# AN ERGONOMICS INVESTIGATION OF THE APPLICATION OF VIRTUAL REALITY ON TRAINING FOR A PRECISION TASK

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Industrial Engineering

by

Delaney M. Bales

June 2017

# COMMITTEE MEMBERSHIP

| | |
|---|---|
| TITLE: | An Ergonomics Investigation of the Application of Virtual Reality on Training for a Precision Task |
| AUTHOR: | Delaney M. Bales |
| DATE SUBMITTED: | June 2017 |
| COMMITTEE CHAIR: | Dr. Reza Pouraghabagher, Ph.D.<br>Professor of Industrial and Manufacturing Engineering |
| COMMITTEE MEMBER: | Dr. Lizabeth Schlemer, Ph.D.<br>Professor of Industrial and Manufacturing Engineering |
| COMMITTEE MEMBER: | Dr. Robert Crockett, Ph.D.<br>Professor and Department Chair of Biomedical Engineering |

# ABSTRACT

## An Ergonomics Investigation of the Application of Virtual Reality on Training for a Precision Task

Delaney M. Bales

Virtual reality is rapidly expanding its capabilities and accessibility to consumers. The application of virtual reality in training for precision tasks has been limited to specialized equipment such as a haptic glove or a haptic stylus, but not studied for handheld controllers in consumer-grade systems such as the HTC Vive. A straight-line precision steadiness task was adopted in virtual reality to emulate basic linear movements in industrial operations and disability rehabilitation. This study collected the total time and the error time for the straight-line task in both virtual reality and a physical control experiment for 48 participants. The task was performed at four different gap widths, 4mm, 5mm, 6mm, and 7mm, to see the effects of virtual reality at different levels of precision. Average error ratios were then calculated and analyzed for strong associations to various factors. The results indicated that a combination of Environment x Gap Width factors significantly affected average error ratios, with a p-value of 0.000.

This human factors study also collected participants' ratings of user experience dimensions, such as difficulty, comfort, strain, reliability, and effectiveness, for both physical and virtual environments in a questionnaire. The results indicate that the ratings for difficulty, reliability, and effectiveness were significantly different, with virtual reality rating consistently rating worse than the physical environment. An analysis of questionnaire responses indicates a significant association of overall environment preference (physical or virtual) with performance data, with a p-value of 0.027.

In general, virtual reality yielded higher error among participants. As the difficulty of the task increased, the performance in virtual reality degraded significantly. Virtual reality has great potential for a variety of precision applications, but the technology in consumer-grade hardware must improve significantly to enable these applications. Virtual reality is difficult to implement without previous experience or specialized knowledge in programming, which makes the technology currently inaccessible for many people. Future work is needed to investigate a larger variety of precision tasks and movements to expand the body of knowledge of virtual reality applications for training purposes.

Keywords: virtual reality, industrial training, disability rehabilitation, haptic feedback, precision task, hand steadiness

# ACKNOWLEDGMENTS

I would like to express special thanks to my advisor, Dr. Reza Pouraghabagher, for mentoring and challenging me through this process. Together, we learned a lot about virtual reality, its capabilities, and its limitations. His contributions and guidance were significant, with a p-value of 0.000. Thank you to Dr. Robert Crockett for introducing me to virtual reality, which became an experience that inspired this research. I would also like to thank Dr. Lizabeth Schlemer for her thoughtful feedback and continued support throughout this process. I appreciate the financial support provided by the Industrial and Industrial and Manufacturing Engineering Student Fee Committee and the Melfred Borzall Project Funds to obtain the equipment needed for this research. Thank you to my friends and family who have supported me through this entire process. Finally, I would like to express special thanks to Johnathan Baird (B.S. Computer Science) for developing the virtual reality program and for his consultation and trusted advice throughout this process.

# TABLE OF CONTENTS

Page

APPENDICES

# LIST OF TABLES

Table Page

# LIST OF FIGURES

# 1. Introduction

Virtual reality (VR) is a rapidly expanding technology that can be used to simulate almost any environment. Current applications include gaming, entertainment, professional training, engineering development, and physical therapy [2]. The believability of a virtual environment is called *immersion*. To create a sense of immersion in VR, multisensory stimuli are needed.

VR has long relied on the senses of sight, hearing, and touch to create convincing experiences. Head-mounted displays (HMDs) provide the visual and sometimes the audial information needed to create a simulated environment. The sense of touch, called *haptics*, is a critical factor of how people physically perceive and feel the world around them [22]. Haptics includes both tactile touch, stimulated through the skin, and proprioception, the sense of one's own bodily positions and motions. Haptic feedback can be difficult to include in a virtual experience and is currently in the early development stage.

The current haptic solution in most consumer-grade VR systems is a hand-held controller. These controllers are usually shaped like a remote control and have ergonomically designed buttons. Active haptic feedback is usually in the form of vibrations from a motor inside the controller. Most games and experiences designed for VR utilize full arm motions while holding the controller and allow for high error tolerances. The haptic feedback component of VR is low resolution, simplified, and concentrated at the hands [9].

Human perception of proprioception in the arms during VR experiences is still a young researched field. As a result, the virtual reality experience may lose its sense of realism or create confusion for its users. VR can be used to train workers, such as machine operators in industrial tasks, while reducing accident risks and eliminating material waste. For some VR simulations, such as surgical procedures, an accurate and realistic simulation is crucial to producing practical results. So far, VR is mainly equipped for arm movement that requires low tolerance [9]. Research needs to be conducted to study detailed tolerances with proprioceptive tasks in virtual reality.

*Cognitive ergonomics* is a field of human factors research that includes the study of human perception of and interaction with their environment [13]. In the current state of VR, people are expected to recognize and respond to the foreign environment. As VR becomes more immersive, people's perceptions and interactions in a simulated world may more closely resemble their real-life responses. The development of VR technology relies on human factors engineering and ergonomics research to achieve a fully immersive experience.

Right now, one of the big deficiencies in VR is the lack of information on performing precision tasks with hand-held controllers [12]. In terms of precision tolerance, a user's expected performance in virtual reality is unknown compared to their performance in physical reality. In addition, there is little empirical information about users' perceptions of a precision tasks in the current VR capability setting. A study of these issues could provide valuable information about needed improvements in VR technology.

This study investigates the effects of virtual reality on training accuracy and human perception for a precision task. A task requiring participants to trace a straight line within a specified gap tolerance was developed in virtual reality. The experiment was conducted to measure the performance of hand steadiness while holding a VR controller. In addition, user perceptions of the two environments were also documented and analyzed to see if VR is a suitable substitute environment for training, and, if so, what factors contributed to this perception.

The purpose of this study is to investigate whether virtual reality can be used to replace physical tasks and emulate training for industrial operations or disability rehabilitation, such as tracing a straight line. The results of this research can assist future VR developers in improving virtual reality programs for factors related to precision, training, and perception. The results also lend itself to further work in the area of precision hand movement tasks in virtual reality.

# 2. Literature Review

The purpose of this literature review is to search what information and research has already been conducted on haptic systems and ergonomics in virtual reality. The term *visuo-haptic* refers to systems that combine visual and touch sensory engagement. It is evident that there is still a lot to be learned in the field of visuo-haptic immersion with virtual reality. Research in the field of haptics and ergonomics begins with background information on human body interactions with haptic feedback and how tactile and proprioceptive senses have been tested in previous studies. This review discusses what is already known in the field of visuo-haptic mixed realities and what future research could benefit the scientific community.

## 2.1 Virtual Reality

The first virtual reality head-mounted display (HMD) was created in 1960 [8]. However, the development of VR was slow until the late 1980’s and 1990’s. For the last 20 years, virtual reality has progressed tremendously. Lots of research has been conducted on virtual reality technology on various topics such as visual quality, interactive controllers, and psychological effects. One topic that has been researched heavily within human-computer interaction is haptics. However, haptics in conjunction with virtual reality specifically is still a new concept and field of study.

## 2.2 Human Physiology for Haptics

Human touch is categorized two different ways: tactile and proprioceptive. Tactile touch is stimulated through the skin and characterizes sensations such as heat, pressure, vibration, slipping, and pain [14]. Proprioception is the awareness of one’s own body placement or limb position. An example of proprioception is when someone is aware of where their arms are reaching in the dark or while their eyes are closed. Proprioceptive touch also provides information on body forces and motion through muscles, tendons, and joints [14]. Haptic feedback affects both tactile and proprioceptive touch.

The human body has two types of skin: glabrous (non-hairy) skin found on the palm of the hands and hairy skin found on the rest of the body. Certain parts of the body, such as the palm of the hands, are more sensitive and perceptive than other parts of the body, such as the outer thigh, due to variances in thickness, vascularity (blood vessels), density, electrical conductivity, and mechanical properties of the skin [16]. A 1968 study shows that women are more sensitive to pressure than men and both the left and right sides of the body are equally sensitive to pressure [24].

Testing tactile sensitivity can be difficult due to the various factors regarding physical limitations and differences, and tactile location, spacing, and strength. Since some parts of the body are more sensitive than others, and due to variability from person to person, the required or even recommended amount of spatial separation between two simultaneous points of contact can vary quite a lot to obtain an accurate perception of reality [16]. However, it has been found that spatial separation is more distinctly perceived “if they are oriented along the transverse rather than the longitudinal axis of the body,” which is visualized below in Figure 1 [16].

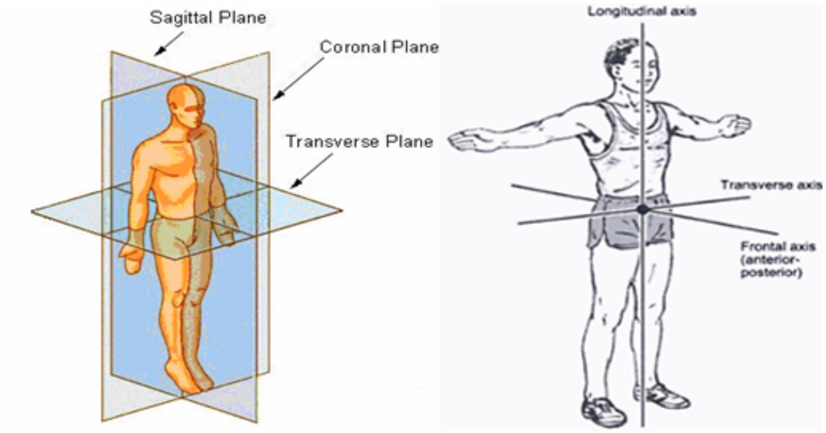

Figure 1. Anatomical human planes of motion (left), and axes of rotation (right) [18].

Human skin has physical limitations associated with tactile perception. According to Myles and Binseel, adaptation and masking are two of these physical limitations. Adaptation is when “a stimulus is presented for a lengthy amount of time” and results in reduced perception [16]. The habituation effect, as MacLean calls it, could heighten a person’s threshold for subsequent trials of tactile stimulation [14]. Masking is when “the perception of a target stimulus is changed by a non-target stimulus that overlaps in time and/or space, and masking can interfere with one’s ability to localize the target stimulus” [15]. Essentially, masking is when one tactile stimulus overrides another tactile stimulus and reduces the perception. MacLean also mentions this masking effect when discussing the spatial resolution of tactile receptors for accurate human perception: “if there is a high density of receptors, the resulting overlap and ‘crosstalk’ reduce effective resolution” [14]. A way to reduce the masking effect is to allow time and space between stimuli, such as reducing the spatial resolution or lowering the frequency of tactile signals.

A commonly used method of tactile feedback is vibration. Vibration is better detected on hairy, bony skin than on soft, fleshy areas [15]. Three separate research studies from 1937, 1996, and 2003 all found that elderly people have a higher vibration threshold in certain areas of the body [6, 11, 23]. The 1937 study by Laidlaw and Hamilton also found that overweight people generally had a higher vibration threshold.

## 2.3 Haptic Hardware

To target the two types of haptic feedback, tactile and proprioceptive, there are two types of haptic devices. Force feedback devices target proprioception by “providing forces that react to our movements in space” [14]. These types of devices are useful for simulating a physical force response virtually, such as dialing a knob that gets gradually more difficult to turn. Tactile displays are designed to provide local feedback to the skin, such as vibration, heat, or pressure. MacLean lists “piezoelectric, voice coil, solenoid, or eccentric motor actuation” as the most popular devices for tactile displays due to their “efficiency, small size, and low power requirements.” Hatzfeld and Kern describe a variety of haptic interaction system types, such as assistive systems, haptic interfaces, manipulators, teleoperators, and comanipulators [7].

There are several examples of existing haptic systems. The ALEx arm exoskeleton is “an upper limb mechanically compliant exoskeleton with low encumbrances, and low friction of the actuation system” which has six degrees of freedom [20]. The ALEx exoskeleton system, which is an arm and shoulder haptic feedback exoskeleton that is paired with a head-mounted virtual reality (VR) display, is shown below in Figure 2.

