to determine the relationship between the reliability of the recognition system and the willingness of the participant to use the gesture. Figure 9 shows the scatterplot of Q2 - Q5 as a function of Q4. The two variables were negatively correlated at a statistically significant level, $r_s(230) = -0.39$, $p < 0.001$.

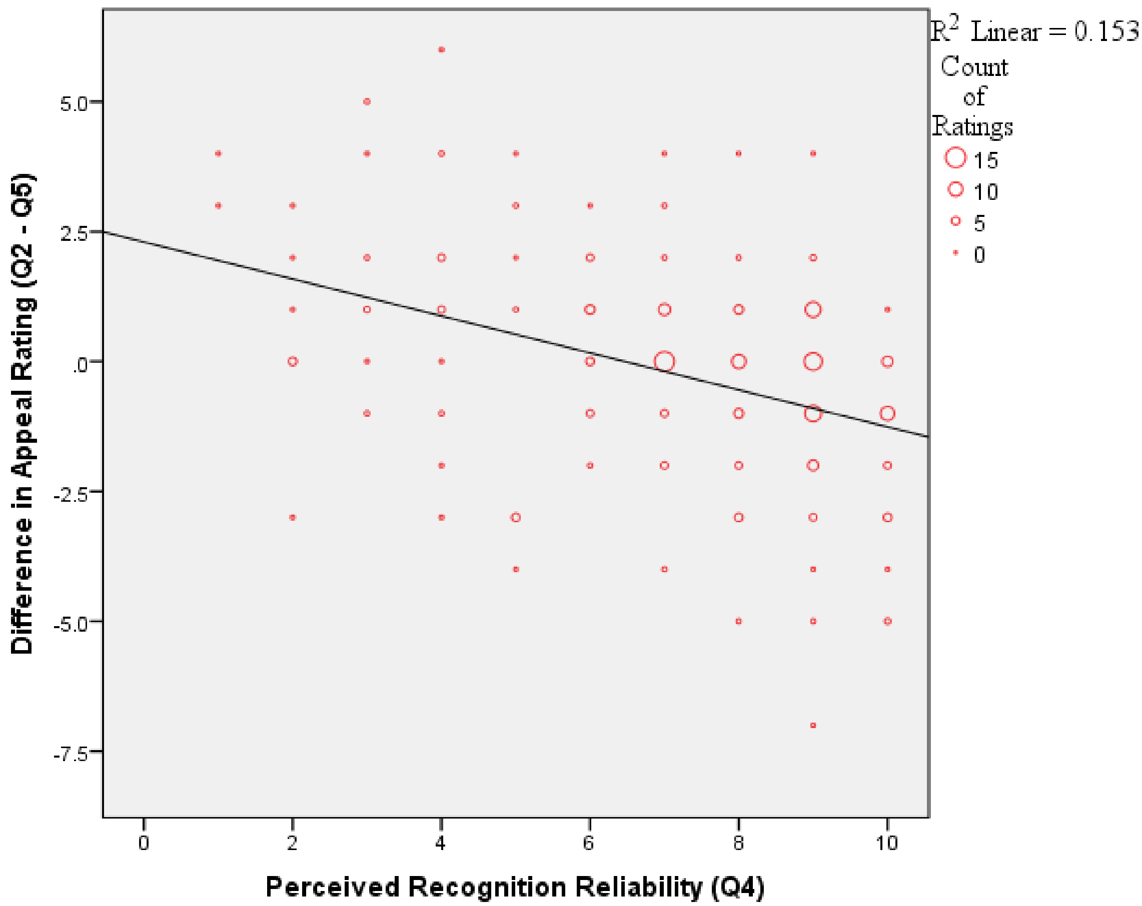

Figure 9. Regression fit plot of projected difference in appeal rating (Q2 – Q5) as a function of perceived recognition reliability (Q4)

Regression on Both Effort and Reliability

Finally, a multivariable linear regression was calculated to predict appeal rating (Q5) based on perceived effort (Q3) and perceived recognition reliability (Q4). As shown in Table 7 and Table 8, the result of the regression indicated that effort and recognition reliability explained a significant amount of the variance in appeal ($F(2, 229) = 99.18$, $p < 0.001$, with an R^2 of 0.464). The analysis showed that both effort and recognition reliability significantly predicted appeal. Specifically, the appeal rating significantly increased as perceived effort decreased ($\beta = -0.33$, $t(230) = -6.47$, $p < 0.001$), and as perceived reliability increased ($\beta = 0.50$, $t(230) = 9.80$, $p < 0.001$).

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.681	.464	.459	1.820

Table 7. Regression summary for appeal rating (Q5) as a function of perceived effort (Q3) and perceived reliability (Q4).

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	p
1	(Constant)	4.219	.500		8.437	.000
	Q3	-.312	.048	-.330	-6.471	.000
	Q4	.540	.055	.500	9.800	.000

Table 8. Parameter estimates of the multivariable regression for appeal (Q5) as a function of both effort (Q3) and reliability (Q4) as predicting variables.

Discussion

Experiment 1 compared the perceived effort, perceived recognition reliability, and participants' willingness to use each gesture in four gesture pairs. The participants' responses to Q1 generally supported the first hypothesis that easy gestures were less tiring than difficult ones. The Wilcoxon signed ranks test on Q1 indicated that pinch, small tap, and small circle were significantly less tiring to perform than their paired difficult gestures. The difference between two swipe gestures were not significant ($p = 0.051$), indicating that these two gestures required comparable effort.

The second hypothesis of this experiment was that users would rate the appeal of easy gestures higher than the appeal of difficult gestures. The Wilcoxon signed ranks test conducted on Q2 supported this hypothesis for the tap and circle pairs, but not for the pinch / grab and swipe pair.

The third hypothesis was that people prefer gestures that take less effort. This was tested with the Spearman's rank-order correlation between a gesture's appeal rating (Q2) and perceived effort (Q1) for all 8 gestures. The strong, negative correlation between these two factors indicated that the more effortful a gesture was, the less appealing it would be.

The fourth hypothesis was that user satisfaction is negatively affected when the system does not consistently and reliably recognize user input. This was revealed in the Spearman's rank-order correlation between the perceived recognition reliability (Q4) and the difference in appeal caused by change in reliability (Q2 – Q5) between the ideal and actual recognition systems. The analysis revealed a negative correlation with an $r_s = -0.39$ ($p < 0.001$), confirming the hypothesis. The r_s value demonstrated a significant, but moderate relationship. Also the direction of the effect indicated that: 1) the more reliably the system recognized a gesture, the smaller the difference between Q2 and Q5; and 2) when gestures were perfectly recognized by the system (when Q4 = 10), participants tended to find the system more appealing than when they simply performed the gesture for the experimenter (Q2 – Q5 < 0).

The final hypothesis was that both effort and system reliability will be important factors in overall user satisfaction with a gesture. This relationship was explored through a multivariable linear regression treating appeal as a dependent variable and both perceived effort and reliability ratings as independent variables. The regression yielded an $R^2 = 0.464$ ($p < 0.001$). It appears that appeal correlates with both perceived effort and reliability. Specifically, the appeal increases with decreases in perceived effort ($\beta = -0.33$, $p < 0.001$) and appeal increases with increases in perceived reliability ($\beta = 0.50$, $p < 0.001$).

Among the four gesture pairs, the differences between effort and satisfaction for the two swipe gestures were not significant. Also the satisfaction rating for the pinch / grab pair was not significant. It may be that the gestures were too similar to reveal the differences observed in the other gestures or that the gestures differed in some undiscovered way that negated the expected effects.

The experiment suffered from several limitations. First, the participants had different fine motor skills. Some participants dexterously manipulated each of their fingers individually without any particular effort, while others found such manipulations to be difficult. Less dexterous participants preferred gestures with larger movement and gestures that were performed with the whole hand (e.g. grab) rather than only finger movements. Because their preference was opposite to the expected effect, the preferences of the less dexterous users reduced the significance level for the Wilcoxon signed ranks tests. However, these participants often rated the difficult gesture, especially grab and circle large as less effortful so it strengthened the negative correlation between effort and appeal rating, which was consistent with our hypothesis that people prefer less effortful gestures. A larger experiment could control for or block user dexterity.

A second limitation of experiment 1 was the relatively small range of effort, compared to full body, or even limb movements, for example. Perhaps 20 repetitions of these hand gestures were not sufficient to reveal perceivable differences in effort for all participants. Some participants had to perform the gestures again when finishing the survey to remind themselves of the difficulty of the gesture. Adding more repetitions might help participants better perceive gesture effort, but would increase the length, tediousness and cost of the experiment.

Finally, although the hypotheses emphasized differences in preference and reliability between the gestures, other facets of the gesture may have affected user preferences. For example, some participants reported that although the grab and large circle gesture took more energy to perform, they would still prefer these gestures to their paired pinch and small circle gestures, because these gestures felt more natural to perform. Naturalness is an interesting psychological facet beyond the scope of this study, but potentially an interesting avenue of research for future studies.

In conclusion, Experiment 1 established that physical effort has an important effect on user satisfaction for gestures. Users preferred the less effortful gestures in two of the three cases where they perceived the difference in effort to be significantly different. In the third case, their preferences were not significantly different. Recognition reliability also impacted user satisfaction. As expected, participants preferred gestures that the system reliably recognized. At last, our data suggested that both effort and reliability affect user preference.

CHAPTER V:
EXPERIMENT 2 – EFFORT AND USER PREFERENCE OF SELECTING GESTURES

Experiment 1 explored the relationship between physical effort and user preference for several gestures useful in constructing commands. Experiment 2 explores these relationships specifically for a three-dimensional tapping gesture useful for selecting objects. In this experiment, participants used the tapping gesture to select among 16 targets in two, three-dimensional task environments. The task environments differed primarily in the physical distance required to navigate from one target to another, using distance as a proxy for physical effort. After completing both sets of 16 selections, participants rated how tiring the experience was and which scenario they preferred

Method

Participants

The participants from experiment 1 also completed this experiment. As in experiment 1, the second cohort performed the procedures in reverse order.

Materials

The same hardware and setup from experiment 1 was used for this experiment. In addition, this experiment involved three-dimensional interaction on an HTML webpage. Three.js JavaScript library was used to create the three-dimensional scenario.

Procedure

The test environment displayed 16, white, spherical targets regularly spaced on a rectangular grid, as shown in Figure 11. A wire frame model of a hand was animated by the participant's hand movements perceived by the Leap system. A round, red selection cursor was rendered just beyond the last joint of the index finger. Shadows of the cursor and targets were projected onto two walls and the ground to provide navigational cues in the three-dimensional space.