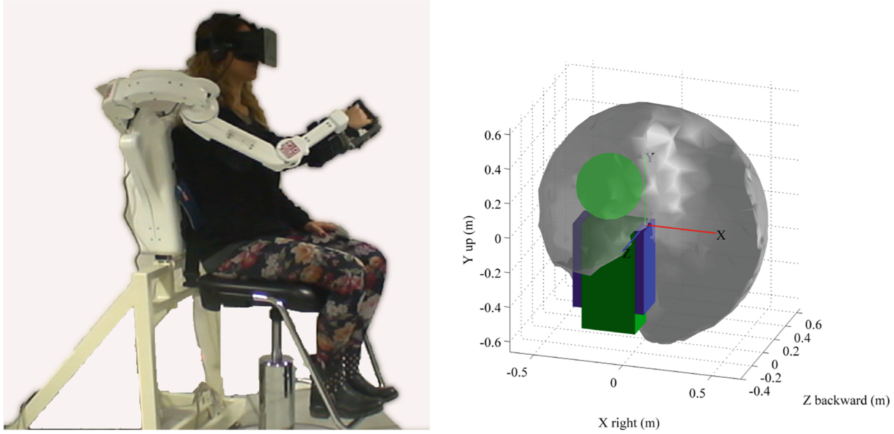

Figure 2. ALEx exoskeleton setup (left), and the virtual workspace of the system in grey with the coordinate system located on the subject's shoulder (right) [20].

In 1989, a haptic display system named SPIDAR ("SPace Interface Device for Artificial Reality") was developed [21]. This string-based design is "composed of a cubic frame that encloses a working space," in which "the main components of the system are four sets of a DC motor, a pulley, an encoder, and a string" [21]. Encoders are used to track position or rotation. The haptic display couples with a virtual environment to provide force feedback when colliding with an object in the virtual world. The SPIDAR system is illustrated in Figure 3 below.

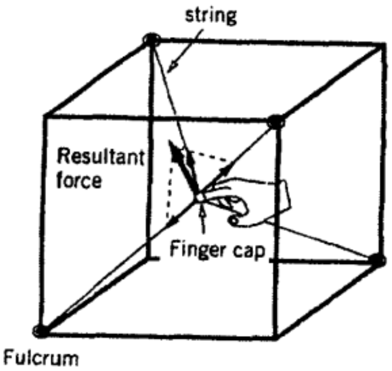

Figure 3. SPIDAR system [21].

Researchers developed the HIRO III haptic simulator, a five-fingered haptic interface robot that connects to a user's five fingertips [5]. This system was developed to avoid the usage of large and clunky exoskeletons and to allow more degrees of freedom than an exoskeleton. Akbar and Jeon call this system “a robotic arm with five fingers [that] provides independent forces to each finger of the user” [1]. The HIRO III design, shown in Figure 4, was tested to show “high-precision force presentation.”

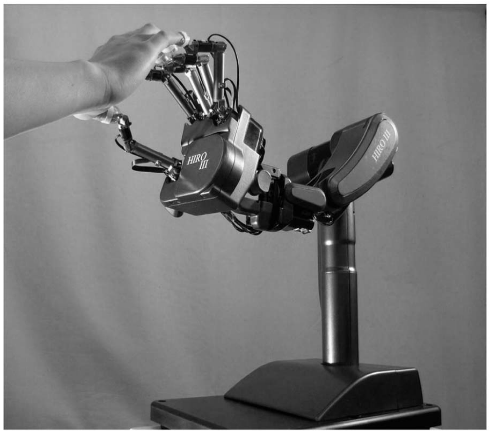

Figure 4. HIRO III haptic simulator [5].

The PHANToM haptic device (currently known as the Geomagic Touch) is a stylus-based design [4]. According to Eck, “the haptic stylus is often augmented with some context-dependent tool like a drill for dental surgery training, a brush for virtual painting, or tools for rapid prototyping” [4]. This device consists of two gimbal joints that allow the stylus to be moved around like a pen or brush. The PHANToM stylus is useful for many applications; however, one downfall is that it is difficult to calibrate.

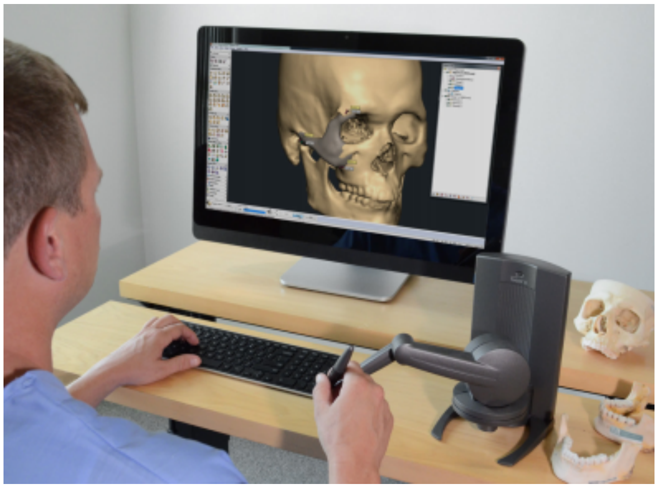

Figure 5. Geomagic Touch (PHANToM) haptic stylus device being used in combination with visual simulation [18].

While there are many options for haptic devices, there are several common limitations. Force feedback devices can suffer from “subtle instability” such as jittering or excess signal [14]. Simulated forces can feel weak or compromised, or respond too slowly, which takes away from the realistic simulation. Physical interactions may be oversimplified or inappropriate for the virtual application, such as having incorrect geometry, size, or degrees of freedom [14]. Differences between physical and virtual haptics can detract from the user experience and authenticity of sensitivity and perception. For owners of haptic devices, equipment can take up a lot of power disproportional to the intended environment and usage, which can be cumbersome and expensive to maintain.

## 2.4 How to Test Haptics: Previous Procedures

A 1950 paper titled *Computing Machinery and Intelligence*, by Alan Turing, questions the ability of machines to exhibit artificial intelligence that is indistinguishable from that of complex human thought. The “haptic Turing test” is one in which the virtually rendered object feels the same as the real object [14]. Researchers have been trying to develop and improve haptic tests and interfaces for more than 80 years. However, only in the last couple decades has the marriage of visual simulation and haptic feedback been studied. Due to the subjective and abstract nature of touch perception, measurements of haptic feedback have been difficult to quantify and collect. A variety of qualitative and quantitative methods have been utilized in the past to measure haptic feedback.

MacLean mentions the “just noticeable difference (JND)” percentage, which is “a common measure of both tactile and proprioceptive sensory resolution” [14]. The exponential resolution curve models the differences in human perception at low versus high torque levels, which makes distinguishing between force levels difficult. The JND measurement attempts to capture small changes in perception but does not answer the question of “whether we discriminate torque or force directly or if we instead sense compliance or something else” [14]. Nonetheless, the JND measurement helps characterize discrimination between torque and displacement.

While testing the ALEx exoskeleton system, researchers assessed the physiological effects both quantitatively and qualitatively. The ALEx system was developed to simulate human interaction with an aggressive virtual avatar in a dark alleyway setting [20]. The researchers monitored the subjects’ heart rate, electrodermal activity, respiration rate, and oxygen saturation for a correlation with heightened sense of danger. Qualitative measurements were taken through a 17-question questionnaire that targeted control, presence, and embodiment [20]. These questions asked about the user’s perception of congruence and persuasiveness—the believability—of the virtual environment compared to a real one. The questionnaire also asked the user if they felt a sensation of danger, oppression, or a need to move their arm during the simulation. The study had two randomized test groups, one with haptics and one without. The questionnaire responses showed several significant differences between the haptic and non-haptic tests, however most responses were similar and showed no significant difference.

Another group of researchers compared the effectiveness of a lightweight wearable exoskeleton with head-mounted display (HMD) to a table-top haptic device with computer monitor display for virtual immersion and sense of presence [10]. In this study, eighteen voluntary test subjects were randomly asked to try one of the two systems and provide their subjective feedback regarding control factors, sensory factors, distraction factors, and realism factors [10]. Most of the data analysis was based on subjective qualitative feedback, and only the time it took to perform the task was recorded quantitatively. The two systems are shown below in Figure 6.

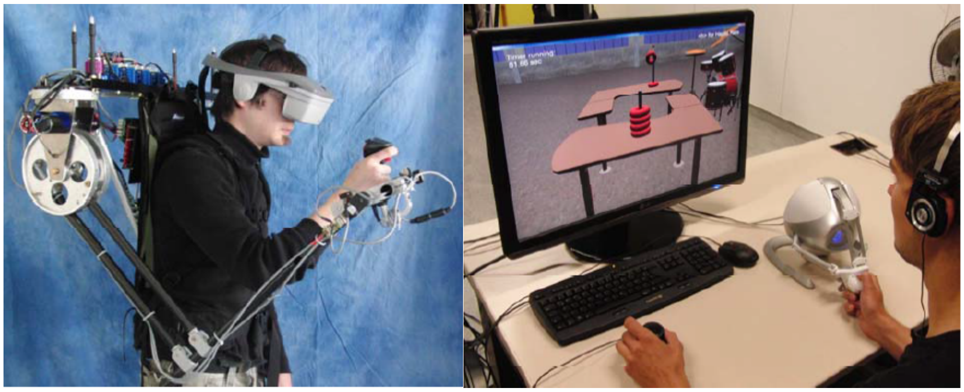

Figure 6. Lightweight wearable haptic exoskeleton paired with HMD (left), and table-top haptic interface paired with computer monitor virtual environment (right) [10].

## 2.5 Visuo-Haptic Research Studies

Several studies in recent decades have conducted different types of haptic feedback tests in conjunction with visual simulation or virtual reality. The previous tests collect and analyze subjective qualitative data (collected through questionnaires) and objective quantitative data (such as physiological measurements) to characterize haptic perception. In these examples, statistical analysis was performed on all test results to attempt characterization and modeling of haptic feedback effectiveness when combined with visual simulation. There are various haptic hardware systems that have been developed for research and consumer use, but the existing options are either underdeveloped or inaccessible, and the variety of systems types is still very broad. Any sign of a preferable or superior haptic test system is still to be determined.

The results of several visuo-haptic research studies point to the conclusion that the quality of visuals in a virtual environment has a greater influence on human perception than the quality of haptic feedback [12, 17]. The first study involves stroke patients performing a task in both physical and virtual environments for physical therapy [12]. The second study studied how people's perception of softness changed just by projecting surface deformations [17]. Both studies found that the visual component of the test was more effective at tricking the brain's perception than the haptics. Another study concluded that a head-mounted display with a wearable exoskeleton was more effective than a table-top haptic device with a computer monitor for virtual immersion [10]. This validity of this study is questionable due to its lack of statistical analysis, omission of its data questionnaire, and poor comparison of "apples to oranges" systems that are too different to be compared.

A 1997 study on "Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction" discusses the issues with interactivity in virtual worlds. The authors attribute the main issues with manipulation and handling in VEs to a lack of haptic feedback limited input information, and limited precision. They also attribute precise manipulation difficulties to the absence of a physical work surface for the person to steady their hand against. Haptic feedback is essential to helping users improve their interactivity with virtual environments [15].

Another study compares the effectiveness of a visuo-haptic virtual environment to a physical environment on the quality of grasping in stroke patients [12]. The twelve subjects had to grasp three differently shaped objects in both physical and virtual environments. The three test methods for grasping the objects was physical environment without a haptic glove, physical environment with a haptic glove, and virtual environment with a haptic glove and HMD. The virtual environment attempted to replicate the physical environment as closely as possible in position, shape, and size. The study found that the subjects responded better to the physical environment. However, in the virtual environment, it was discovered that the quality of the visual graphics was more relevant to the success of the individuals grasping the object than the haptic feedback from the glove [12]. This research suggests that the quality and accuracy of the visual component of a virtual environment should be well-developed, whereas the haptic feedback component can be less developed and still effective.

In the ALEx exoskeleton study in which a person interacts with an aggressive virtual avatar, the results of a haptic group were compared to that of a non-haptic group for each of the 17 questionnaire questions. The results of the two groups were similar for most questions regarding presence and embodiment, except for several significant differences. One significant difference between the two groups was that the haptic group was more likely to feel the sensation that the avatar opponent could have grabbed their arm [20]. The haptic group was more likely to passively move their arm at the beginning and end of the experiment. Another result is that the haptic group felt more congruence between their different sense perceptions, which suggests that mixed visuo-haptic systems are more immersive than strictly visual environments [20].

The 2011 study that was conducted on the usability of a wearable versus a table-top visuo-haptic device claims that the wearable device was preferable [10]. Data was collected qualitatively via a subjective questionnaire and quantitatively via measurement of the amount of time it took each subject to complete a task. The questionnaire results and time data showed that the wearable exoskeleton with HMD was preferable to the table-top haptic simulator and computer screen. However, little to no statistical analysis for significance was performed and the questionnaire questions were not disclosed, so the results and conclusions bring up several questions. To verify and validate the results, statistical analysis and evaluation of the unbiased nature of the questionnaire should be performed.

A system exists that uses visual projections to manipulate a person's perception of softness without haptic feedback. This system, called SoftAR, simulates different objects by projecting surface deformations onto a blank soft surface [17]. A user can perceive a different material softness from the actual softness of the interface through the surface deformation effect (SDE), which is a form of augmented reality where spatial projections are applied to a soft surface. This research indicates that while the physical surface has not changed, a person's perception of softness can be manipulated visually through surface deformation projections.

Only a few research studies paired visual and haptic feedback and analyzed the effectiveness of their combination on user immersion and perception. Of the studies that did not, most of them suggest pursuing further research that combines visual and haptic interfaces [1, 4, 10].