Each participant moved his or her hand to control the cursor. If the cursor sphere intersected a target sphere, the target turned to green, then back to white when the cursor moved away. The target sphere was yellow. Selection was performed by making a tap gesture so that, at the lowest point of the tap motion, the cursor intersected the target. After a successful selection, the current target was removed from the environment and the next target sphere was turned yellow. Target order was randomized.

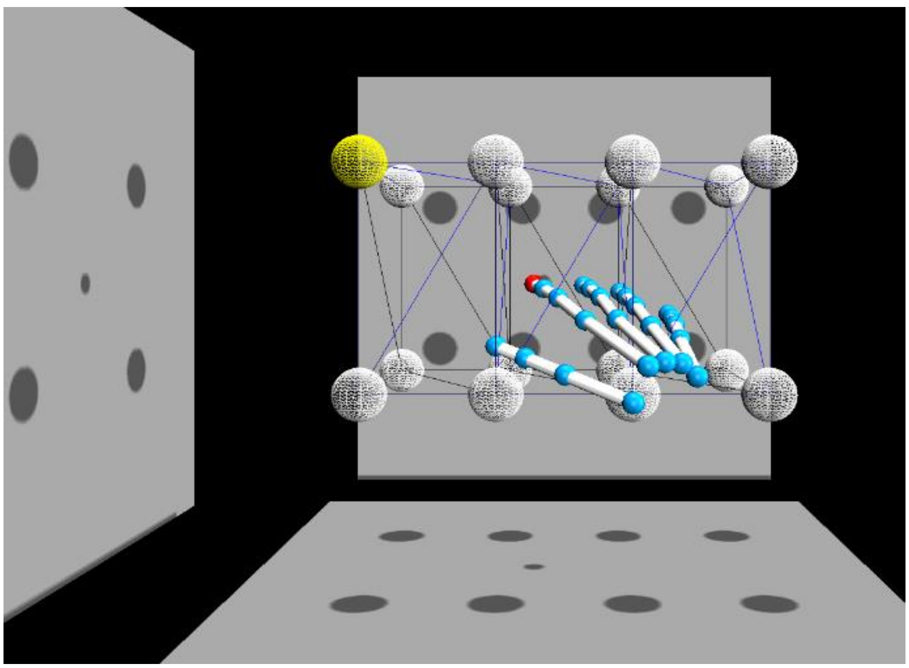

Figure 10. Test scenario of experiment 2

This test was designed to hold selection reliability constant, under two different conditions of physical effort. The two conditions compared in Experiment 2, conditions A and B, are presented in Figure 11 and Figure 12. Selection reliability is the frequency that, given that the system correctly recognized a selection gesture, the cursor intersected the target at the bottom of the gesture motion. Selection reliability is distinct from gesture reliability, the frequency that the system correctly recognizes the user's tapping motion as a tap. Selection reliability is concerned with the success of a correctly recognized gesture, which is mostly related to the location of the gesture rather than the pattern of the motion.

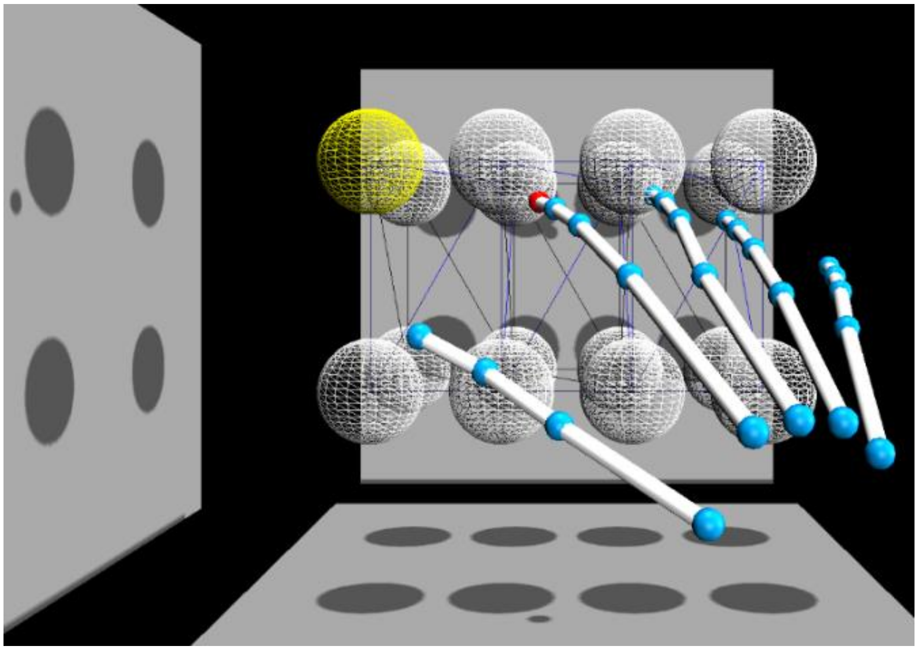

Figure 11. Test scenario of condition A.

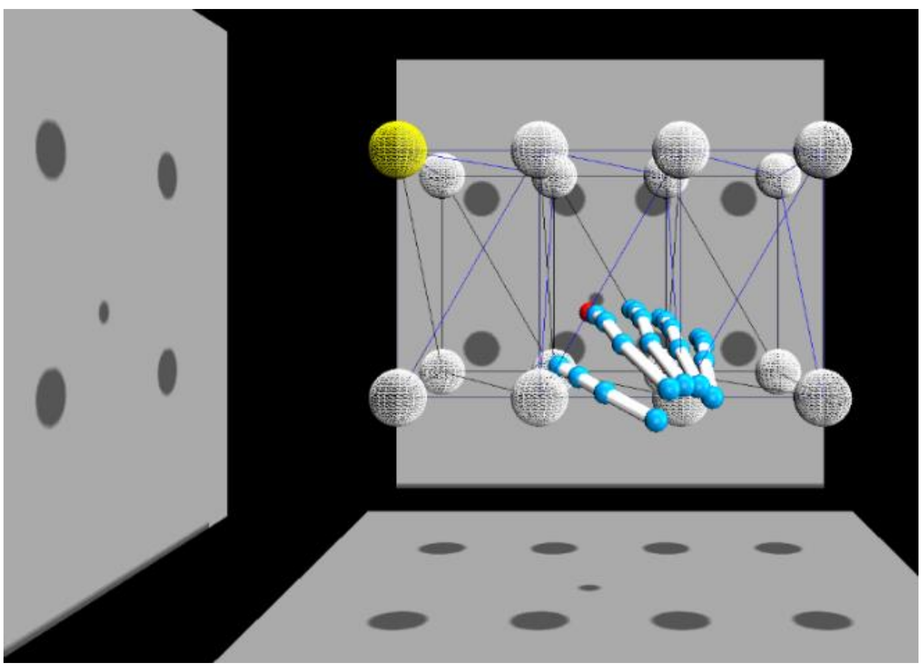

Figure 12. Test scenario of condition B.

Equalizing selection reliability requires that the targets in the two conditions, despite their different visual size and distance, have the same likelihood of being correctly selected. This was achieved by enlarging both the size of the targets and the size of the model hand in condition A (hand size A:B = 2.2:1; target size A:B = 2:1), so that the hand-to-target size ratio, which in turn determined the selection reliability, remained very close in two conditions (hand-to-target size A:B = 1.1:1). The ratios were not set as exactly the same because the performance of the Leap controller is inconsistent across its working space. As Guna et al. (2014) pointed out, the Leap controller is less accurate for samples taken more than 250mm above the controller's surface. In this test, the rectangular grid where the targets located covered an approximate 5.5 x 3.2 x 3.7 inches' space in real world for condition A, and roughly a 12 x 7 x 8 inches' space in real world for condition B. In other words, condition A used only the very center area of Leap's working space, thus obtaining better sensor performance. To offset this performance inconsistency, we slightly increased the hand size to make the target a bit more difficult to click. Specifically, clicking a target in condition A was like clicking a 1.46-inch diameter ball in the real world, whereas the targets in condition B was equivalent to 1.6-inch diameter balls in the real world.

While the selection reliability was kept equal in both conditions, movement distance required to finish the selection task was 2.2 times larger in B than A. Because moving a shorter distance requires less energy, A was expected to be less effortful.

When participants finished the selection task in both conditions they were asked to complete a survey to rate their experience. Survey questions were:

Q1. Please rate how physically tiring this task was on a scale from 0 representing not at all tiring to 10 for very fatiguing.

Q2. Did you experience difficulties in finishing the tasks precisely? Please rate how difficult / ambiguous the interaction was on a scale from 0 for not difficult to 10 for very difficult.

Q3. How appealing is this interaction for the given task? Are you satisfied with the interaction design? Please rate your feeling on a scale of 0 for not appealing (I hate it) to 10 for very appealing (I like it).

As an objective measurement of selection reliability, the number of selection gestures each participant performed was also recorded. For example, consider a participant who tapped 100 times to complete the 16-ball selection task, with 36 taps successfully recognized by the system as tapping gestures. This participant's total attempt count would be recorded as 36, including 16 successful selections and 20 unsuccessful selection gestures. Because there are always 16 successful selections to complete the task, fewer selection gestures are associated with higher selection reliability.

The specific hypotheses for experiment 2 were:

1. The selection reliability is not significantly different in two conditions. This hypothesis is tested with a paired t-test of each participant's total attempts on condition A and B. The null hypothesis is that there is no difference in selection reliability between condition A and B.

2. Participants perceive condition A as less effortful. This hypothesis would be tested with a Wilcoxon signed ranks test on the difference in participant responses to Q1 on each of the two conditions. The null hypothesis is that there is no difference in perceived effort between two conditions.

3. Participants prefer condition A, the less effortful condition. A Wilcoxon signed ranks test of the participant responses to Q3 would confirm this hypothesis. The null hypothesis is that there is no difference in the user preference between conditions.

4. Effort affects user's preference for an interface; participants prefer interactions that require less effort. A correlation between Q3 and Q1 will test this hypothesis, with a significant negative correlation confirming the expected relationship.

Results

Objective Selection Reliability Measurement

The selection reliability was calculated by dividing the number of targets (16 in this test) by the number of recognized attempts for each trial. The difference between condition A and B was tested with a paired t-test. Although close to significant, the difference was not significant at the p = 0.05 level.