## 2.6 Background Summary

Research on the combined effect of visual simulation and haptic feedback, dubbed visuo-haptics mixed realities, has been grazed by several studies over recent decades. A review of the literature reveals several previously used methods for testing subjective haptic feedback as well as existing results and conclusions related to visuo-haptic interfaces. Some studies conclude that the visual component of an immersive virtual reality environment is more effective at establishing authenticity than haptic feedback. Other studies attempt to quantify and qualify haptic feedback perception. The recurring suggestion for future work in the field of visuo-haptics is a strong sign that research in this area is needed and valuable to the scientific community.

Precision work in virtual reality could be applied to physical rehabilitation for disabled persons, such as stroke patients. Current virtual reality studies for precise haptic applications involve the use of a haptic glove or specialty equipment, rather than a handheld controller [12]. Handheld VR controllers could be used to emulate experiences such as training for industrial tasks. The lack of research on precision tasks in virtual reality applications indicates a void in the body of knowledge. Therefore, the objective of this human factors research is to investigate the application of virtual reality on training for a precision task using handheld controllers.

# 3. Design

## 3.1 Overview

Current virtual reality systems are not designed for detailed, fine-tuned tasks. Current VR applications are usually games that involve full body movement or full arm interactions, such as throwing a stick, shooting zombies, or picking up objects on a desk. Rarely do these tasks require extreme precision. To investigate this void, a straight-line accuracy test was adapted for VR, using the HTC Vive handheld controllers. The straight-line test was created in both a physical (real-life) environment and a virtual environment to compare participants' results between the two.

The physical environment was the "reality" experiment and the control consisting of a real room with physical equipment. The participants performed an ergonomics test using equipment from a university ergonomics laboratory. The virtual environment was a replica of the room and equipment in an interactive virtual reality setup. In the virtual environment, the participants wore a VR headset and performed the same task using handheld controllers.

Participants were screened for those who have little to no previous experience with virtual reality and are right-hand dominant. A total of 24 males and 24 females were selected to participate in the experiment. During the experiment, the participant sat in a chair at a desk. The experimenter sat on the other side of the desk, with a computer, and gives instructions. Each participant completed the precision task in both environments in a randomly assigned order and then completed a questionnaire.

## 3.2 Physical Environment

The purpose of the physical environment was to test how participants performed in a real-life setting. This was the control experiment.

The straight-line test in the physical environment utilized the Lafayette Instruments Groove Type Steadiness Tester. This straight-line steadiness tester has two parallel, adjustable, and 25cm long metal plates that are mounted on top of a mirrored glass surface, which is secured to the board. The gap width between the two plates was set to either 4mm, 5mm, 6mm, or 7mm. The participant used an electronic test lead to draw a line on the glass from left to right inside the gap between the two plates, without touching the metal. If the test lead contacted the metal, an electric circuit was completed and a clock counter kept track of how long the participant was in contact with the metal. This was recorded as the "error time," in milliseconds. The total time it took a participant to start and stop drawing the line was also recorded. An error ratio, equal to the error time over the total time, was calculated and averaged across three samples for each participant. The error ratio was a value between 0.0 and 0.1.

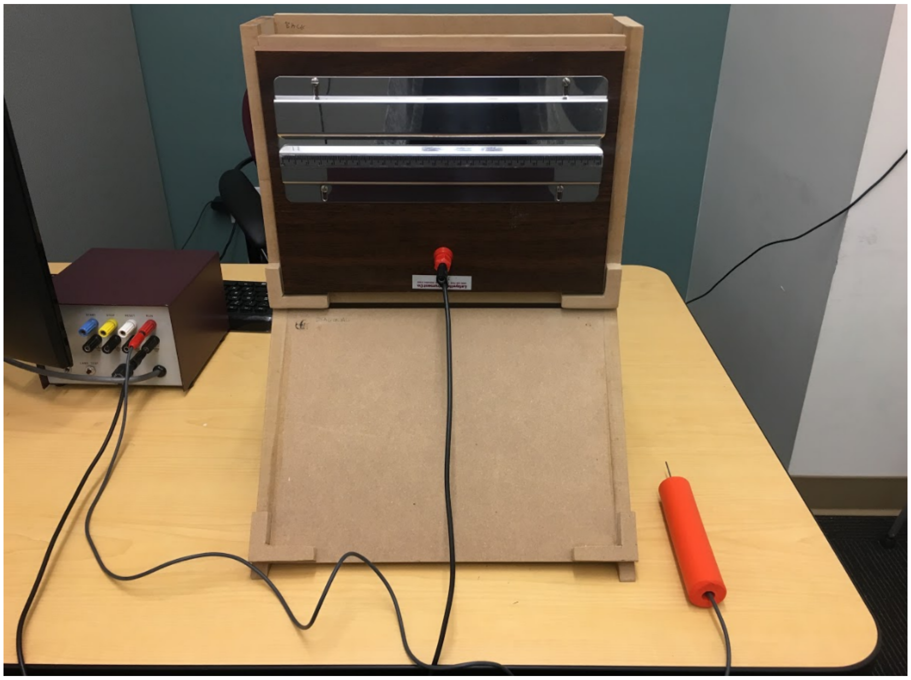

Figure 7. Physical test setup.

The straight-line tester was mounted on a vertical board in front of the participant. It was positioned in the same place for every participant. The stylus was an electronic test lead that was embedded inside a foam tube of 1.125in diameter. The participant was instructed to hold the stylus like they would a remote control, with their fingers curled around the bottom of the foam grip and their thumb on top. They were instructed to raise their arm comfortably to touch the board with the tip of their stylus.

In the physical environment, the height, length, and depth of the target area was controlled by the equipment. Depth was controlled by the back of the straight-line test board. Length was controlled to 25cm. Participants concentrated on staying within the height of the gap width and not touching the metal guides. They were instructed to keep a pace of three seconds to trace the full 25cm line. The experimenter recorded the actual times and had participants redo the task if their pace was too fast or too slow.

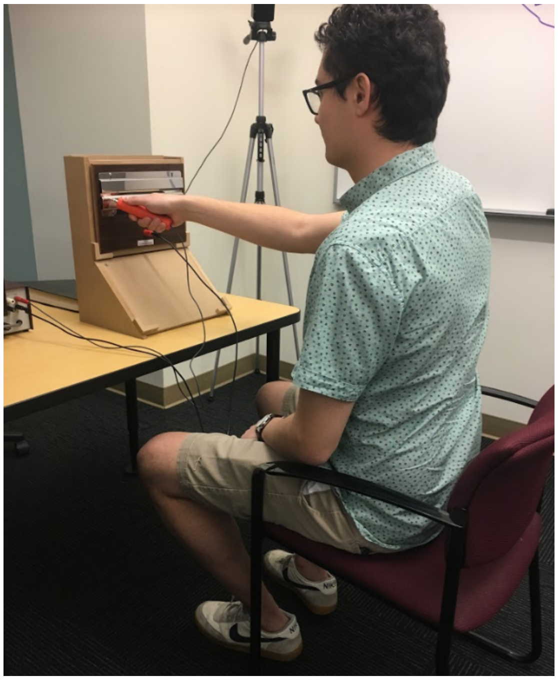

Figure 8. Participant performing test in the physical environment.

Figure 9. Close-up of participant tracing a line in the physical environment.

### 3.3 Virtual Environment

The purpose of the virtual environment was to test how participants performed the task in virtual reality. The participant wore the head-mounted display and picked up the HTC Vive controller with their right hand (Figure 10). The VR program recorded the same information (error time and total time) as the physical equipment. In adapting to virtual reality, there were several differences between the virtual and physical environments. The changes were meant to reduce the cognitive load of the virtual environment so that participants were not distracted while performing the task.

Figure 10. Participant performing test in the virtual environment.

The room, table, chairs, and straight-line test equipment were replicated in VR. The experimenter, the computer, and the VR motion tracking base stations were removed from the virtual environment to de-clutter the visual space.

The position and orientation of the virtual controller matched wherever the controller was in real time. The controller was modelled in VR to look the same as the physical stylus. The participant was instructed to hold the stylus like they would a remote control, with their fingers curled around the bottom of the foam grip and their thumb on top. They were instructed to raise their arm comfortably to touch the board with the tip of their stylus. The participant could not see their hand or body in VR.

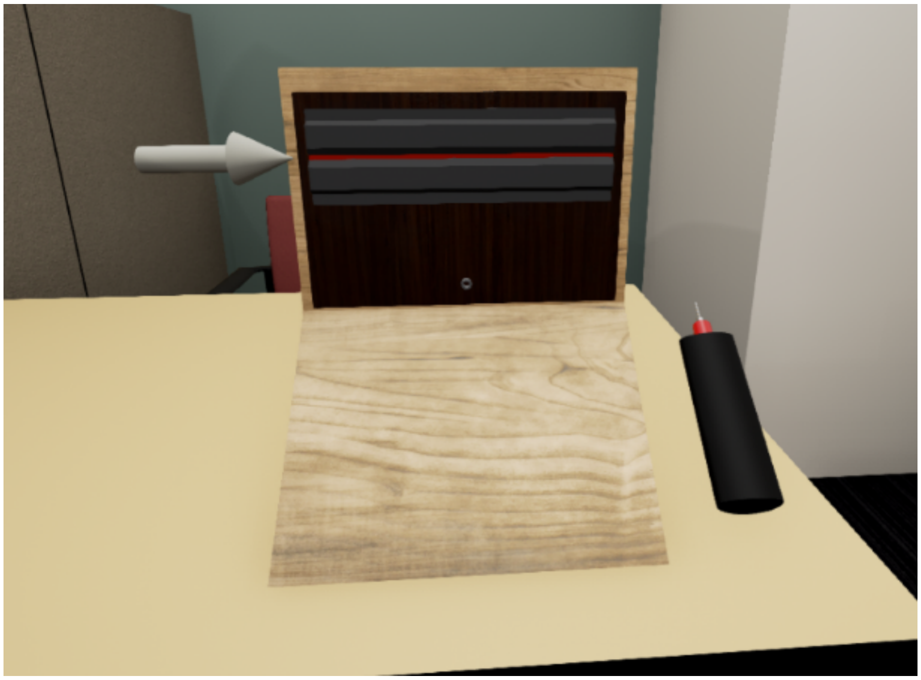

Figure 11. Test setup in virtual reality.

No vibrational feedback was implemented in the hand controller. Vibrations would shake the participant’s hand and risk increasing their error time. Error time from systematic vibrational feedback would be impossible to discern from the participant’s natural unsteadiness, resulting in an artificially inflated error time and poor results.

The same rationale applies to the elimination of pressing buttons on the controller. Even though the VR controller had buttons, the participant did not need to press any buttons during the experiment. The participant simply moved their controller to the target area and the experimenter controlled when the program started and stopped recording.

Although it was considered, a physical board was not installed for the participant to hit with their controller when they touched the virtual board, as a form of haptic feedback. The calibration procedure was not precise enough to ensure perfect alignment of the physical board with the virtual board, which could result in failure to complete the task or skewed results.

Finally, to replace the absence of any haptic feedback, visual feedback was implemented in the virtual environment. Rather than haptic cues, color-coded cues helped the participant perceive their performance. The colors red, yellow, and green indicated depth and accuracy (Figure 12). Green meant the tip of the pointer was on the surface and within the accurate range. Yellow meant the tip of the pointer was on the surface, but outside of the accurate range, such as touching the metal guide bars. Red meant the tip of the pointer was not on the surface at all. Participants were instructed to correct themselves based on the color indicators.

During the experiment, participants were instructed to keep a pace of three seconds to trace the full 25cm line in virtual reality. The experimenter checked the results and had participants redo the task if their pace was too fast or too slow.

Figure 12. Red, yellow, and green color indicators, respectively.

## 3.4 Technical Development

The experimental task in the physical environment was designed by this author. The experimental task in the virtual reality environment was designed and directed by this author, and developed by an experienced Computer Scientist. The virtual reality experience was developed using UnReal Engine, a gaming engine used for developing 3D interactive spaces [3].

Figure 13. Virtual environment designed in UnReal Engine.

# 4. Methodology

## 4.1 Design of Experiment

The three main factors were gender (either Male or Female), environment (either Physical or Virtual), and gap width (4mm, 5mm, 6mm, or 7mm). In part of the analysis, the environment factor was eliminated by taking the difference between each participant's virtual error ratio and physical error ratio. This gave the total change in performance per person, which illustrates how much the environment affected each participant's performance better than raw scores.

Forty-eight subjects (24 males and 24 females) were recruited from four different major departments at Cal Poly State University, San Luis Obispo. There was an equal number of males and females for every type of test. Twelve subjects were assigned to each of the four gap widths (4mm, 5mm, 6mm, and 7mm). Of the six males assigned to 4mm, half of them started in the physical environment and the other half started in the virtual environment. The same assignment method applied to all males and females at all gap widths. Every participant did the task in both the physical and virtual environment. At the end of the experiment, each participant completed a questionnaire about their experience and perceptions of various human factor dimensions.

## 4.2 Hypothesis

My hypothesis is that the virtual environment will have higher error ratios than the real environment. The goal of this experiment is to test if there is a significant effect of virtual reality on error ratios for a detailed task. It is hypothesized that:

- There is a significant difference between corresponding error ratios in the physical environment and the virtual environment

- There is a significant difference of error ratios between each gap width in either environment
- There is no difference between genders in either environment
- Those who prefer the virtual environment in the survey also have a lower error ratio

## 4.3 Variables

### 4.3.1 Independent Variables

There are four independent variables.