	Paired Differences			t	p
	Mean	Std. Deviation	Std. Error Mean		
Correct Rate A - B	-.04	.12	.02	-1.82	.079

Table 9. Result of paired t-test in success rate between condition A and B.

Participants' Response to Survey Questions

Mean participant response to survey question Q1 through Q3 for both conditions is presented in Figure 13 and Table 10. Three separate Wilcoxon signed ranks tests showed that there were significant differences between condition A and B in perceived effort (Q1), $Z = -4.35$, $p < 0.001$, perceived accuracy (Q2), $Z = -3.49$, $p < 0.001$, and appeal (Q3), $Z = -3.95$, $p < 0.001$.

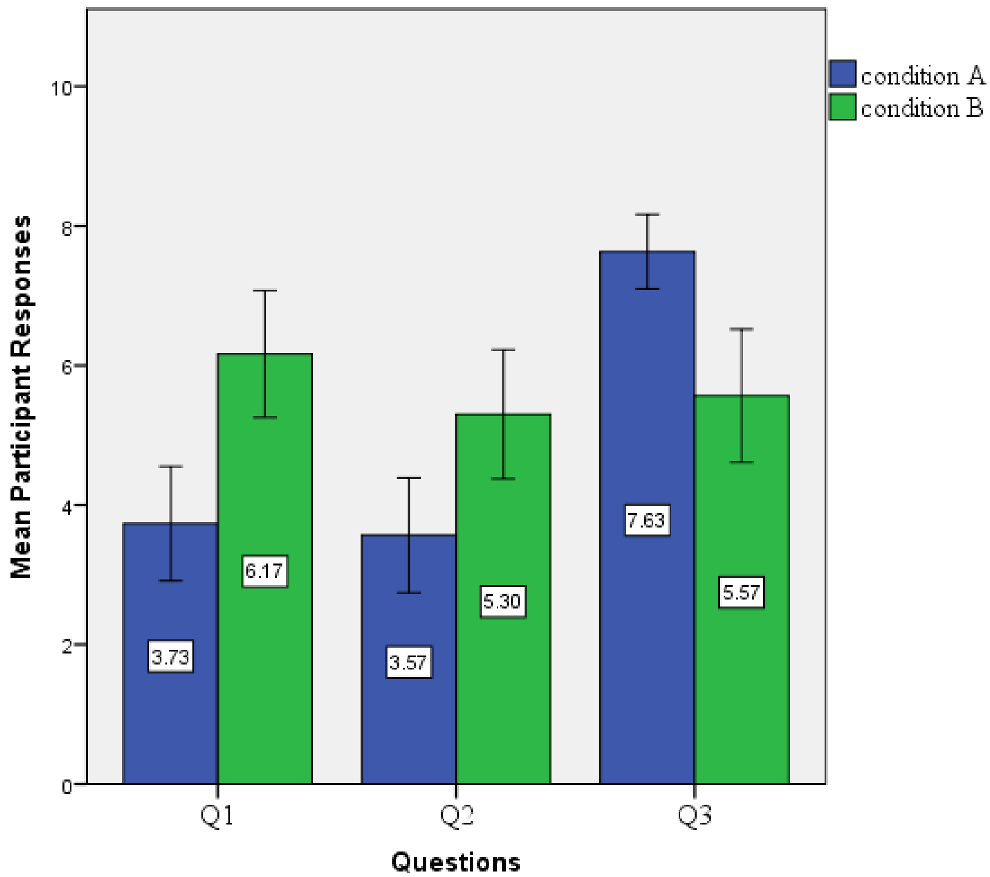

Figure 13. Mean participant responses to Q1 through Q3 in experiment 2.
Error bar: 95% confidence interval.

	Q1A	Q1B	Q2A	Q2B	Q3A	Q3B
Mean	3.73	6.17	3.57	5.30	7.63	5.57
Std. Deviation	2.20	2.44	2.21	2.48	1.43	2.56

Table 10. Means and Standard Deviations for the participants' response to Q1 through Q3 in experiment 2.

Correlation between Effort and User Satisfaction

To test the hypothesis that people prefer interactions that are less effortful, a Spearman's rank-order correlation with samples from both conditions was run to determine the relationship between the participants' perceived effort (Q1) and user preference (Q3). Figure 14 shows the plot of user satisfaction as a function of perceived effort. There was a strong, negative correlation between the two variables, which was statistically significant, $r_s(58) = -0.71$, $p < 0.001$, indicating that the more effortful a gesture was, the less likely it would be preferred by the participants.

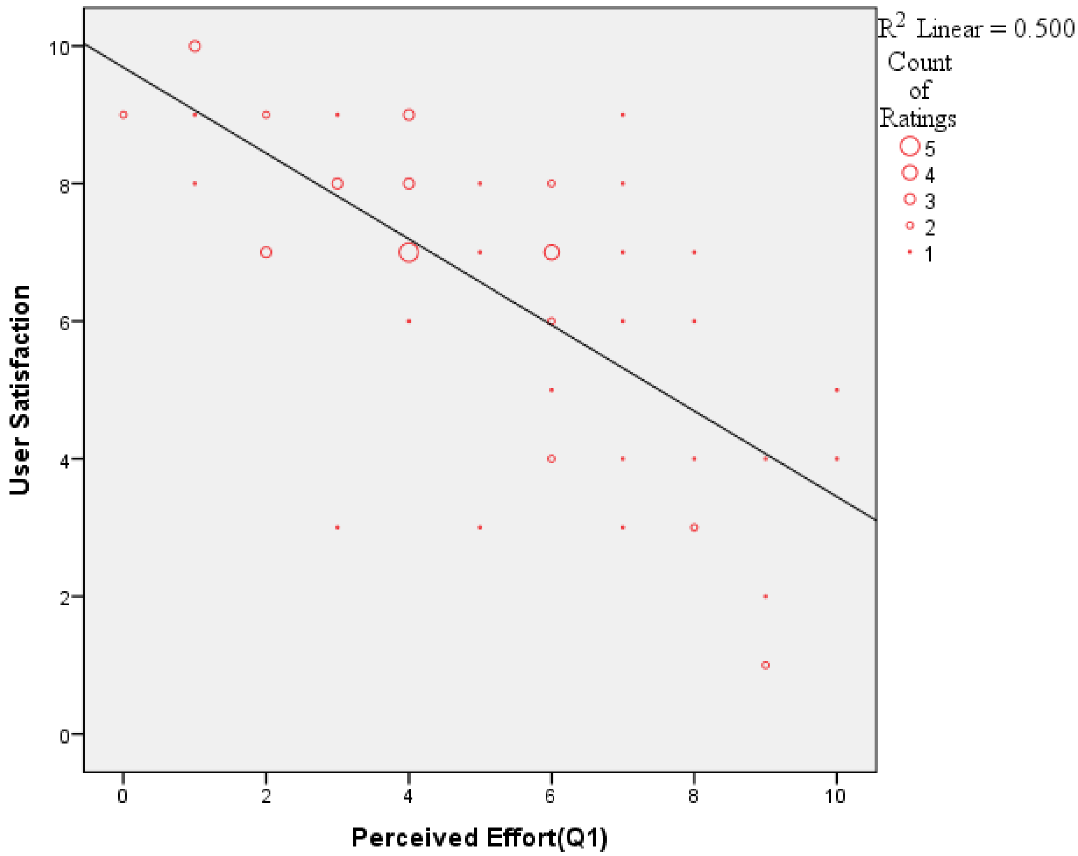

Figure 14. Regression of user satisfaction as a function of perceived effort in experiment 2.

Discussion

This experiment explored user preferences for two conditions with similar selection reliability but different levels of gesture effort. To test that the selection reliability was equal in both conditions, a paired t-test compared the total number of recognized gesture attempts in each condition for each participant, with the expectation that there would be no difference. Although the difference was not significant, the p-value was small enough ($p = 0.079$) to draw attention. Participants tended to have more unsuccessful selections in the low-effort condition. Further analysis of this result suggests that this may have been caused by a design decision when creating the experiment.

As described previously, the target and hand were both enlarged in condition A, while the hand-to-target ratio was kept constant, in order to achieve consistent selection reliability. However, an element of the skeleton model, specifically the diameter of the cursor at the fingertip of the model, was not enlarged (See Figure 11 and Figure 12, note how “thin” the hand is in condition A). This dimension was chosen to avoid visually obstructing the test and because participants in the pilot test disliked the larger hand dimensions, which were varied with the cursor dimension. In retrospect, it is apparent that because a successful selection was determined by the collision between the cursor and the target, a larger cursor would have increased the consistency in selection reliability between the two conditions. Fortunately, the effect size was small enough to not be significant.

Our hypothesis that the shorter movement distance in condition A would lead to a less tiring interaction was confirmed by the Wilcoxon signed ranks test of participants’ responses to Q1. The perceived effort of condition A was significantly lower than that of condition B. A separate Wilcoxon signed ranks test showed that there was a significant difference in participants’ responses to Q3, which asked how satisfactory the interaction was. The less effortful condition (a.k.a condition A), enjoyed a significantly higher user satisfaction rating.

More importantly, there was a strong correlation between the perceived effort (Q1) and overall satisfaction (Q3), which supported our hypothesis that people prefer interactions that are less effortful. This provides evidence that for pointing and navigational gesture tasks, effort appears to be an important factor to user satisfaction.

CHAPTER VI:
EXPERIMENT 3 – USER BALANCE BETWEEN EFFORT AND RELIABILITY

The results of experiment 1 and 2 establish that effort is an important factor of a user's overall satisfaction to a gesture interface. It is also clear that people appreciate reliable systems. It remains unclear how people behave when faced with an effort-reliability tradeoff. We hypothesize that people balance the two factors without going to either extreme. Thus, in this experiment we present 10 conditions with different effort-reliability tradeoffs and ask participants to decide which tradeoff they prefer. Then we describe this balance using different measures obtained from the test.

Method

Participants

The participants from experiments 1 and 2 also took part in this experiment. Unlike the first two experiments, all participants followed the same test procedure.

Materials

This experiment employed the same scenario and task as experiment 2. The target size and visual distance between targets on the screen were consistent, but there were 10 different settings for hand size (see Figure 15).

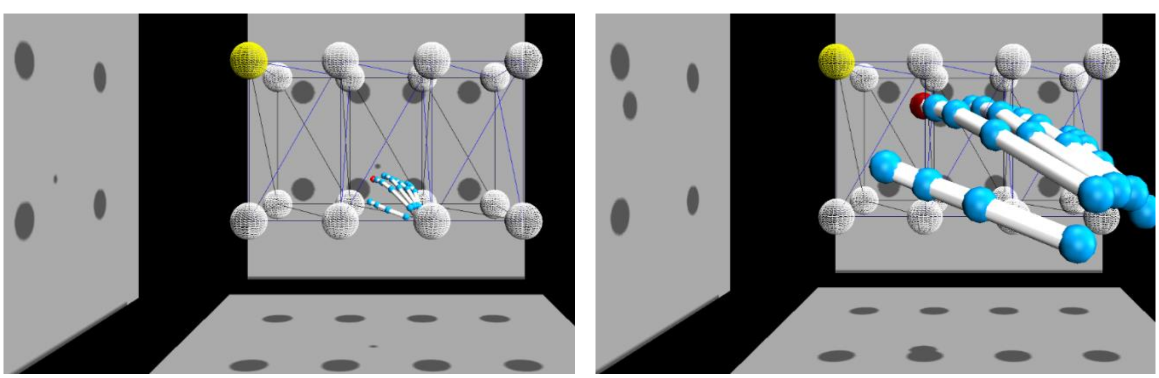

Figure 15. Scenario of the smallest (option 1, left) and largest (option 10, right) hand size settings for experiment 3. There are 8 other settings in between.

When the hand is very small, participants must move their hand a long distance to finish the task, but they enjoy greater control over the cursor position and it is fairly easy to tap the target precisely due to the relatively big target. When the hand is big, participants need only move their hand a short distance, but the target is relatively small and it is difficult to select the interior of the target precisely. It is evident from experiment 2 that the longer the movements required to finish a task, the more effort the interface will require. Thus, as the hand size increases from option 1 through 10, the effort needed to finish the task decreases, while the difficulty to select a target increases. The equivalent real world sizes of the movement and size of the targets in different options are presented in Table 11.

Option	**Relative Scale**	**Work Envelope (inches)**			**Target Diameter (inches)**	**Cursor Diameter (inches)**
		x	**y**	**z**		
1	0.8	15.00	8.75	10.00	1.20	0.43
2	1.1	10.91	6.36	7.27	0.87	0.43
3	1.5	8.00	4.67	5.33	0.64	0.43
4	1.9	6.32	3.68	4.21	0.51	0.43
5	2.3	5.22	3.04	3.48	0.42	0.43
6	2.7	4.44	2.59	2.96	0.36	0.43
7	3.1	3.87	2.26	2.58	0.31	0.43
8	3.5	3.43	2.00	2.29	0.27	0.43
9	3.9	3.08	1.79	2.05	0.25	0.43
10	4.3	2.79	1.63	1.86	0.22	0.43

Table 11. The equivalent real world size of the movement and size of the targets of different options in Experiment 3. Numbers are in inches, except relative scale.

Procedure

During the experiment each participant was instructed to “imagine this is one of your daily applications in which you need to interact with the computer by clicks. Among the ten hand sizes, which one is most preferable?" Participants were then allowed ample time to familiarize themselves with all settings and pick the one they preferred.

Once participants had selected a preferred setting, they were asked to finish in that specific setting the selection task described in experiment 2. We measured the total attempts to select all 16 balls, and upon completion, asked them to answer the following survey questions:

Q1. Please rate how physically tiring this task was on a scale from 0 representing not at all tiring to 10 for very tiring.

Q2. Did you experience difficulties in finishing the tasks precisely? Please rate how difficult / ambiguous the interaction was on a scale from 0 for not difficult at all to 10 for very difficult.

The hypothesis for this experiment was that participants would choose neither of the extreme conditions, and tended to find a compromised point between effort and system reliability. The hypothesis would be verified by a positive correlation between perceived effort and system reliability established in the user preferred options.

Results

Participant preference over 10 options is presented in Figure 16. Option 1, 8, and 10 were never selected as the most preferable option. Option 2 and 9 were selected only once. It appeared that participants tended to prefer option 4 through 7 over others (mean = 5.03, standard deviation = 1.61).

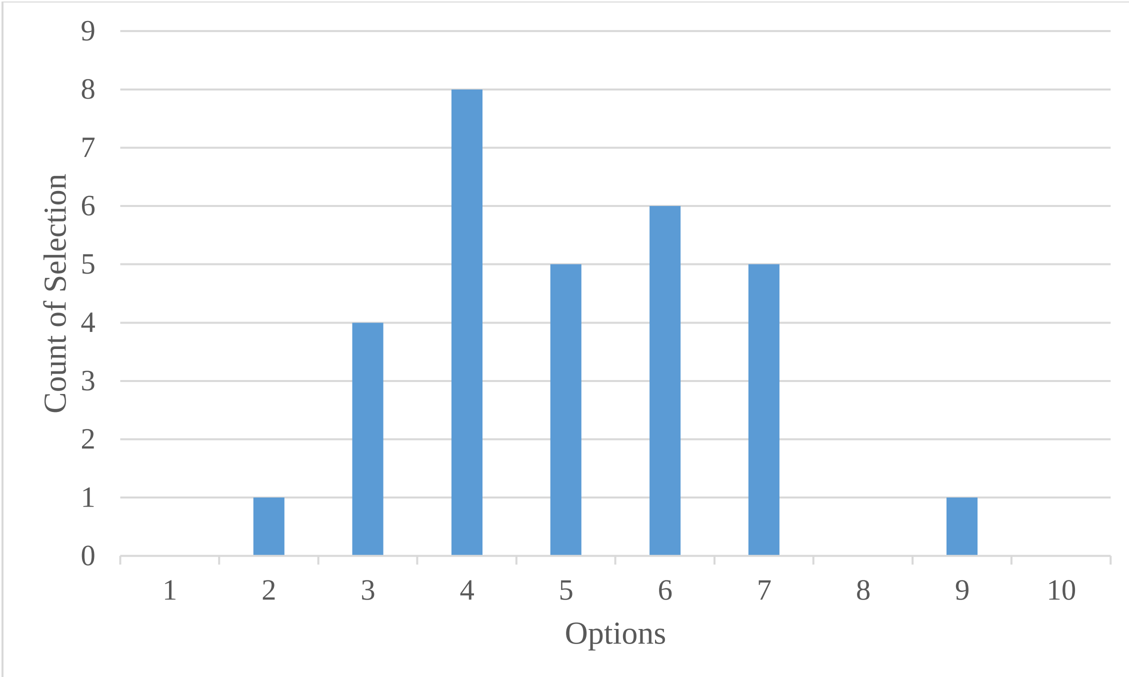

Figure 16. Distribution of the most favorable option in experiment 3

The aim of this experiment was to find out how participants balanced effort with reliability when selecting a preferred option. The total attempts and the responses to survey questions reflected participants' experience and evaluation in this matter. Table 12 displays these results.

	Total Clicks	Q1 - Effort	Q2 - Difficulty
N	30	30	30
Mean	23.10	4.23	4.40
Median	22.00	4.50	4.00
Std. Deviation	4.759	2.192	2.143
Range	18	7	10
Minimum	17	1	0
Maximum	35	8	10

Table 12. Descriptive Statistics of total clicks, survey questions 1 and 2 for experiment 3.

There was a strong positive correlation between effort (Q1) and difficulty (Q2), $r_s(29) = 0.69$, $p < 0.001$ (see Figure 17 and Table 13, described in Discussion).

		Q2
Spearman's rho	Q1	.686
p		.000
N		30

Table 13. Spearman's correlation between Q1 and Q2 for experiment 3.

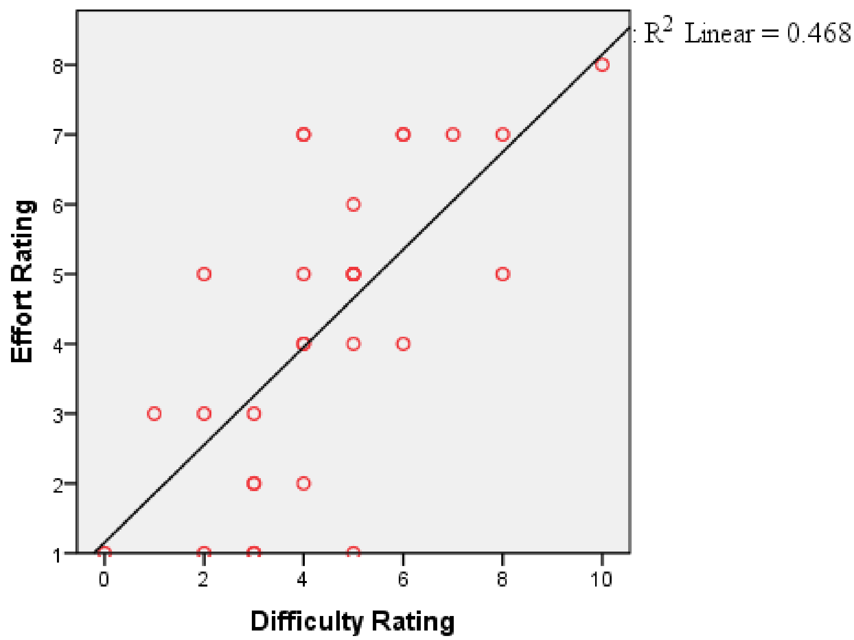

Figure 17. Plot of effort rating (Q1) as a function of difficulty rating (Q2)

Discussion

This experiment presented 10 different levels of effort-reliability tradeoffs. Participants decided which tradeoff they preferred. Clearly, participants did not choose either of the extreme conditions. Twenty-four out of thirty (80%) participants chose option 4 through 7.

Our hypothesis that there is a point that makes users comfortable in dealing with the effort-to-reliability tradeoff was supported by the *positive* correlation between participants' responses to Q1 and Q2. Q1 reflected the participants' estimation of how tiring the interaction has been. Q2 represented participants' attitude on how difficult the task was to be completed precisely, or the reverse of the system reliability. In the coordinate with difficulty as the x-axis and effort as the y-axis, as shown in Figure 17, it is expected that each participant's rating for option 1 through 10 would distribute along the direction from the top left corner, indicating a tiring but reliable option, to the bottom right corner, indicating a less effortful but also less reliable option. The points in Figure 17 each reflect a participant's balance in the matter of the effort-to-reliability tradeoff, and were distributed along the fit line for a positive correlation. This positive correlation evidences that participants were looking for a balance between the two factors—they would prefer a condition that was less tiring (lower Q1) in case they could finish the task without too much ambiguity (lower Q2); and they would sacrifice some effort and go for a more effortful setting (higher Q1) only if the system did not allow them to finish the task reliably (higher Q2). If this was not the case, the participants should have preferred a different option where they must sacrifice more effort to gain better reliability, or vice versa.