- Gap Width (4mm, 5mm, 6mm, or 7mm)– Each subject was randomly assigned to a gap width and performed all their tests with the same gap width.
- Gender (Male or Female)– An equal number of males and females performed all tests.
- Environment (Physical or Virtual)– Each subject performed the same tests in both the physical environment and the virtual environment. They were randomly assigned to start in one environment, and then continue to repeat the task in the other environment.
- Direction of Randomized Test Sequence (Physical to Virtual, or Virtual to Physical)– The test sequence refers to what order of environments the participant was randomly assigned to during the experiment. The two possible test sequences are (1) physical to virtual, and (2) virtual to physical. An equal number of males and females were randomly assigned to each test sequence.

### 4.3.2 Dependent Variables

- Error ratio (Equal to error time over total time)
- Change in average error ratio between environments
- Questionnaire results

The error ratio was calculated by dividing the error time over the total time. Three samples of error ratios were taken for each participant in each environment and then averaged. Each participant ended up with two scores, average physical error ratio and average virtual error ratio. The change in average error ratio between environments was also calculated and used for part of the analysis.

In the physical environment, error time was defined by amount of time spent touching the top or bottom metal guides. The total time was measured as the time between when the experimenter said “go” and when the participant said “stop” as soon as they reached the end of the line. The experimenter started and stopped a timer on these verbal cues.

In the virtual environment, the program measured both error time and total time. Error time was counted when the user was either not directly on the base surface or when they were touching or outside of the bounds. The total time was counted when the stylus was in contact with the base plate. The total time was controlled by the experimenter, who cleared the clock when they said “go” and stopped the clock as soon as the participant reached the end of the line. Since it was possible for the user to stick the stylus through the base plate, the total time detection zone started at the surface and descended into the base by a few centimeters (see Figure 14).

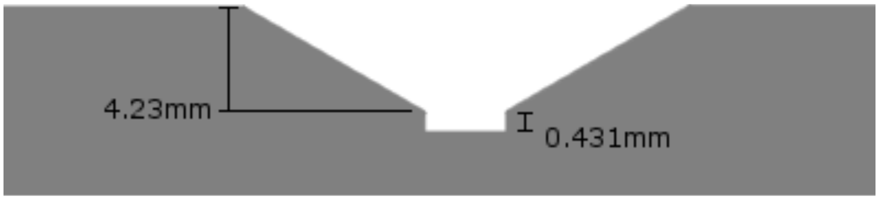

Figure 14. Geometry of cross section of virtual board on its side.

In addition to the quantitative error ratios, the dependent variables also included the questionnaire results from each participant. Participants were asked to rate their perceptions of certain human factor measures, such as reliability and comfort, on a scale of 1 to 5. They were also asked to choose whether they preferred the physical or virtual environment for two scenarios: overall and specifically for training for this task. They were also solicited for comments to explain their ratings, likes, and dislikes about the environments.

### 4.3.3 Controlled Variables

- Location: The cubicle in the back-left corner of the Cal Poly Ergonomics Lab was used for all experimental data collection.
- Lighting: All room lights were on and window shades were closed.
- Set-Up: Inside the cubicle was a table, a participant chair, an experimenter chair on the other side of the table, a desktop computer facing the experimenter, and either physical equipment or virtual reality equipment set up in front of the subject.
- Apparatus: HTC Vive headset and controller, Lafayette Instruments straight-line steadiness tester, an electronic test lead glued inside a 1.125in diameter foam tube
- Pace: All participants were instructed to keep a pace of 3 seconds to trace the line. They were instructed to count to 3 seconds, either out loud or to themselves. The

average and standard deviation of actual total time were 3.027±0.764 seconds. The median of actual total time was 2.936 seconds.

- Procedure: Subjects were randomly assigned to start in one environment, and then switch to the other environment. All other procedural steps, including introduction, instructions, and conclusion, were the same for all participants.
- Hand Dominance: All participants were screened and selected for right hand dominance.
- Gender: An equal number of males and females were selected.
- Participant Consent Form (See Appendix A)
- Script and Instructions (See Appendix B)
- Post-Experiment Questionnaire (See Appendix C)

## 4.4 Participants

Subjects were recruited for this experiment through department emails to students in industrial engineering, mechanical engineering, biomedical engineering, and kinesiology. Subjects volunteered on a first-come, first-serve basis to fill open time slots. No incentive, other than the opportunity to experience virtual reality for the first time, was offered to recruit participants via email. However, donuts were provided at the end of the actual experiment as a surprise token of appreciation.

- Background: Cal Poly Student
- Majors: 28 mechanical engineering, 7 industrial engineering, 4 biomedical engineering, 3 kinesiology, 6 other majors
- Age: 18-26 years old (Sample average is 20.4±1.9 years old)
- Gender: 24 males and 24 females

The subjects were screened to meet the following criteria.

- Handedness: Right hand dominant
- Experience: Either no experience or only one prior experience with virtual reality
- Vision: Normal or corrected-to-normal vision, such as either contacts or glasses. All participants who wore glasses kept their glasses on while wearing the VR headset.

## 4.5 Equipment

- Lafayette Instrument Co. Model 54035 Clock/Counter
- Lafayette Instrument Co. Model 32010 Groove Type Steadiness Tester (straight-line steadiness test in Figure 15)
- Electronic test lead, glued inside a 1.125in diameter foam tube to replicate the grip size of the HTC Vive controller
- HTC Vive headset and controller (see specifications below)
- HTC Vive motion-tracking base stations setup, including tripods
- Desktop computer used by the experimenter
- Hygienic wipes to sanitize the headset, controller, and stylus between participants

Figure 15. Lafayette Instruments Groove Type Steadiness Tester.

Specifications for the HTC Vive are as follows:

- Display: OLED
- Resolution: 2160x1200 (1080x1200 per eye)
- Refresh Rate: 90 Hz
- Field of View: 110°
- Tracking Area: 15ft. x 15ft.

The HTC Vive virtual reality system was chosen, as opposed to the Oculus Rift system, due to its availability on the market at the time this experiment began in November 2016. The HTC Vive system included two handheld controllers in the system, whereas the Oculus Rift had not yet released its Oculus Touch controllers. The complete HTC Vive system was available for immediate shipment in November 2016. The purpose of this experiment is to study precision tasks using existing, available VR hardware. No specialized equipment, such as haptic gloves or a haptic stylus, was purchased for this experiment due to budget constraints and availability constraints. A haptic stylus was quoted for more than $2,000. Haptic gloves were deemed too specialized for the broader purpose of this study and capabilities of the experimenter’s resources. Funding from the

Industrial and Manufacturing Engineering Student Fee Committee and the Melfred Borzall Project Fund was applied for and granted to this experiment.

It should be noted that the desktop computer was built specifically for this experiment to meet the PC specifications of the HTC Vive. It was more economical to build a PC from separate components than to buy an off-the-shelf computer by Dell or a similar brand. All equipment purchased for this virtual reality system will stay in the Industrial and Manufacturing Engineering department for future work related to human factors engineering and ergonomics.

It should also be noted that a Leap Motion controller was also purchased during the early development of this thesis experiment. The Leap Motion controller is designed to integrate with the HTC Vive headset. The controller tracks free hand and finger position and movement. It took several iterations of programming with the Leap Motion before starting the first pilot tests. The first pilot tests proved unsuccessful and the Leap Motion controller was ditched and replaced with the HTC Vive controllers. More information about the pilot tests can be found in section 4.9 Pilot Tests.

## 4.6 Task

The task was to draw a straight line from left to right, 25cm long, while staying within a certain gap tolerance and not touching the top or bottom guides. Participants were instructed to use their right hand to hold the stylus like a TV remote control, with their palm on the right side of the barrel, fingers curled around the bottom, and thumb on top. If the tip of the stylus contacted the top or bottom metal guides, an electrical circuit was completed and a clock counter recorded how long the subject was in contact with the top. The experimenter counted aloud three seconds to demonstrate the pace needed to trace the entire line before starting the timer. The participant counted aloud to themselves or in their head during the actual task.

The actual total time it took to draw the line was recorded by the experimenter in a spreadsheet. The experimenter had participants redo the task if their pace was too fast or too slow. The task was repeated three times, and the average error ratio (error time over total time) was calculated for each environment. Each participant has an average physical error ratio and an average virtual error ratio for their assigned gap width.

## 4.7 Experimental Procedure

The experiment lasted approximately 10 minutes and the questionnaire took approximately 10 minutes to complete. Participants were scheduled into 25min time slots.

1) Greeting
    a. When the participant arrived, they set their belongings on a chair away from the experimental setup to eliminate distractions. The participant took a seat at the table and were given the informed consent form (Appendix A) to read and sign.
    b. The experimenter read the introduction script (Appendix B).
2) Environment 1
    a. The participant was randomly assigned to start in either the Physical or the Virtual environment. The correct equipment was already set up to begin.
    b. The experimenter read the instructions for this task.
    c. The participant was given 3 warmup trials that were not recorded to ensure they understood the task.
    d. The participant was asked if they had any questions and if they were ready to begin.

   e. The recorded task began. The participant repeated until there were 3 sufficient samples of data for this environment.

3) Switch environments
   a. When the participant was done with Environment 1, the experimenter set up the equipment for the next environment.
4) Environment 2
   a. The experimenter read the instructions for this task.
   b. The participant was given 3 warmup trials that were not recorded to ensure they understood the task.
   c. The participant was asked if they had any questions and if they were ready to begin.
   d. The recorded task began. The participant repeated until there were 3 sufficient samples of data for this environment.
5) Conclusion and Questionnaire
   a. After the participant had completed both tasks, the experimenter read a concluding statement.
   b. The participant was given a written post-experiment questionnaire (Appendix C) to fill out and return.
   c. The participant took the questionnaire to a different cubicle so they could concentrate and fill it out privately and at their own pace.

Figure 16. The experimenter giving instructions to the participant during the virtual environment task.

Figure 17. The experimenter collecting data during the virtual environment task.

## 4.8 Measures

The main measurements of data were total time to complete the task and total error time. This was measured with a timer in the physical experiment and with the software in the virtual experiment. The other measurements were obtained through the questionnaire. These measurements include rating various user experience dimensions, such as difficulty, comfort, reliability, strain, and effectiveness, on defined scales of 1 to 5. Participants were also asked to state their preferences and write comments and suggestions for improvement.

## 4.9 Pilot Tests

The very first iteration of this experiment used a Leap Motion controller in conjunction with the HTC Vive headset. The Leap Motion was meant to replace the HTC Vive handheld controllers due to its advertised functionality of tracking finger movement in free space. After completing all the software development for this test setup, pilot tests were run. These first pilot tests showed that this experiment would be impossible to collect reliable data with the Leap Motion controller. Most of the issues encountered with Leap Motion involved the virtual hand, which jittered asynchronously with the user’s hand, jumped to different positions in space, and randomly inverted upside-down. Neither the participant nor the experimenter could control when these idiosyncrasies would occur or fix them. Thus, the Leap Motion controller was taken out and the entire experiment was redesigned using only the handheld controllers that came with the HTC Vive system.

It was assumed that organizations using virtual reality for training purposes will buy the HTC Vive kit alone, without separately purchasing the Leap Motion controller. The HTC Vive headset and controllers are already meant to work together, so it was natural to use the included handheld controllers without Leap Motion. This experiment should be helpful for people who are trying to use commercially available virtual reality systems, such as the HTC Vive, rather than customized or excessively expensive equipment.

The second iteration of pilot tests with the HTC Vive controllers was a success. Four males and three females participated. Very few changes were made to the experiment after the pilot tests. The results from the pilot test are not included in the results section of this report.

# 5. Results

The results are segmented into quantitative results, qualitative results, and observations. In the Quantitative Results section, the data is presented in two forms: (1) average error ratio per environment per person and (2) the change in average error ratio between environments per person. The average error ratio per environment is an indication of each person's "raw score" in each environment. It is calculated by taking the participant's error time and dividing by total time. The error times varied per person, but the actual total times for the entire sample were 3.027±0.764 seconds.

The change in average error ratio between environments is an indication of how much a person's individual performance was affected by switching to a different environment. With this data, the magnitude of the effect can be compared across all participants. The change in average error ratio between environments was calculated per person using the following formula.

$$\Delta\ Avg\ Error\ Ratio = (Virtual\ Avg\ Error\ Ratio) - (Physical\ Avg\ Error\ Ratio)$$

The results of 48 participants (24 males and 24 females) were used in the quantitative analysis. There were three additional participants, for a total of 51 people, who had completed the experiment, but a glitch in the virtual reality system recorded incorrect times for these three people. Their quantitative data was thrown out and new participants were recruited to replace the results. The three original participants were not made aware of the glitch and continued with the experiment and questionnaire normally. Since they were not aware of the glitch and could comment candidly on their experience, some of their feedback regarding comfortability, likes, and dislikes were saved and included in the qualitative results section.

## 5.1 Quantitative Results

### 5.1.1 Average Error Ratio Per Environment

Data must pass a normality test before further analysis. The residuals of this data passed the normality test with a p-value of 0.183, which allowed the analysis to continue (Figure 18). The test for equal variances of the residuals versus gap, gender, and environment factors yielded a p-value of 0.109, which means we failed to reject the null hypothesis that all variances are equal (Figure 19). The assumptions to run an ANOVA were met at an alpha level of 0.01. A summary of the collected data is best shown graphically in Figure 20, which shows the average error ratios in each environment at different gap widths.