This balanced point between effort and reliability can be described by either measurement. The absolute value of effort is hard to measure. Rather, participants' perceived effort (mean = 4.23, standard deviation = 2.19) could be a measurement for this balanced point. An alternative, more objective measurement would be the selection reliability defined in Experiment 2, calculated by dividing successful selection gesture by total recorded selection attempts (mean selection reliability: 16 / 23.1 = 69.26%). These numbers are expected to have a predictive power in the development and evaluation of future gesture interfaces. Although these measurements could vary greatly for different system, at least for the selection tasks in our specific test condition, we quantify our findings as:

1) The users did not want to use a system that was perceived as more effortful than the level of “neutral” (in this study the point of 5 in our 10-point scale was marked as “neutral”). Preferably the user’s perceived effort to use the system should be below or close to 4.23 in a 10-point scale.

2) The users did not like a system that had selection reliability lower than 70%. Preferably a gesture interface should allow its user operation to be reliably executed to maintain a success rate of at least 70%.

CHAPTER VII:
EXPERIMENT 4 – THE EFFECT OF POINTING ON SPATIAL WORKING MEMORY IN A 3D GESTURE INTERFACE

Experiment 1 through 3 examined how effort and recognition reliability affected user satisfaction of a gesture interface. Experiment 4 examines the influence pointing has on spatial working memory by comparing participant's recall in *move* and *no-move* conditions in our three-dimensional virtual environment. In the move condition, participants are asked to point to, and memorize a series of locations. In the no-move condition, participants are asked to passively view, and memorize the spatial array. The accuracy of each participant's recall and the time taken during the recall phase were recorded for analysis.

Method

Participants

Thirty-three University of Iowa students and staff members (13 males, age Mean = 29.3, SD = 9.3) underwent individual 35-minute sessions in exchange for a $5 gift card. All participants had normal or corrected-to-normal vision, and were able to make hand gestures for daily communication purposes without difficulties. One participant didn't finish the test and the data were discarded.

Materials

The experiment, programmed in JavaScript, was individually conducted on a PC computer with a 12.5-inch IPS display at a resolution of 1366 x 768 pixels in a room with soft light. The experiment was presented in the Chrome web browser in full-screen mode. The display was adjusted to each participant's eye level. Participants were seated approximately 20 inches from the computer screen. The Leap Motion controller, which captured participants' hand movements, was placed in front of the participant to allow them to comfortably point to a target as needed.

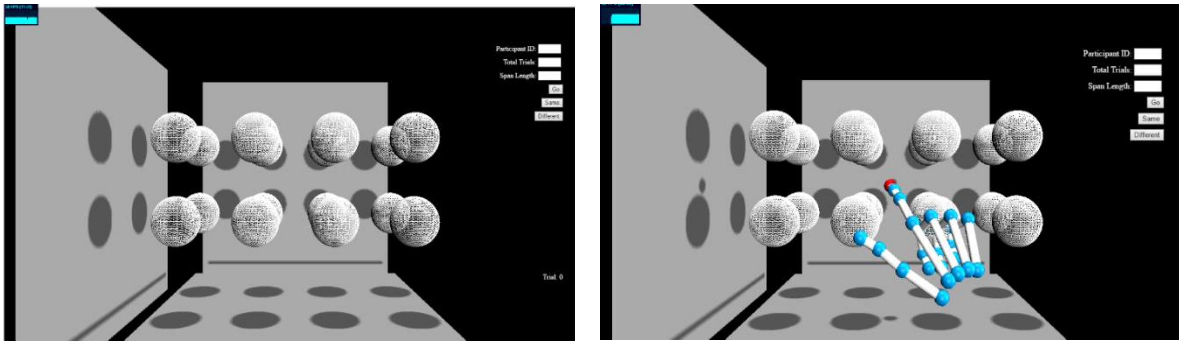

Figure 18. Test Scenario. Left - when hand is not in use. Right - when participants put a hand on top of the sensor, a model was animated in the scenario.

The test interface, as shown in Figure 18, contains 16 white, semi-transparent balls that are arranged on a 4 x 2 x 2 basis on the x, y and z axes. A wire frame model of a hand was animated by the participant's hand movement as perceived by the Leap controller, when a gesture was required. A round, red selection cursor was rendered just beyond the last joint of the index finger. Shadows of the cursor and targets were projected onto two walls and the ground to provide navigational cues in the three-dimensional space. Participants moved their hand to control the cursor for the pointing action. If the cursor sphere intersected a target sphere, the target flashed in green once as a visual feedback. The three-dimensional virtual environment projected to a 12 x 3 x 3-inch space in the real world. For example, to move the cursor from the ball at the top left corner to the ball at the top right corner, one needed to move his or her fingertip by 12 inches to the right in the real world.

Procedure

The basic sequence of events is depicted in Figure 19. Each trial included a study phase, a mask, and a test phase. In the study phase of each trial, one spatial array consisting of six items was presented. The items in each array were presented sequentially at one of 16 possible target areas, and their locations were randomized with the constraint that no item occurred in a location previously occupied by another item. In the move condition, the items were marked as red, whereas in the no-move condition, yellow was used to indicate the to-be-remembered locations. In both conditions each item was presented for 1300ms, then resumed to normal status (turn back to white) for 200ms, followed by the next item.

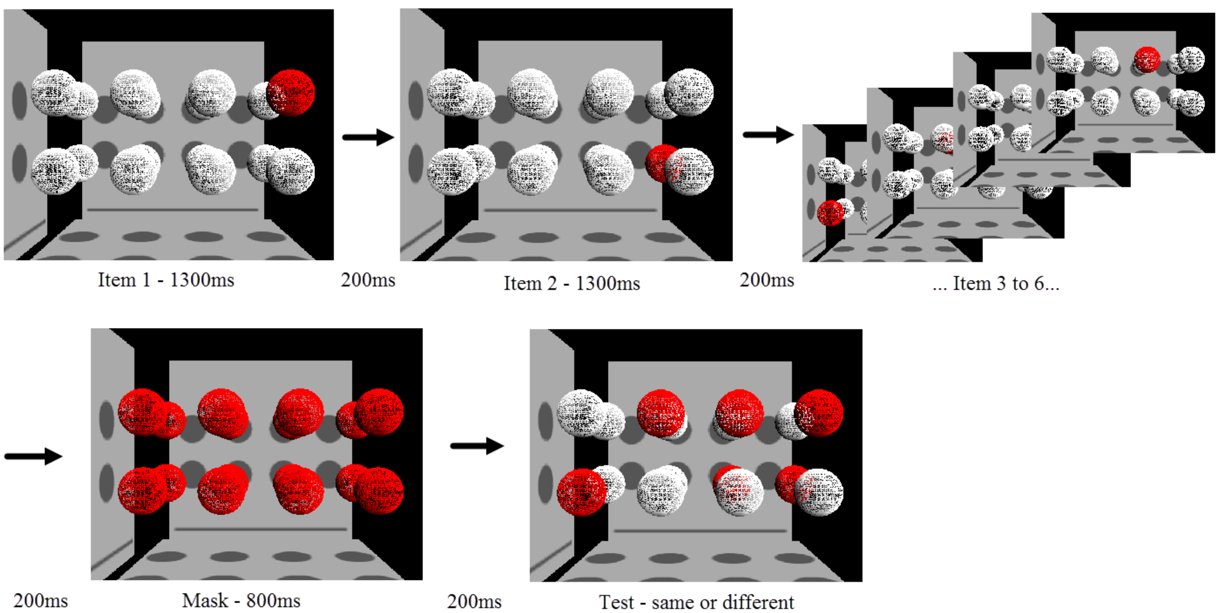

Figure 19. Depiction of a single pointing condition trial, with a study phase consisting of 6 items, a mask, and a test phase. Participants were asked to point to the highlighted balls. Hand model not included in this figure.

In the move condition, participants were instructed to remember the spatial locations of the items while at the same time pointing to the locations of each item during the time it was presented. Each move trial was preceded by a warning text "Point to Targets!" for 1500ms in red to get participants ready for the action. In the no-move condition, on the other hand, participants were instructed to simply watch the sequentially presented items and attempt to remember their locations. No-move trials were preceded by a yellow warning text of “Watch, Do Not Point!” for 1500ms, indicating that participants had to passively observe the spatial array. The move / no-move trials were mixed and their order was randomized separately for each participant. The only constraint was that no more than three consecutive trials of the same type could appear in a row.

Immediately after the study phase, there was a mask in which all possible balls were highlighted for 800ms. Then a test array was presented containing 6 items. Participants had to judge whether the items presented in the test array matched what had been seen during the study phase. Response of “same” or “different” was given by pressing “s” or “d” key on the keyboard. The participants were given a 6000ms time limit to respond. The response time, measured in millisecond between the moment the test array was presented and the moment either “s” or “d” key was pressed, was recorded.

Each participant finished 100 trials in total. Half of the trials were move trials. In both move and no-move conditions, half of the test arrays matched what had been presented during the study phase (i.e. the correct answer for that trial was “same”) while the other half of the trials had one item shifted to a previously unoccupied position. It is worth noting that each participant was given ample time to practice the pointing gesture in the gesture interface until they felt ready for the test. Then a warm-up session including 8 to 12 trials which were identical to the formal experimental trials were conducted to get them familiarized with the test.

Design

This experiment followed a within-subject design. Participant performance, under both move and no-move conditions, was measured by the recognition accuracy, which was computed as the percentages of trials correctly recognized. In addition, the response time of each trial, measured from the moment that the test array was presented to the moment that an answer was given, was recorded to determine each trial's cognitive load.

Results

Mean recognition accuracy for both move and no-move conditions is shown in Figure 20. A Shapiro-Wilk test confirmed that participant recognition accuracy was normally distributed. A paired t-test revealed a significant difference in the accuracy for move (mean = 79.4%, standard deviation = 8.29%) and no-move (mean = 87.5%, standard deviation = 6.28%) conditions; $t(31) = -5.36$, $p < 0.001$, with participants exhibiting better memory during the no-move condition.

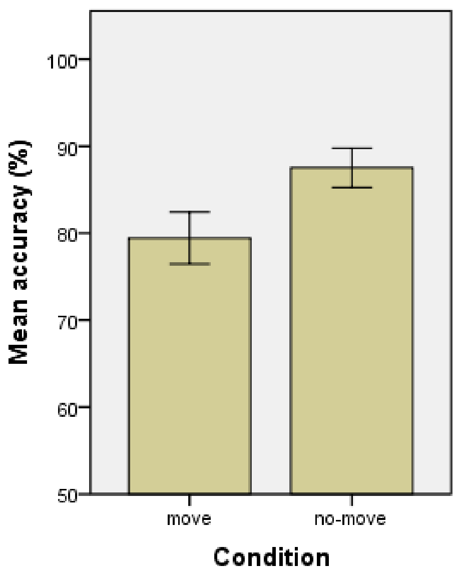

Figure 20. Recognition accuracy of move and no-move condition. Error bars depict 95% confidence intervals.

To examine participants' change in performance over time, mean recognition accuracy as a function of condition and experiment progress, depicted by trial numbers, is presented in Figure 21.

Figure 21. Recognition accuracy as a function of experiment progress (grouped by 5 trials).

During the test phase, participants had to determine whether the test array presented matched the previously-presented test array. To determine whether performance differed between matching test arrays and arrays that had shifted one position, mean recognition accuracy based on different condition and test array types is illustrated in Figure 22. Data were then analyzed with a 2 (condition: move vs. no-move) x 2 (test array type: same vs. different) repeated measures ANOVA. There was a significant main effect of condition, $F(1, 31) = 27.86$, $p < 0.001$, indicating the participant performed better under the no-move condition. There was also a significant main effect of test array type, $F(1, 31) = 60.04$, $p < 0.001$, indicating participants made more mistakes when the test array was different than the study array (when the correct answer should be "different"). Finally, there was a significant interaction between condition and test array type, $F(1, 31) = 25.79$, $p < 0.001$.

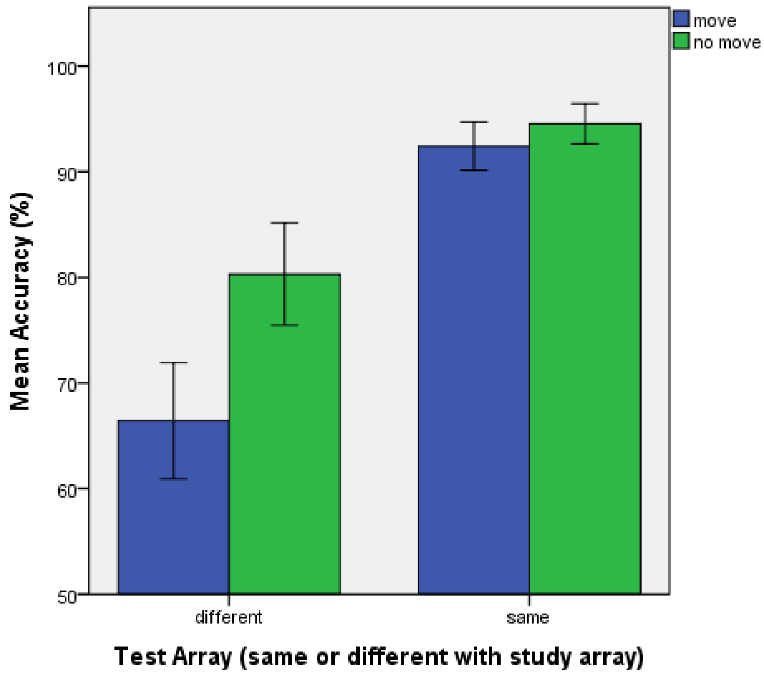

Figure 22. Recognition accuracy as a function of pointing instruction and test array type. Error bars depict 95% confidence intervals.

For all unmatched trials, in which the test array had one item shifted from the study array, the mean recognition accuracy as a function of condition and the serial position of the shifted item is shown in Figure 23. A two-way repeated measures ANOVA revealed that there was a significant main effect of condition, $F(1, 27) = 22.87$, $p < 0.001$. There was also a significant main effect of serial position, $F(5, 135) = 3.78$, $p = 0.003$, and the interaction of condition and serial position, $F(5, 135) = 4.07$, $p = 0.002$. Post hoc comparisons with the Bonferroni adjustment showed that under the move condition, the accuracy when the 6^{th} position was shifted was significantly higher than when any of the 1^{st} – 4^{th} item was shifted ($p < 0.04$ for all factors). No other differences achieved significance.

Figure 23. Recognition accuracy as a function of condition and serial position of the item shifted in unmatched trials (in which test array did not match study array)

Due to the 3D nature of this experiment, we were able to determine whether the shift of an item on the z-axis had an effect on recognition accuracy. Figure 24 shows the mean recognition accuracy as a function of condition and shift direction (shifts on the z axis vs. shifts on the x and y axes only). A two-way repeated measures ANOVA showed no significant difference between different shift directions.

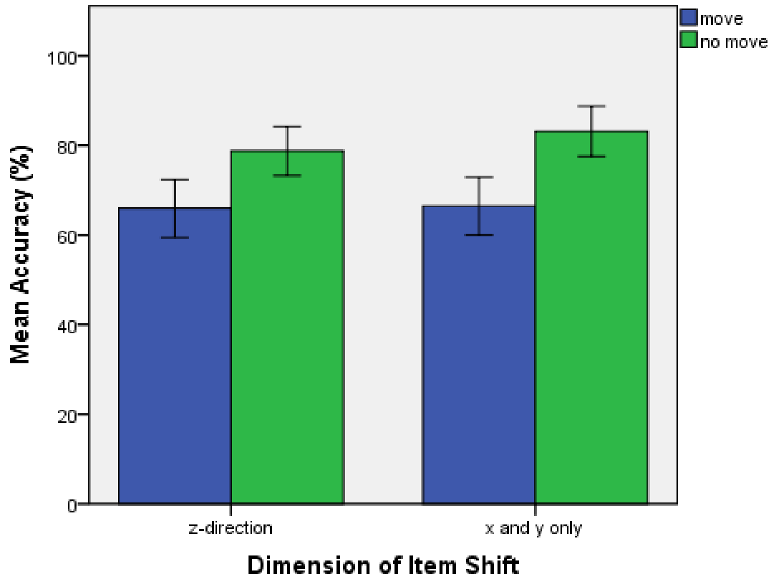

Figure 24. Recognition accuracy of unmatched trials (where test array did not match study array), divided by the direction of the shifted item (whether the shifted item was shifted along the z-axis, or along the x and y axes only). Error bars depict 95% confidence intervals.

Mean response time in milliseconds for both move and no-move conditions are shown in Figure 25. In the move condition, participants took more time to judge whether or not the test array matched the study array. A Shapiro-Wilk test was conducted and confirmed the normality of the data. A paired t-test then showed a significant difference in the response time for move (mean = 1922, standard deviation = 396) and no-move (mean = 1847, standard deviation = 374) conditions; $t(31) = 3.7$, $p = 0.001$.

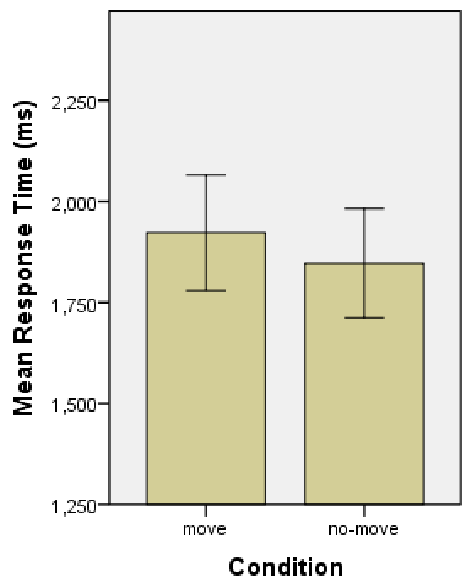

Figure 25. Mean response time of move and no-move conditions. Error bars depict 95% confidence intervals.

Researchers have proposed the use of time as a measure of cognitive effort (Cooper-Martin, 1994; Koponen, Aziz, Ramos, & Specia, 2012). Mean response time, grouped by correct and incorrect recalls for both move and no-move conditions are shown in Figure 26. Data were analyzed with a 2 (pointing condition: move vs. no-move) x 2 (recognition correctness: correct vs. incorrect) repeated measures ANOVA. There was a significant main effect of correctness, $F(1, 31) = 64.42$, $p < 0.001$.

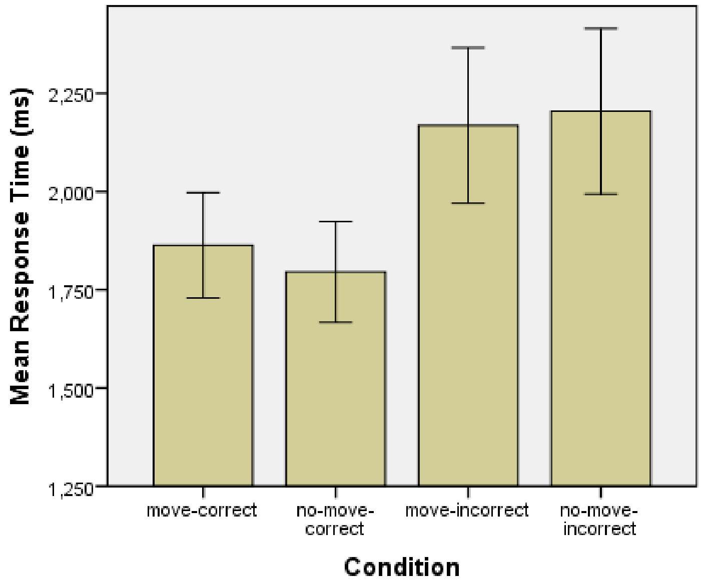

Figure 26. Mean response time of move and no-move condition, divided by correct and incorrect responses. Error bars depict 95% confidence intervals.

Discussion

The aim of this experiment was to determine how pointing influences spatial working memory in a three-dimensional virtual environment. We found that when participants were asked to point to the to-be-remembered locations, their recall of those locations was worse than when they were asked to only watch and memorize the locations. Interestingly, when the test array matched the study array, participants were able to achieve a relatively high recognition accuracy, and the difference between move and no-move conditions was not significant ($p = 0.111$), as shown in Figure 22. In contrast, when the test arrays did not match the study arrays, the recognition accuracy was significantly lower, and there was a significant difference between move and no-move conditions. Participants tended to answer “same” more when they were unsure about a trial. This was as expected since only one item out of six was shifted for unmatched arrays, making the shift hard to detect in some cases. As a result, the unmatched trials seemed to be more sensitive in determining the effects on memory.