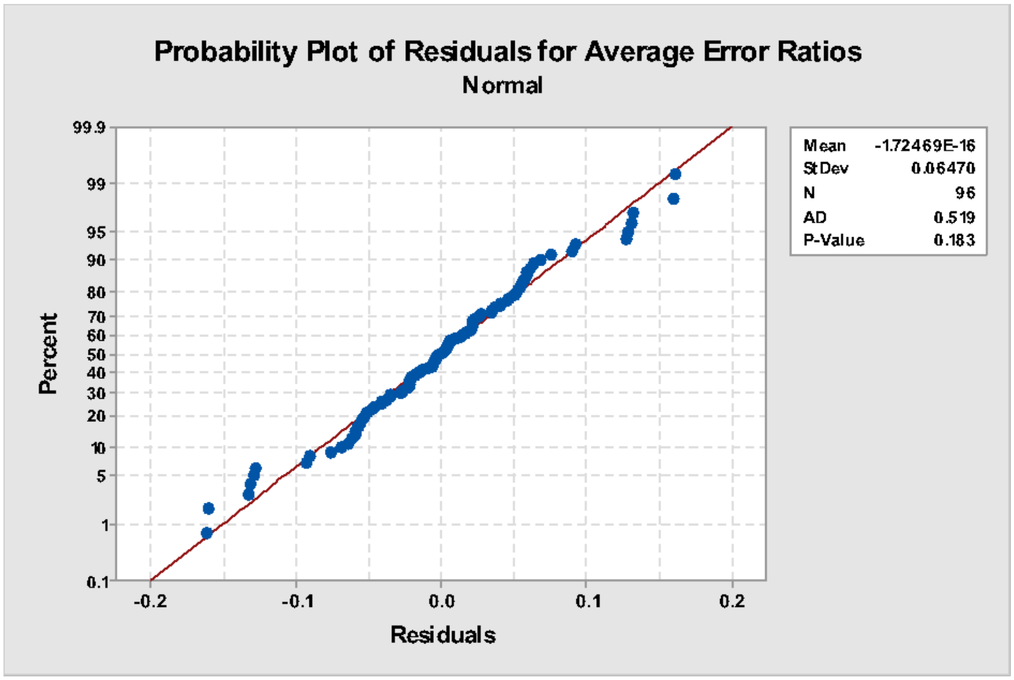

Figure 18. Normality test of residuals for average error ratios.

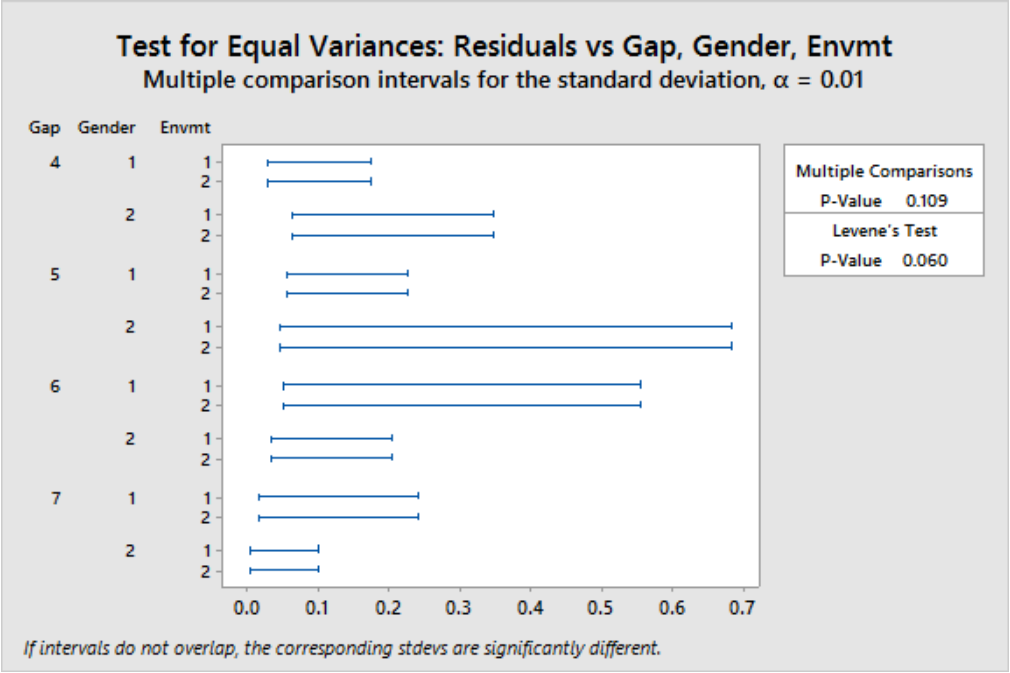

Figure 19. Test for equal variances of residuals of average error ratios vs. gap, gender, environment.

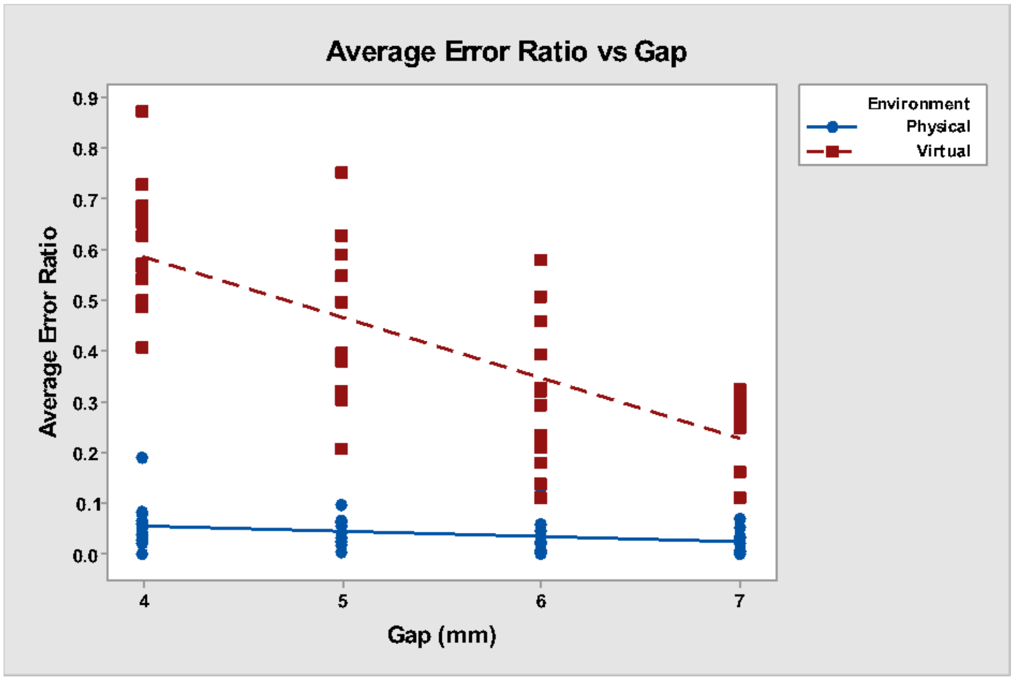

Figure 20. Scatterplot of average error ratios in each gap category, separated by environment.

As expected, the average error ratios in the virtual environment are higher than the average error ratios in the physical environment at all gap widths, as shown in Figure 20. The range of average error ratios is much wider in the virtual environment than the range in the physical environment. This means that between participants of all genders, there is greater variability of accuracy in virtual reality.

Table 1. ANOVA for Data: Average Error Ratio. Effects: Gap, Gender, Environment, and Block(Gap, Gender). Interaction between Environment and Gap is significant.

```
Analysis of Variance

Source                     DF   Adj SS   Adj MS  F-Value  P-Value
  Gap                       3  0.53005  0.17668    17.77    0.000
  Gender                    1  0.00108  0.00108     0.11    0.743
  Environment               1  3.23299  3.23299   325.19    0.000
  Gap*Gender                3  0.00692  0.00231     0.23    0.873
  Environment*Gap           3  0.36600  0.12200    12.27    0.000
  Environment*Gender        1  0.01605  0.01605     1.61    0.211
  Environment*Gap*Gender    3  0.00770  0.00257     0.26    0.855
  Block(Gap, Gender)       40  0.34773  0.00869     0.87    0.663
Error                      40  0.39767  0.00994
Total                      95  4.90620

Model Summary
        S    R-sq  R-sq(adj)  R-sq(pred)
0.0997085  91.89%     80.75%      53.31%
```

The interaction between Environment and Gap has a p-value of less than 0.001. This result indicates that the interaction between Environment and Gap has a significant effect on the average error ratio. The interaction plot in Figure 21 shows that average error ratios simultaneously increase and spread out when Environment and Gap change. A Tukey Pairwise Comparison test is performed in the next section to find which means are significantly different from each other between gap widths.

As the gap width decreases from 7mm to 4mm, it is assumed that task difficulty increases. Without interaction, we might only see an incremental increase in average error ratio for each gap width, with no difference between environments. Without interaction, we might also observe a constant increase in average error ratios between environments. However, these are not the case. Interaction between both Environment and Gap has the most significant effect on average error ratio.

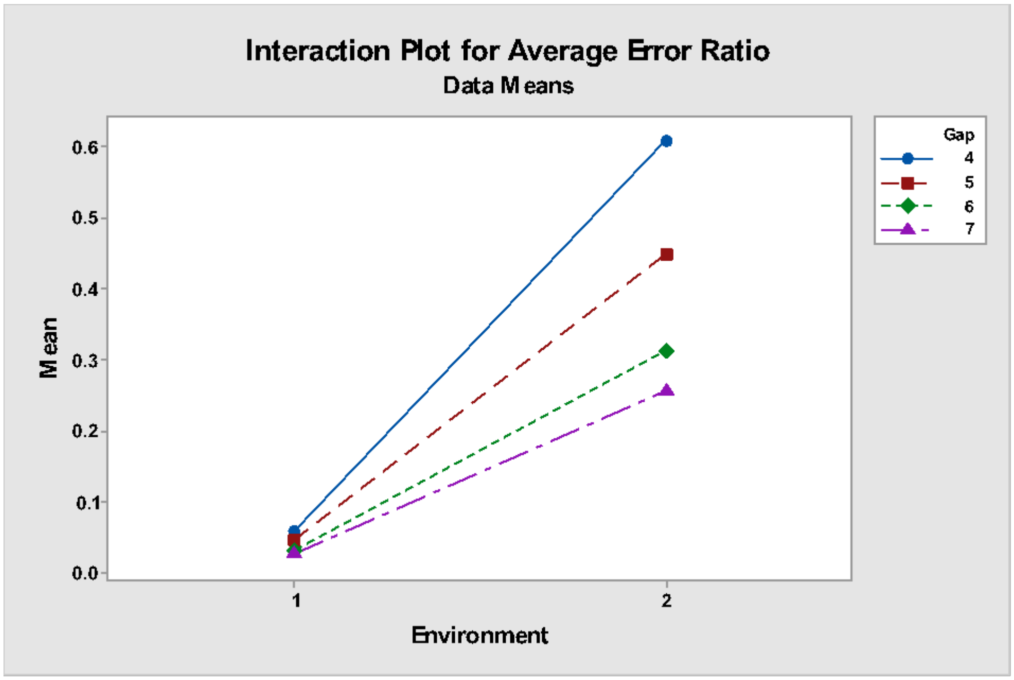

Figure 21. Interaction plot between Environment and Gap effects. Environment 1 is the Physical Environment and environment 2 is the Virtual Environment.

The effect of Direction of Randomized Test Sequence (whether a participant started in the virtual or physical environment) was also run in an ANOVA, and found to have no significant effect on results, with a p-value of 0.710 (Table 8, Appendix D). Balancing the Direction factor was meant to mitigate the effects of “learning” in the repeated measures design. An example of the learning effect would be if all participants started in the physical environment, learned how to do the task, and then performed the task better in the virtual environment, thus skewing the results. The ANOVA results indicate that Direction, and therefore the learning effect, had no significant effect on the results in a balanced repeated measures design of experiment.

The interaction plot above shows a significant increase of average error ratios in the virtual environment. A Tukey Pairwise Comparison test was used to find which means are significantly different from each other at a 97% confidence level. The test was applied to compare means across the four different gap widths specifically in virtual reality. The test shows that the mean at 4mm is not grouped with any of the other means. Subsequently, the means for 5mm and 6mm are grouped together and the means for 6mm and 7mm are grouped together. According to the test output, in Table 2, the mean at 4mm is significantly different from the means of all other gap widths in the virtual environment. This comparison also indicates that the mean average error ratios, if expressed as error percentages, are 61% at 4mm, 45% at 5mm, 31% at 6mm, and 26% at 7mm.

Table 2. Tukey Pairwise Comparison for Average Error Ratios only in VR

```
Tukey Pairwise Comparisons:
Response = Virtual Average Error Ratio, Term = Gap

Grouping Information Using the Tukey Method and 97% Confidence

Gap2   N      Mean  Grouping
4     12  0.608500  A
5     12  0.448333     B
6     12  0.312000     B  C
7     12  0.256333        C

Means that do not share a letter are significantly different.
```

### 5.1.2 Change in Average Error Ratio Between Environments

The change in average error ratio (Δ average error ratio) is the virtual error ratio minus the physical error ratio. The average error ratios in the virtual environment were always higher than the ratios in the physical environment, so all Δ average error ratios were positive values.

The residuals of this data passed the normality test with a p-value of 0.247, which allowed the analysis to continue (Figure 22). The test for equal variances of the residuals versus gap and gender factors yielded a p-value of 0.040 (Figure 23). At an alpha level of 0.01, we failed to reject the null hypothesis that all variances are equal. The assumptions to run an ANOVA were met. A summary of the collected data is best shown graphically in Figure 24, which shows the Δ average error ratios at different gap widths.