There was a change in performance over time. The recognition accuracy was lowest at the beginning of the test, especially for the move-condition. Although warm-up sessions were given prior to formal tests, participants needed more trials to become acquainted with the tasks. After about 20 trials, participants' performance became relatively stable, but did not rise further as the experiment progressed. In fact, recognition accuracy decreased slightly as the experiment proceeded, until finally increasing during the last 10 trials. We argue that the decrease in performance was due to the fact that participants were growing tired for the highly focused tasks. The experimenter gave each participant a reminder when there were about 10 trials remaining, which may have encouraged participants to finish more actively, thus causing an increase in accuracy.

In the study by Spataro et al. (2015), the authors studied whether pointing movements facilitated the serial recall of spatial positions. They reported that pointing impaired the recall of the initial and middle positions more than the recall of the final position. Our data revealed similar findings, albeit from a different perspective. Participants were able to identify the difference between study and test arrays significantly better when the item shifted in test arrays was the last item in the study array, as opposed to when any of the first four items shifted. As Spataro et al. suggested, this result may be concerned with the recency effect of visual memory that applies only to last one or two items, which is completely removed by just 3 seconds of mental arithmetic (Phillips & Christie, 1977). This component of visual-spatial memory, compared to the more stable component that can be associated to visual long-term memory, which encompasses the first four items, seemed to be disrupted less by pointing.

This study had two important differences when compared to previous similar studies (e.g., Chum et al., 2007; Dodd & Shumborski, 2009; Rossi-Arnaud et al., 2011; Spataro et al., 2015). First, the experiment involved the spatial memory and movement in three-dimensional space, while previous studies focused on spatial arrays in 2D space. Surprisingly, our data provided no evidence that the pointing movements and the shift of items on the z-axis had a significant effect on spatial memory, as compared to the x- and y-axis conditions. Thus, it seemed that the cognitive effort necessary to memorize locations did not increase when depth information was involved. Or, alternatively, participants may have used only two-dimensional encoding for our test, since the experiment was presented on a 2D computer display.

The second major difference with previous studies is that participants were given 1.5 seconds to memorize a location in our study, while all previous experiments gave only 1 second or shorter. For example, in one study the to-be-remembered item disappeared when it was touched, or until 1 second had elapsed, whichever came first (Chum et al., 2007). Extra time was given mainly to ensure that participants could point to the targets in three-dimensional space without feeling rushed. For a three-dimensional gesture interface, there was an extra cognitive process where participants had to map their hand movement in the real world along with the one in the virtual interface. It is possible that the lower recognition accuracy obtained from the move condition was due to this additional cognitive load, and the difference might diminish by allowing more time for the tasks. However, the 1.5 second duration was selected by a pilot test, and was long enough that almost all participants followed the pace without difficulties. The participants were also told to try following the pace, but simply move on in case they missed a target.

In addition, it is proved that even with half of the arrays to memorize, fewer items to memorize, and enough time to finish all cognitive tasks, participants still displayed consistently better memory for passively viewed objects (Dodd & Shumborski, 2009). Thus, we argue that the difference observed between the move and no-move conditions were mainly an effect of the pointing movement, not the gesture interface. Further studies might examine the role of the interface in greater detail by manipulating pointing movements in real-world 3D space.

At last, we examined the time taken to give a response in both conditions. Participants used significantly more time to give a response in move conditions. This may suggest that the move conditions cost more mental effort / resources in the recall phase of memory.

CHAPTER VIII:
CONCLUSION

Gesturing is a natural modality of human-computer interaction. Some challenges faced by gesture interface developers include the lack of an effective evaluation method, and the general fatigue that users experience in many projects. This work bridges this gap, examining the role effort plays in a gesture interface, and tries to provide a clearer understanding of what consists of an effective gesture interface. It provides evidence for the need to evaluate gesture systems such that effort is taken into account, calling researchers' and practitioner's attention to the evident importance of effort in gesture interfaces.

Effort has been shown, through this work, to be an important factor to user's satisfaction of a gesture interface. For both gesture command tasks, and pointing and selecting tasks that are often found in a menu system, effort showed a great impact on a user's willingness to use the system. Our findings suggest that designers must consider and balance a gesture's physical effort when designing an interface. Effort should also be taken into account when evaluating a gesture interface. Perhaps one reason effort has not received much attention is the absence of a commonly accepted measurement to evaluate it. Novel methods and techniques to measure effort, especially for small actions, are needed to further clear the obstacle for gesture interface research.

Another contribution of this work is that it explored the relationship between effort and reliability, and how these two factors together impacted a gesture's overall user experience. There is no doubt that people prefer reliable systems. The importance of reliability was confirmed in this work, and it was found that there were cases that user satisfaction correlated more strongly with reliability than with effort. Achieving a higher reliability has always been one of the primary goals for gesture system developers, especially when new technologies and algorithms were still emerging. At the same time, the experiments reported here demonstrate that users are willing to sacrifice some reliability in exchange for a less effortful interaction, in case they have to face such an effort-reliability tradeoff and the two factors could not be optimized at the same time. In this study, users avoided extreme conditions. At one extreme lies a condition that is perfectly accurate, but awfully tiring. At the other extreme lies a condition that causes minimal fatigue, but is terribly ambiguous. Neither extreme was preferred by participants. Our data suggest that participants balance effort and reliability, which has been previously overlooked. The study suggests that the balance between effort and reliability should be more carefully reviewed to create a more effective evaluation for gesture interfaces from a user's perspective.

Finally, this work examines the effect of pointing on spatial working memory in a three-dimensional gesture interface. Pointing was found to have a negative effect on memory, even if it could have activated both visual and motor encoding. The results also showed that the introduction of the third dimension did not significantly impact memory. A future study that involves pointing to physical objects in the real world may further verify this finding. At last, the response time used during the recall phase was significantly longer when pointing was manipulated in encoding. This may be because the worse memory in the move condition increased the participant's cognitive effort during the recall phase.

This study suffers from several limitations. One is the lack of effective measurements for physical effort. Objectively measuring the physical effort of a gesture interface is extremely difficult because the movements are small. In addition, individual variation may greatly impact the objective measure of physical effort. Some people are stronger than others. Some move more gently and gracefully than others, for example.

Standard questionnaires that use user-reported measurements are a more practical way to obtain a gesture task's physical effort. A list of standard questionnaires for system usability was reviewed, including the System Usability Scale (SUS), the Post-Study Usability Questionnaire (PSSUQ), the Questionnaire for User Interaction Satisfaction (Q.U.I.S) (Chin, Diehl, & Norman, 1988), and the NASA Task Load Index (TLX). However, none of these standard questionnaires focus on the evaluation of the physical effort of a gesture interface, and do not fit the specific research interest of this study. For example, the NASA TLX includes measurements of a task's mental demand, physical demand, and effort, among other attributes. It measures demands beyond the scope of this study and seemed poorly suited for the subtle physical demands explored in this work. Also, the NASA TLX questionnaire neglects system reliability, a critical factor in this work. As a result, this project utilizes its own non-standard questionnaires that use the participant's perceived effort to measure gesture task effort. The use of a non-standard questionnaire is a limitation since it reduces the comparability of the data. Perhaps in the future, it would be worthwhile to integrate portions of the NASA TLX questionnaire to see if it is useful in measuring the subtle physical demands.

Another limitation this project has is due to the great variability of gesture application and technologies. The results, particularly the balance point we obtained from Experiment 3, can hardly be generalized to different types of applications. For example, users may wish to invest more effort in order to ensure high levels of reliability, such as when an error would cause an important setback or risk. The importance of errors were not specifically controlled in this experiment. Nevertheless, this project provides evidence on how effort and reliability can influence a gesture interface.

For future work, there are several directions to pursue. One direction is to explore how system preferences differ among individuals: male versus female or teens versus elders, for example. Another interesting direction is to explore how perceived effort and preference evolve over time, particularly in the context of training. A third is exploring the relationship between gestures and memory, particularly investigating the relationship between pointing gestures and three-dimensional spatial memory in the real world and how people encode the three-dimensional information presented in a two-dimensional computer display. Finally, gestures may have a diminished effect on spatial working memory as the number of items increases; a study that explicitly manipulates array sizes could reveal more about this interesting effect.

In summary, a primary goal of Human Computer Interaction is to create a usable, user-friendly interaction between users and computers. Designing with users in mind has long been a core concept in this field. This work fills a gap in gesture interface research literature, defining and exploring the tradeoff between the needs of the gesture recognition system and the comfort and difficulty of the gesture for users. The evidence and deeper understanding provided by this work may raise attention for the need to evaluate gesture effort, and ultimately, may lead to the design of better, more usable gesture interfaces.

REFERENCES

Alexander, R. M. (1997). A minimum energy cost hypothesis for human arm trajectories. *Biological Cybernetics*, *76*, 97–105.

Amft, O., Amstutz, R., Smailagic, A., Siewiorek, D., & Tröster, G. (2009). Gesture-controlled user input to complete questionnaires on wrist-worn watches. In *Human-Computer Interaction. Novel Interaction Methods and Techniques* (pp. 131–140). Springer.

Baudel, T., & Beaudouin-Lafon, M. (1993). Charade: remote control of objects using free-hand gestures. *Commun. ACM*, *36*, 28–35.

Bekkering, H., & Neggers, S. F. (2002). Visual search is modulated by action intentions. *Psychological Science, 13*, 370–374.

Bellmore, C., Ptucha, R., & Savakis, A. (2011). Interactive display using depth and RGB sensors for face and gesture control. In *Image Processing Workshop (WNYIPW), 2011 IEEE Western New York* (pp. 1 –4).

Bolt, R. A. (1980). *"Put-that-there": Voice and gesture at the graphics interface* (Vol. 14). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=807503%C3%DC

Cabral, M. C., Morimoto, C. H., & Zuffo, M. K. (2005). On the usability of gesture interfaces in virtual reality environments. In *Proceedings of the 2005 Latin American conference on Human-computer interaction* (pp. 100–108). New York, NY, USA: ACM.

Cassell, J. (1998). A framework for gesture generation and interpretation. *Computer Vision in Human-Machine Interaction*, 191–215.

Cerlinca, T., & Pentiuc, S. (2012). Robust 3D Hand Detection for Gestures Recognition. *Intelligent Distributed Computing V*, 259–264.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 213–218). New York, NY, USA: ACM.