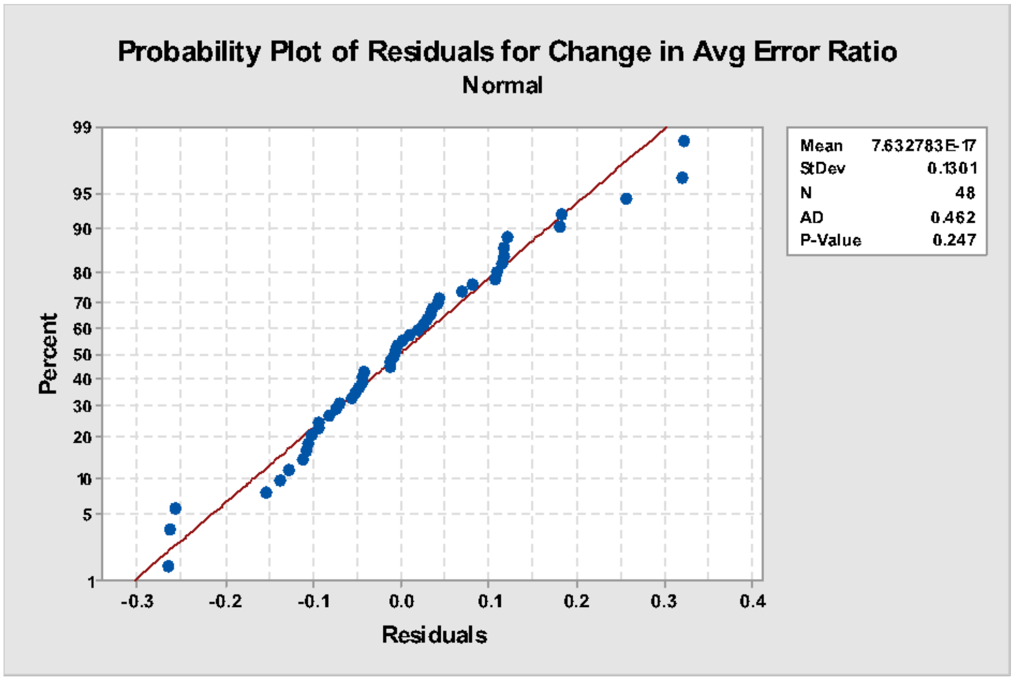

Figure 22. Normality test of residuals for change in average error ratios.

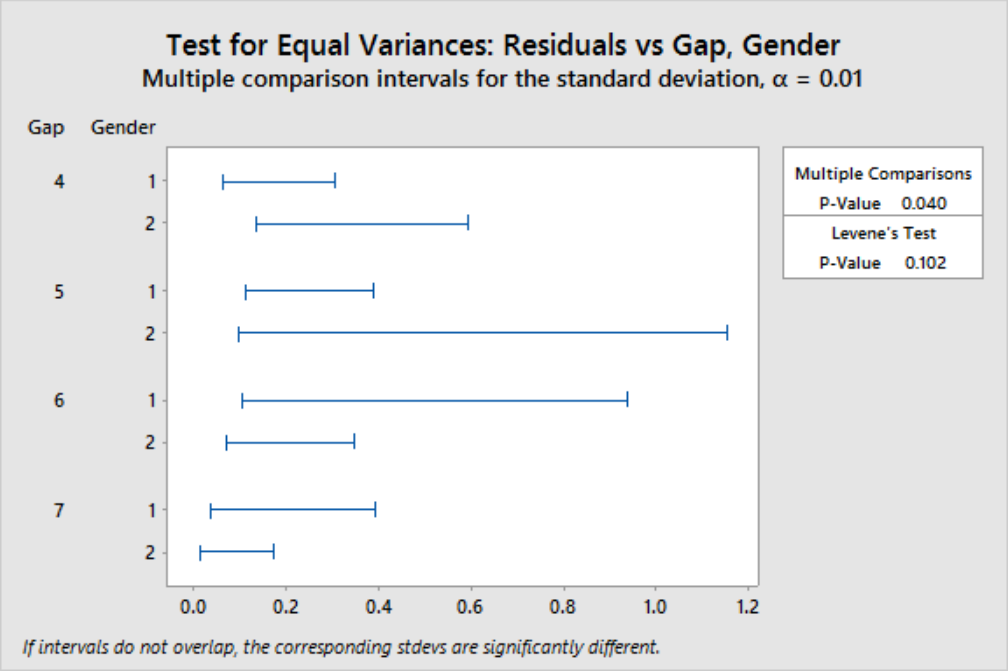

Figure 23. Test for equal variances of residuals of change in average error ratio vs. gap and gender.

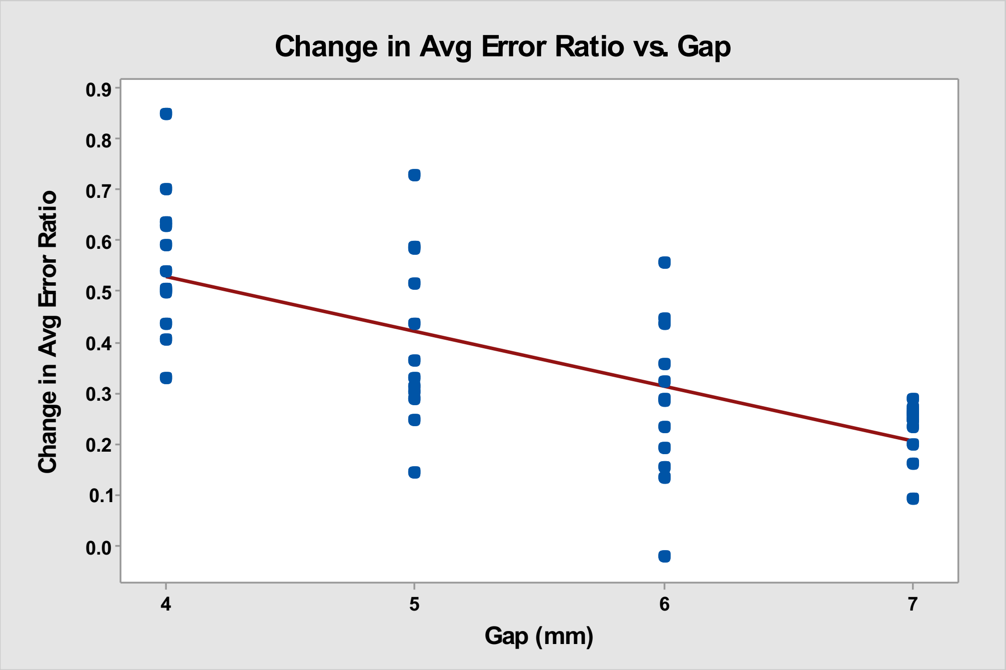

Figure 24. Scatterplot of change in average error ratio (virtual - physical) at each gap width.

The distribution of the changes in average error ratio is shown above in Figure 24. As the gap size increased, the difficulty of the test decreased. The participant's change in performance improved with an increase in gap size. This means that at larger gap widths, participant's performance in the virtual environment was closer to their performance in the physical environment.

The change in average error ratio represents how much each participant's performance was affected between the two environments. The ANOVA for the Change in Average Error Ratio data produced the same results as the Average Error Ratio data. The ANOVA results for Change in Average Error Ratio can be found in Table 9 in Appendix D.

A Tukey Pairwise Comparison test was used to find which means are significantly different from each other at a 97% confidence level. The test was applied to compare means across the four different gap widths. The test shows that the mean at 4mm is grouped with the mean at 5mm. Subsequently, the means for 5mm and 6mm are grouped together and the means for 6mm and 7mm are grouped together. According to the test output, in Table 3, the mean at 4mm and the mean at 7mm are significantly different. This means that a physical gap width increase of 3mm, which is a relatively small measurement, produced significantly better performance results.

Table 3. Tukey Pairwise Comparison for Changes in Average Error Ratio

```
Tukey Pairwise Comparisons:
Response = Change in Avg Error Ratio, Term = Gap

Grouping Information Using the Tukey Method and 97% Confidence

Gap   N      Mean  Grouping
4    12  0.551365  A
5    12  0.403636  A   B
6    12  0.281750      B  C
7    12  0.231352         C

Means that do not share a letter are significantly different.
```

### 5.1.3 Preference versus Performance

Participants responded to a questionnaire at the end of the experiment. The questionnaire asked for the participant's overall preference of the physical environment or virtual environment. The participant was also asked which environment they would recommend for training somebody else in the same task. It is expected that those who overall prefer the virtual reality environment have less change in their performance between the environments. One participant did not choose an overall preference.

The change in average error ratios was used to represent the change in performance per participant. The preferences and opinions of each participant from the questionnaire were corresponded with their change in average error ratio. The residuals passed the normality test with a p-value of 0.489 to allow further analysis under this assumption.

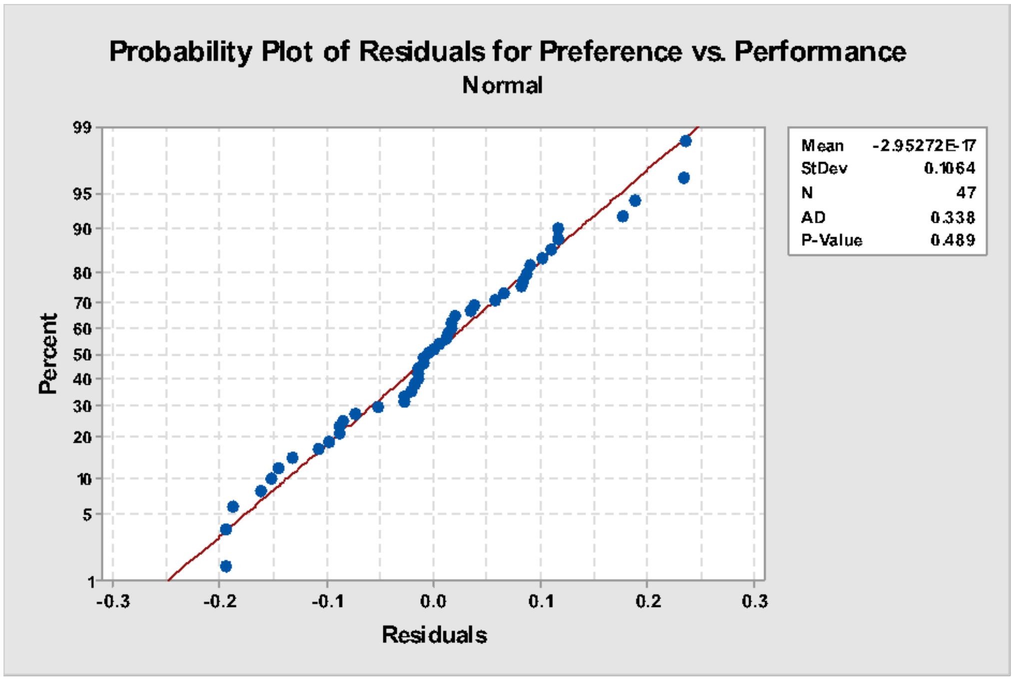

Figure 25. Normality test of residuals for preference versus performance.

A two-sample t-test was performed to see if there was a significant difference between the mean change in average error ratios of participants who overall preferred the physical environment and participants who overall preferred the virtual environment. With a p-value of 0.332, there is not a significant difference between the average change in performance for participants who preferred the physical environment and those who preferred the virtual environment (Table 4). Notably, 30 participants preferred the physical environment while 17 participants preferred the virtual environment, and one person did not respond. Participants' feedback, expanded in the Qualitative Results section, gives some indication of their likes and dislikes which may have influenced their preference.

Table 4. Two-sample T-test for Change in Average Error Ratio and Environmental Preference

```
Two-sample T for Change in Average Error Ratio

OverallPref   N   Mean  StDev  SE Mean
Physical     30  0.382  0.184    0.034
Virtual      17  0.328  0.179    0.044


Difference = μ (Physical) - μ (Virtual)
Estimate for difference:  0.0541
95% CI for difference:  (-0.0576, 0.1657)
T-Test of difference = 0 (vs ≠):
T-Value = 0.98  P-Value = 0.332  DF = 34
```

Participants were asked which environment they preferred overall (Overall Preference) and which environment they preferred specifically for training in this task (Training Preference). The Overall Preference of certain environment has a significant association with the change in average error ratios at a p-value of 0.027 (Table 5). In general, the participants who preferred the virtual environment had less change in performance between the two environments. Gap width also had a significant association with change in performance at a p-value of less than 0.001, which is consistent with earlier findings. It was found that Training Preference had no significant effect on performance, with a p-value of 0.855 (Table 11, Appendix D).

Table 5. ANOVA for Data: Change in Average Error Ratio. Effects: Overall Environmental Preference, Gender, Gap.

```
Analysis of Variance

Source                     DF   Adj SS     Adj MS  F-Value  P-Value
  OverallPref               1  0.09058   0.090578     5.39    0.027
  Gender                    1  0.02826   0.028261     1.68    0.204
  Gap                       3  0.70148   0.233825    13.92    0.000
  OverallPref*Gender        1  0.00909   0.009087     0.54    0.468
  OverallPref*Gap           3  0.06150   0.020500     1.22    0.319
  Gender*Gap                3  0.02207   0.007356     0.44    0.727
  OverallPref*Gender*Gap    3  0.05131   0.017102     1.02    0.398
Error                      31  0.52060   0.016793
Total                      46  1.52637
```

## 5.2 Qualitative Results

Each participant responded to a questionnaire after completing the experiment. The questionnaire asked participants to rate the physical and virtual environments for five user experience dimensions: difficulty, comfort, strain, reliability, and effectiveness. The following Table 6 shows the average score of each dimension for each environment. A two-sample t-test was performed on each user experience dimension. At an alpha level of 0.05, the ratings of difficulty, reliability, and effectiveness were significantly different between environments. The ratings of comfort and strain had low p-values of 0.063 and 0.086, respectively, but were not considered significantly different between environments. In every category, virtual reality scored worse than the physical environment based on the defined scale from 1 to 5. It is surprising that comfort and strain were not significantly different because users had to wear a headset in virtual reality, which could have been uncomfortable for some participants.

Table 6. Average Ratings of User Experience Dimensions

| | Physical Environment Average Score | Virtual Environment Average Score | P-value of Two Sample T-Test |
|---|---|---|---|
| Difficulty<br>(1=difficult and 5=easy) | 3.729 | 2.729 | 0.000 |
| Comfort<br>(1=uncomfortable and 5=very comfortable) | 4.250 | 3.936 | 0.063 |
| Strain<br>(1=least strain and 5=most strain) | 2.146 | 2.500 | 0.086 |
| Reliability<br>(1=least reliable and 5=most reliable) | 4.452 | 4.000 | 0.005 |
| Effectiveness<br>(1=least effective and 5=most effective) | 4.292 | 3.250 | 0.000 |

Out of all the questionnaire responses, the following statements are noteworthy:

- 32 participants said they were better able to perceive depth in the physical environment due to the haptic feedback of touching the board
- In both the physical and virtual environments, 41 participants felt minor strain in their hand, forearm, or arm while holding up the controller and stylus during the task
- 16 participants said the mirrored background and the metal were difficult to differentiate on the physical equipment, making the task more difficult. They also mentioned that the screeching sound of metal on glass was distracting. (Note: This noise occurred in only a handful of trials.)
- 29 participants said depth perception was difficult in the virtual environment due to the lack of haptic feedback

- 25 participants said being able to use virtual reality was “cool”, “fun,” or “interesting”
- 15 participants stated that they liked the color-changing indicators in virtual reality because it gave immediate feedback of their performance
- 14 participants said the VR headset felt heavy or uncomfortable
- 13 participants suggest adding haptic feedback to improve the virtual environment
- 5 participants did not like not being able to see their hand or the experimenter in the virtual environment
- 4 participants said the headset screen was pixelated
- 1 participant said she or he felt uncomfortable knowing that someone is near them but not being able to see that person in the virtual environment
- 1 participant said, “Even though [virtual reality] looks unreal, I felt my brain wanted to think it was real.”

## 5.3 Observations

The experiment design process took longer than expected because of the limitations in virtual reality that had to be considered and worked around. Some of the decisions made to consciously differentiate the environments (i.e. color-coded indicators, removing vibrational feedback) were the best way to compensate these limitations at the time of the experiment.

Since virtual reality is a new and expensive technology, it was easy to find 51 people who had never experienced it before to participate in the experiment. After the experiment was finished, many participants asked questions about the technology and wanted to learn more about future opportunities involving VR in the college. The following is a list of observations made throughout the experiment.

- Two participants had extreme difficulty getting the VR test to work. When the timer started, the colored background would turn red. This issue was resolved after trying again several times.
- Three trial runs were given to each participant in the instructions. In the physical environment, most participants used all three trial runs. In the virtual environment, most participants asked for and used more than three trial runs, approximately five or six.
- The software developer who programmed the VR environment is color blind. He chose colors that he could see, so it was assumed that the experiment was accessible for color blind participants. One participant said she or he is color blind in the post-experiment questionnaire, which could have affected their ability to recognize the color indicators in VR. Their results were within the normal range and were not thrown out.
- Two participants noted that they have genetic or above average natural hand shakiness. One participant noted that she or he has tendonitis. Their results were within the normal range and were not thrown out.
- Five participants wore glasses. The headset fit over their glasses, so they all kept their glasses on during the experiment. Three participants commented on the discomfort of wearing glasses under the headset.
- Four participants noted “slight disorientation” when they first put on the virtual reality headset, but they all said they adjusted quickly. Only one person said his or her eyes felt strained after. Another person said she or he felt a little dizziness after taking the headset off, but not during the experiment.

# 6. Discussion

This experiment measured and analyzed the performance of participants conducting a straight-line precision task in virtual reality. Forty-eight participants were asked to trace a straight line within a certain gap tolerance. The task was designed to be unfamiliar because of the way the stylus was held, gripped like a TV remote rather than like a pencil, to trace the line. The participants were training how to do the task in both reality and virtual reality to emulate precise motions of certain industrial operations or physical therapy. The total time to trace the line and the error time were recorded to determine average error ratios. The average error ratios of each participant represented relative scores of how well each participant performed in each environment.

The data showed that in both environments, error increased as the difficulty of the task increased. At lower gap widths, the average error ratios were higher. The data also showed that a participant's average error ratio was always higher in the virtual environment than in the physical environment. These two observations do not act independently. There is a significant interaction between gap width and environment, at a p-value of less than 0.001. The interaction indicates that performance was significantly affected by a combination of gap width and environment factors.

Each participant had two average error ratios, one for the physical environment and one for the virtual environment. The difference between these two ratios quantifies how much a person's performance changed between environments. Every participant had a higher error ratio in the virtual environment than in the physical environment. In the straight-line test, the gap width represented precision tolerance, measured at 4mm, 5mm, 6mm, or 7mm. The change in average error ratio at 4mm is significantly higher than the change in average error ratio at 7mm, at a 97% confidence level. The results suggest that as the difficulty of the task increases (i.e. the gap width shrinks from 7mm to 4mm), the change in performance is significantly worse.

The mean average error ratios in virtual reality show just how much error VR developers could expect when designing a precise linear target like the straight-line test. If the mean average error ratio in VR is represented as an error percentage, VR developers could expect 26% error for 7mm-wide targets, 31% error for 6mm targets, 45% error for 5mm targets, and 61% error for 4mm targets (Table 2). Depending on the purpose of the VR experience, perhaps a 4mm target tolerance is preferred for more difficult tasks, whereas a 7mm target tolerance is preferred for easier tasks.

When looking at user preferences, there was a significant association of overall environment preference with the change in average error ratio (p-value of 0.027). In general, the participants who preferred virtual reality had less change in their error performance between the two environments. This could indicate that participants who liked the virtual environment also performed better in the virtual environment. This is an interesting finding that suggests that users could be motivated to perform better in virtual reality due to positive expectations and opinions of the technology. Future work would be needed to further investigate this association.

## 6.1 Limitations of the Study and Sources of Error

The development of this study included many iterations of tasks and software programming. Due to the experimenter's unfamiliarity with game engine programming for virtual reality, a consultant with a Bachelor of Science in Computer Science was recruited to assist with the technical development of this experiment. Despite online tutorials and software help functions, it can be difficult for the average person to learn how to create a sophisticated program in UnReal Engine without formal education or extensive experience. The steep learning curve of VR software is a major inhibitor to widespread adoption of virtual reality. It has also created a very lucrative market and high demand for VR programmers and developers. The success of this VR program was reliant on the abilities of the computer scientist to replicate the physical environment and equipment functions, as well as to develop creative work-arounds for issues such as lack of haptic feedback.

The experimenter had strong interest in, but limited experience with, virtual reality before starting this thesis project. That was not enough to start this project without outside guidance from experts in virtual reality. An extensive literature review and several discussions with local VR experts were helpful and essential to understanding the complexities and challenges of virtual reality. Part of the purpose of this experiment was to study the effectiveness of off-the-shelf hardware systems for a precise task so that others can create and improve future training programs with similar consumer systems, such as the HTC Vive. The development process of this study indicates that virtual reality for training for a precision task is not something that one implements overnight. Rather, it is difficult to implement without previous experience.

A potential source of error in data collection was the method of tracking time data. The physical timer was dependent on the experimenter clicking the timer button at the right time. The experimenter controlled when the timer started because the experimenter said “go” and started the timer simultaneously. However, the experimenter waited until the participant said “stop” to stop the timer, but could not see the participant’s hand because they were on opposite sides of the table. The experimenter read instructions on how to keep the correct pace of 3 seconds, but there was variation of approximately ±1 second. The experimenter used discretion in judging whether a sample was too fast or too slow and had the participant repeat the test if the results were too far off.

In the virtual reality program, the timer began when the program detected the proximity of the pointer to the target area. To start the task, the participant would hold the pointer to the left of the straight line. The experimenter zeroed the timer as soon as they said “go” and the participant started to move his or her hand. The experimenter could see a replica of the virtual program in real time on the computer monitor, and stopped the timer as soon as the participant reached the end of the line, without any verbal cues. Again, the experimenter used discretion in judging whether a sample was too fast or too slow.

One factor that was not analyzed is the tradeoff between speed and accuracy during the straight-line steadiness test. Because total times varied in every sample, even with the same participant, the error ratio was calculated to normalize the data. However, looking at the raw data, there is a possibility that participants who were slower to trace the line had less error because they could concentrate. This effect was not tested, but it is nonetheless relevant and possibly significant. The median total time was 2.936 seconds, which is slightly faster than the recommended 3 seconds pace. The effect of speed versus accuracy may have some effect, however it was not accounted for in this study.

The HTC Vive controller has several embedded motion tracking sensors throughout the remote. The sensors located in the handle portion are far enough back that they could be affected by the wrist position and angle of the user. In this study, the precise positional relationship between the sensors in the controller and the user’s wrist were not measured. The movement of the virtual controller corresponds with the sensors in the physical controller, so there could be some source of error that derives from the way the controller was modeled in VR.

Calibration of the equipment is another potential source of error, but no statistical analysis was performed to address this. The experiment was set in a shared classroom and the data collection period spanned over nine days due to scheduling availability.

Between two to fourteen participants were scheduled for each day of data collection. The classroom is locked and a class of 24 students meet once a week in the room. Due to the semi-public nature of the classroom, all test equipment was set up and taken down every day to ensure equipment protection and security. The positions of all the equipment were measured and marked with tape to ensure consistency between days. The virtual reality equipment was calibrated the same way every single day. Despite best efforts to maintain consistency, we recognize that there could have been small differences in calibration due to the frequency of setting up and taking down the equipment. No statistical analysis was performed to investigate this effect.

### 6.2 Participation and Feedback

Participants were recruited from engineering and kinesiology departments at Cal Poly via department emails that linked to an online form. As soon as the department emails were sent out, there was an overwhelming and quick response to fill up available time slots. The emails screened participants for right-handedness and no prior experience with virtual reality. It should be noted that the 24 available spots for males filled up at a rate nearly twice as fast as the 24 available spots for females. No incentives were offered to participants in the recruitment email.

Feedback collected from the questionnaires was fairly consistent. More than half of participants described virtual reality as “cool,” “fun,” or “interesting.” The functionality of physical equipment is limited, whereas virtual reality can be programmed to do almost anything. Even though the virtual setup was not exactly the same as the physical setup, the additional features of color-changing and decluttering the visual space (by eliminating wires, non-essential objects, and the experimenter) were added to help the user perform the task. Color-changing and decluttering the visual space would be difficult to implement in the physical environment at the same speed and ease as virtual reality.

Despite the possibilities of virtual reality, participants still struggled. A majority, 32 out of 48, of participants noted that it was difficult to perceive depth in virtual reality. This could be due to a variety of reasons, such as the lack of haptic feedback in the controller, the pixelated resolution of the HTC Vive head-mounted display, or the non-reflective surface of the straight-line test equipment in virtual reality. This feedback suggests that virtual reality needs better visual quality for people to perform precision tasks. Haptic feedback was not implemented in this study because the vibration of the controller might shake the user's hand involuntarily and induce additional error. The haptic feedback component of gripped devices, like the HTC Vive controller, needs additional research for precision tasks in virtual reality applications.

# 7. Conclusion & Future Work

This study investigates the effects of virtual reality on user performance while training for a precision task. In general, virtual reality yielded higher error ratios than the physical test environment. As the difficulty of the task increased, the performance in VR degraded significantly. A combination of environment and gap width factors significantly affected users' average error ratios. As a result, the differences between the mean error of each gap width were magnified and much more noticeable in virtual reality than in physical reality.

The results of this research point to the importance of haptic feedback in virtual reality. For an unfamiliar and precise task, the current handheld controllers are not the best interaction tool. Versatile shapes of controllers may need to be experimented with to determine how best to marry technological capability of the developer and functionality for the end user. When the Leap Motion controller was briefly used in the development of this experiment, the output was too jumpy and unpredictable to produce reliable data for analysis. The Leap Motion would have allowed precise finger tracking, but did not work the way in which it was intended.

Virtual reality has great potential in applications for abled and disabled persons. This study focused on training for a linear precision task. Linear movement is a basic motion that can be found in industrial operations, such as cutting or soldering, and in disability rehabilitation, such as physical therapy for stroke patients. The results of this experiment show that virtual reality still needs significant improvements to get the same results as a physical task, even for something as simple as tracing a straight line.