Chum, M., Bekkering, H., Dodd, M. D., & Pratt, J. (2007). Motor and visual codes interact to facilitate visuospatial memory performance. *Psychonomic Bulletin & Review*, *14*, 1189–1193.

Cooper-Martin, E. (1994). Measures of cognitive effort. *Marketing Letters*, *5*, 43–56.

Crowley, J., Berard, F., & Coutaz, J. (1995). Finger tracking as an input device for augmented reality. In *International Workshop on Gesture and Face Recognition, Zurich* (pp. 195–200).

De Ruiter, J. (1998). Gesture and speech production. Retrieved from http://pub.uni-bielefeld.de/publication/2530982

Diamond, W. J. (2001). *Practical Experiment Designs: For Engineers and Scientists*. John Wiley & Sons.

Dodd, M. D., & Shumborski, S. (2009). Examining the influence of action on spatial working memory: The importance of selection. *The Quarterly Journal of Experimental Psychology*, *62*, 1236–1247.

Du, H., & To, T. H. (2011). HAND GESTURE RECOGNITION USING KINECT. Retrieved from http://iss.bu.edu/data/jkonrad/reports/HDTT11-04buece.pdf

Engelbrecht, S. E. (2001). Minimum principles in motor control. *Journal of Mathematical Psychology*, *45*, 497–542.

Fant, L. J. (1972). *Ameslan: An introduction to American Sign Language*. National Association of the Deaf. Retrieved from http://www.getcited.org/pub/101345558

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

Flash, T., & Hogan, N. (1985). The coordination of arm movements: an experimentally confirmed mathematical model. *The Journal of Neuroscience*, *5*, 1688–1703.

Forza Motorsport (series). (2014, March 25). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Forza_Motorsport_(series)&oldid=601132269

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*, 419–429.

Goldin-Meadow, S., & Morford, M. (1990). Gesture in early child language. In *From gesture to language in hearing and deaf children* (pp. 249–262). Springer.

Guna, J., Jakus, G., Pogačnik, M., Tomažič, S., & 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*, 3702–3720.

Hannus, A., Cornelissen, F. W., Lindemann, O., & Bekkering, H. (2005). Selection-for-action in visual search. *Acta Psychologica*, *118*, 171–191.

Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, *396*, 228–228.

Jain, J., Lund, A., & Wixon, D. (2011). The future of natural user interfaces. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (pp. 211–214). New York, NY, USA: ACM.

Jeff Sauro. (2012). *Quantifying the user experience : practical statistics for user research / Jeff Sauro, James R. Lewis.* Waltham, MA: Elsevier/Morgan Kaufmann.

Kelley, J. F. (1984). An Iterative Design Methodology for User-friendly Natural Language Office Information Applications. *ACM Trans. Inf. Syst.*, *2*, 26–41.

Kemps, E. (2001). Complexity effects in visuo-spatial working memory: Implications for the role of long-term memory. *Memory*, *9*, 13–27.

Kinect Sports. (2014, March 15). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Kinect_Sports&oldid=578601127

Koponen, M., Aziz, W., Ramos, L., & Specia, L. (2012). Post-editing time as a measure of cognitive effort. *Proceedings of WPTP*. Retrieved from http://157.56.13.76/AMTA2012Files/html/13/13_paper.pdf

Kouroupetroglou, G., Pino, A., Balmpakakis, A., Golematis, V., Chalasyanis, D., Ioannou, N., & Koutsoumpas, I. (2011). Performance evaluation of gesture-based 2D and 3D pointing tasks. In *9th Int. Gesture Workshop-Gesture in Embodied Communication and Human-Computer Interaction* (pp. 10–15).

Kraemer, H. C., & Thiemann, S. (1987). *How Many Subjects?: Statistical Power Analysis in Research*. SAGE.

Krauss, R. M., & Hadar, U. (1999). The role of speech-related arm/hand gestures in word retrieval. *Gesture, Speech, and Sign*, *93*. Retrieved from http://www.columbia.edu/~rmk7/PDF/K%26H.pdf

Leap Motion. (n.d.). Retrieved April 16, 2015, from https://www.leapmotion.com/product

Lee, H., Jeong, H., Lee, J., Yeom, K.-W., Shin, H.-J., & Park, J.-H. (2008). Select-and-point: a novel interface for multi-device connection and control based on simple hand gestures. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems* (pp. 3357–3362). New York, NY, USA: ACM.

Lewis, J. R. (1991). Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. *ACM SIGCHI Bulletin*, *23*, 78–81.

Li, Y. (2012). Hand gesture recognition using Kinect. In *Software Engineering and Service Science (ICSESS), 2012 IEEE 3rd International Conference on* (pp. 196–199).

Logie, R. H. (1995). Visuo-spatial working memory. Retrieved from http://doi.apa.org/psycinfo/1995-97088-000

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press. Retrieved from http://books.google.com/books?hl=en&lr=&id=3ZZAfNumLvwC&oi=fnd&pg=PA6&dq=Hand+and+mind:+What+gestures+reveal+about+thought.&ots=oHe9SCsIbx&sig=iMfb1UGID1XsRz_LYLOO2eI2zwk

McNeill, David. (2008). *Gesture and thought*. University of Chicago Press. Retrieved from http://books.google.com/books?hl=en&lr=&id=N0SmyU4TKRwC&oi=fnd&pg=PR7&dq=gesture+and+thought&ots=9a63x5behx&sig=36j1OdJtd_rUSnaF2zkZg6rfgYM

Mistry, P., & Maes, P. (2009). SixthSense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches* (pp. 11:1–11:1). New York, NY, USA: ACM.

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

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology*, 411–424.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics*, *46*, 135–147.

Nguyen, D. D., Pham, T. C., & Jeon, J. W. (2009). Fingertip detection with morphology and geometric calculation. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on* (pp. 1460–1465). IEEE.

Nickel, K., & Stiefelhagen, R. (2007). Visual recognition of pointing gestures for human–robot interaction. *Image and Vision Computing*, *25*, 1875–1884.

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

Phillips, W. A., & Christie, D. F. M. (1977). Components of visual memory. *Quarterly Journal of Experimental Psychology*, *29*, 117–133.

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

Quek, F. K. H. (1996). Unencumbered gestural interaction. *IEEE MultiMedia*, *3*, 36 –47.

Quinn, J. G. (1994). Towardsa clarification of spatial processing. *The Quarterly Journal of Experimental Psychology Section A*, *47*, 465–480.

Rekimoto, J. (2001). Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Wearable Computers, 2001. Proceedings. Fifth International Symposium on* (pp. 21–27).

Ren, Z., Meng, J., Yuan, J., & Zhang, Z. (2011). Robust hand gesture recognition with kinect sensor. In *Proceedings of the 19th ACM international conference on Multimedia* (pp. 759–760). New York, NY, USA: ACM.

Riener, A., Rossbory, M., & Ferscha, A. (2011). Natural DVI based on intuitive hand gestures. In *Workshop UX in Cars, Interact* (p. 5).

Rimé, B. (1982). The elimination of visible behaviour from social interactions: Effects on verbal, nonverbal and interpersonal variables. *European Journal of Social Psychology*, *12*, 113–129.

Rogers, W. T. (1978). The Contribution of Kinesic Illustrators Toward the Comprehension of Verbal Behavior Within Utterances. *Human Communication Research*, *5*, 54–62.

Ronchetti, M., & Avancini, M. (n.d.). Using Kinect to emulate an Interactive Whiteboard. Retrieved from http://latemar.science.unitn.it/segue_userFiles/LITSA/Avancini_Mattia_138793_Thesis.pdf

Rossi-Arnaud, C., Spataro, P., & Longobardi, E. (2011). Effects of pointing on the recall of simultaneous and sequential visuospatial arrays: a role for retrieval strategies? *Psychological Research*, *76*, 699–712.

Spataro, P., Marques, V. R. S., Longobardi, E., & Rossi-Arnaud, C. (2015). Does pointing facilitate the recall of serial positions in visuospatial working memory? *Cognitive Processing*, *16*, 377–381.

Starner, T., & Pentland, A. (1995). Real-time american sign language recognition from video using hidden markov models. In *Computer Vision, 1995. Proceedings., International Symposium on* (pp. 265–270).

Stern, H. I., Wachs, J. P., & Edan, Y. (2006). Human Factors for Design of Hand Gesture Human - Machine Interaction. In *IEEE International Conference on Systems, Man and Cybernetics, 2006. SMC '06* (Vol. 5, pp. 4052 –4056).

Suarez, J., & Murphy, R. R. (2012). Hand gesture recognition with depth images: A review. In *2012 IEEE RO-MAN* (pp. 411 –417).

Trigueiros, P., Ribeiro, A. F., & Lopes, G. (2012). Vision-based hand segmentation techniques for human-robot interaction for real-time applications. Retrieved from http://repositorium.sdum.uminho.pt/handle/1822/15263

Uno, Y., Kawato, M., & Suzuki, R. (1989). Formation and control of optimal trajectory in human multijoint arm movement. *Biological Cybernetics*, *61*, 89–101.

Van den Bergh, M., & Van Gool, L. (2011). Combining RGB and ToF cameras for real-time 3D hand gesture interaction. In *Applications of Computer Vision (WACV), 2011 IEEE Workshop on* (pp. 66–72).

Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology*, *95*, 57–79.

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

Wah Ng, C., & Ranganath, S. (2002). Real-time gesture recognition system and application. *Image and Vision Computing*, *20*, 993–1007.

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology*, *114*, 591.

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin, 84*, 963.

Wilson, A. D. (2010). Using a depth camera as a touch sensor. In *ACM International Conference on Interactive Tabletops and Surfaces* (pp. 69–72). New York, NY, USA: ACM.

Wright, M., Lin, C.-J., O'Neill, E., Cosker, D., & Johnson, P. (2011). 3D gesture recognition: an evaluation of user and system performance. In *Pervasive Computing* (pp. 294–313). Springer.

Yang, C., Jang, Y., Beh, J., Han, D., & Ko, H. (2012). Gesture recognition using depth-based hand tracking for contactless controller application. In *2012 IEEE International Conference on Consumer Electronics (ICCE)* (pp. 297 –298).

Yang, M.-H., & Ahuja, N. (2001). Recognizing hand gestures using motion trajectories. In *Face Detection and Gesture Recognition for Human-Computer Interaction* (pp. 53–81). Springer.

Zimmerman, T. G., Lanier, J., Blanchard, C., Bryson, S., & Harvill, Y. (1987). A hand gesture interface device. In *ACM SIGCHI Bulletin* (Vol. 18, pp. 189–192).