The development of this study was challenging for the experimenter without substantial prior experience with virtual reality. It would be difficult for the average person to develop an interactive virtual reality program, such as a precision task for training, without formal computer science education or experience. The steep learning curve of VR software is a major inhibitor to the widespread adoption of virtual reality. Even with consumer-geared products such as the HTC Vive, the setup of virtual reality for the purpose of training poses several challenges. The currently available equipment does not seem adequate to train people in precision tasks in VR.

## 7.1 Future Work

The straight-line test focused on a simple, linear, and precise movement. Recommendations for future work extend to a variety of precision tasks that have more levels of complexity and different movements. Tasks for future work can be three-dimensional, such as pick and place. Perhaps different postures such as standing or reclining could affect task performance in virtual reality and mimic industrial operations such as training in manufacturing or maintenance work. Two-handed tasks could investigate hand-eye coordination of precise or small-scale assembly work in VR.

In this study, training in VR was studied as an alternative for physical training, rather than as a preparation tool for a physical task. Perhaps VR is not ready to replicate the exact motions and environment that one would experience in a physical system, but it may be an effective tool to prepare someone for a physical task. Future research could investigate the exact role and purpose of VR in training experiences.

Haptic feedback is clearly still an issue for virtual reality in precise applications. To continue future work in this experiment, vibrational feedback should be added in the handheld controller to test whether this type of feedback is an adequate depth indicator for participants. The results with and without vibrational feedback should be compared to see whether vibrations created more error or prevented error.

At the time of this study, the VR hardware and software were limited in its capabilities. The resolution in the head-mounted display was pixelated, which made it difficult to see the tip of the stylus. The field of view was limited to 110°, which could be widened to improve the peripheral view and sense of immersion. The handheld controllers could only be held a certain way and the design could be more ergonomic or interchangeable. Once new designs are released, researchers can test for performance of precision tasks in virtual reality using improved equipment.

Virtual reality is being developed at a rapid pace. Each iteration of the technology has a short life span and incrementally improves upon the previous version. The intensity of development in the field of virtual reality points to far-reaching questions for future work: How can precision tasks in virtual reality match real life tasks? How can virtual reality be improved to help training and rehabilitation? How can visuo-haptic feedback be optimized for an immersive experience? The opportunities capable through virtual reality are promising, and future work is needed to investigate the human factors implications of this new frontier.

# BIBLIOGRAPHY

[1] Akbar, N., & Jeon, S. (2014). Encountered-Type Haptic Interface for Grasping Interaction with Round Variable Size Objects via Pneumatic Balloon. *Haptics: Neuroscience, Devices, Modeling, and Applications, 1*(1), 192-200.

[2] Applications of Virtual Reality. (n.d.). Retrieved from https://www.vrs.org.uk/virtual-reality-applications/

[3] Baird, J. (2017). Delaney's Project [UnReal Engine]. Unpublished software program.

[4] Eck, U., Pankratz, F., Sandor, C., Klinker, G., & Laga, H. (2015). Precise Haptic Device Co-Location for Visuo-Haptic Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics, 21*(12), 1427-1441.

[5] Endo, T., Kawasaki, H., Mouri, T., Ishigure, Y., Shimomura, H., Matsumura, M., & Koketsu, K. (2011). Five-Fingered Haptic Interface Robot: HIRO III. *IEEE Transactions on Haptics, 4*(1), 14-27.

[6] Goble, A., Collins, A., & Cholewiak, R. (1996). Vibrotactile Threshold in Young and Old Observers: The Effects of Spatial Summation and the Presence of a Rigid Surround. *Journal of the Acoustical Society of America, 99*(4), 2256-2269.

[7] Hatzfeld, C., & Kern, T. (Eds.). (2009). *Engineering Haptic Devices: A Beginner's Guide* (2nd ed.). Springer

[8] History of Virtual Reality. (2016). Retrieved from http://www.vrs.org.uk/virtual-reality/history.html

[9] Hollerbach, J. M. (2000). Some current issues in haptics research. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on* (Vol. 1, pp. 757-762). IEEE.

[10] Kossyk, I., Dorr, J., Raschendorfer, L., & Kondak, K. (2011). Usability of a virtual reality system based on a wearable haptic interface. *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems,* 3474-3479.

[11] Laidlaw, R., & Hamilton, M. (1937). On the vibrational sensitivity in different regions of the body surface. *Acta Physiologica Scandinavica, 31*, 285-289.

[12] Levin, M., Magdalon, E., Michaelsen, S., & Quevedo, A. (2015). Quality of Grasping and the Role of Haptics in a 3D Immersive Virtual Reality Environment in Individuals with Stroke. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 23*(6), 1047-1055.

[13] Long, J. (1989). *Cognitive ergonomics and human-computer interaction* (Vol. 1). Cambridge University Press.

[14] MacLean, K. (2008). Haptic Interaction Design for Everyday Interfaces. *Reviews of Human Factors and Ergonomics, 4*(1), 149-194. Retrieved from http://web.stanford.edu/class/me327/readings/1-MacLean08-RHFE-Design.pdf.

[15] Mine, M., Brooks Jr, F., & Sequin, C. (1997). Moving objects in space: exploiting proprioception in virtual-environment interaction. *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 19-26. ACM Press/Addison-Wesley Publishing Co.

[16] Myles, K., & Binseel, M. (2007). The Tactile Modality: A Review of Tactile Sensitivity and Human Tactile Interfaces. *Army Research Laboratory*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a468389.pdf.

[17] Punpongsanon, P., Iwai, D., & Sato, K. (2015). SoftAR: Visually Manipulating Haptic Softness Perception in Spatial Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics, 21*(11), 1279-1288.

[18] Retrieved from http://www.geomagic.com/files/6114/3940/9416/Haptic_Device_brochure-8-2015- final.pdf.

[19] Retrieved from http://www.quintic.com/education/case_studies/Anatomical_Movements.html.

[20] Ruffaldi, E., Barsotti, M., Leonardis, D., Bassani, G., Frisoli, A., & Bergamasco, M. (2014). Evaluating Virtual Embodiment with the ALEx Exoskeleton. *Haptics: Neuroscience, Devices, Modeling, and Applications, 1*(1), 133-140.

[21] Sato, M. (2002). SPIDAR and virtual reality. *2002 Proceedings of the 5th Biannual World, 13*, 17-23.

[22] Srinivasan, M. A. (1995). What is haptics?. *Laboratory for Human and Machine Haptics: The Touch Lab, Massachusetts Institute of Technology*, 1-11.

[23] Stuart, M., Turman, A., Shaw, J., Walsh, N., & Nguyen, V. (2003). Effects of Aging on Vibration Detection Thresholds at Various Body Regions. *BMC Geriatrics, 3*(1). Retrieved from http://www.biomedcentral.com/1471-2318/3/1.

[24] Weinstein, S. (1968). Intensive and Extensive Aspects of Tactile Sensitivity as a Function of Body Part, Sex, and Laterality. In *The Skin Senses* (pp. 195-222). Springfield, IL: Charles C. Thomas.

# APPENDICES

## Appendix A. Informed Consent Form

### INFORMED CONSENT TO PARTICIPATE IN A STUDY TO TEST THE EFFECTS OF PHYSICAL AND VIRTUAL ENVIRONMENTS ON TRAINING FOR AN UNFAMILIAR TASK

A research project on human factors and virtual reality is being conducted by Delaney Bales, a student in the Department of Industrial Engineering at Cal Poly, San Luis Obispo, under the supervision of Dr. Reza Pouraghabagher. The purpose of the study is to observe and analyze the effects of physical and virtual environments on training for an unfamiliar task.

You are being asked to take part in this study by performing a task in two different environments, physical and virtual, followed by a questionnaire. Your participation will take approximately 25 minutes. Please be aware that you are not required to participate in this research, you may discontinue your participation at any time without penalty, and you may omit any items you prefer not to answer in the questionnaire at the end.

By participating in this study, there is a risk of experiencing headaches, dizziness, or nausea as a result of virtual reality, called "virtual reality sickness." All possible risks associated with participation in this study include head discomfort, confusion, disorientation, headaches, dizziness, nausea, overheating, spreading germs, strained muscles, hitting something, static shock, sharp objects, overexertion, and glasses damage. Wearing glasses while wearing a virtual reality head-mounted display is not recommended, and you assume all risk if your glasses are damaged during the experiment. Please note that these risks have been minimized and are unlikely to occur, however individual response to virtual reality may vary. Do not wear the virtual reality head-mounted display longer than 20 minutes. If you should experience physical or psychological distress, please be aware that you may contact Health Services at (805) 756-1211 or Counseling Services at (805) 756-2511 for assistance.

Your confidentiality will be protected by the assignment of a subject number in place of your name. Your name will not be used in data collection or in reports of this research. Your subject number will be used to identify your individual results in data collection and in reports of this research. Your responses to the questionnaire will also remain confidential to protect your privacy. Although there are no direct benefits to you by participating in this study, your participation may result in a greater understanding of virtual reality in designing industrial training methods for unfamiliar tasks.

If you have questions regarding this study or would like to be informed of the results when the study is completed, please feel free to contact researcher Delaney Bales at dmbales@calpoly.edu or 916-616-9305, or Dr. Reza Pouraghabagher at rpouragh@calpoly.edu. If you have concerns regarding the manner in which the study is conducted, you may contact Dr. Michael Black, Chair of the Cal Poly Institutional Review Board, at (805) 756-2894, mblack@calpoly.edu, or Dr. Dean Wendt, Dean of Research, at (805) 756-1508, dwendt@calpoly.edu.

If you agree to voluntarily participate in this research project as described, please indicate your agreement by signing below. Please keep one copy of this form for your reference, and thank you for your participation in this research.

| | |
|---|---|
| ______________________________ | ______________ |
| Signature of Volunteer | Date |
| ______________________________ | ______________ |
| Signature of Researcher | Date |

## Appendix B. Experiment Instructions

### Participant Screening

1) Are you right-handed?
2) How many times have you used Virtual Reality before? (Must be never or only once)
3) How is your vision? (Must be normal or corrected-to-normal)

### Greeting & Introduction

Hello, I'm Delaney and I am an Industrial Engineering grad student. Thank you for participating today. The purpose of my thesis experiment is to investigate the effects of virtual reality on training for an unfamiliar but simple task. The experiment is broken into two parts--Reality and Virtual Reality--followed by a questionnaire about your experience. In both parts, your task is to trace a straight line from left to right while holding a remote. (Show physical equipment.) Your goal is to complete the task as quickly and accurately as possible. You may stop the experiment at any time if you do not want to continue. Do you have any questions?

### Instructions for Physical Environment

For this part of the experiment, you will be holding this pointer to trace a straight line from left to right. There are two metal plates on here that will detect when you touch it. Behind it is a piece of glass that you're supposed to trace on. The goal is to not touch the metal and trace the line as accurately as possible. When you pick up the pointer, hold it like a remote control with your thumb on top. Extend your arm in a comfortable position until you can reach the board, but without locking your elbow. Please make sure you are sitting up straight.

We will start with 3 practice runs. Make sure you start on the left edge. Keep the pace by counting to three seconds… 1 thousand, 2 thousand, 3 thousand, and you should be done tracing the whole line. Go ahead and start your practice runs. (Do 3 practice runs.) Are you done with your practice runs? We will now start recording. (Clear the timer)

Remember the goal is to not touch the metal. I will count to 3 and say Go. It is important that as soon as you finish, you must say Stop and I will stop the timer. Any questions? Ready? 1, 2, 3, Go. (Record the error and total time. Clear the timers. Repeat three times.) We are now done with this part of the experiment. Please give me a moment to setup the next experiment.

### Instructions for Virtual Environment

For this part of the experiment, you will be wearing a VR headset and using a controller to trace a straight line from left to right. Before we begin, there are several things you should know.

1) You will only be wearing the headset for about 5 minutes or less.
2) The headset will completely obstruct your vision, but you will see where the controller is.
3) The headset is secured to your head, but you will feel about a pound of weight as it rests on your face.
4) I wipe the cushion after every use, but you may also put a tissue around the cushion if you'd like.
5) (Glasses Optional) Please keep your glasses on at first. You can take them off if it is uncomfortable or if your vision is not affected.

First, I need you to measure your eye distance to customize the fit. You can change the distance with this knob on the bottom left of the headset after you put it on. When you pick up the controller, hold it below the trigger and do not pull the trigger or press any buttons. If you accidentally press something, it won't do anything except make your arm shake and affect the accuracy of the line. Extend your arm in a comfortable position until you can reach the board, but without locking your elbow. Please make sure you are sitting up straight. Now, it's time to put on the headset. (Help them adjust the headset and then check.) How does everything feel? Can you see? You can use this knob to adjust the vision.

Now pick up the controller. Do you see the red line where the arrow is pointing to? Try to touch It with your pointer. You will be using the pointer to trace a straight line from left to right on the colored surface. You want the tip of the pointer on the surface and inside the grey guides. There are 3 colors to help you perceive depth and accuracy.

- Green is good, it means you are on the surface and within the accurate range.
- Yellow is medium, it means you are on the surface, but not within the accurate range.
- And Red means you are not touching the surface at all, meaning the depth of the pointer is too far away or too far into the board.

The goal is to stay on the surface and keep it Green as much as possible and trace the line as accurately as possible. If you feel uncomfortable at any time, please tell me and I will stop the experiment.

We will start with 3 practice runs. Make sure you start on the left edge. Keep the pace by counting to three seconds… 1 thousand, 2 thousand, 3 thousand and you should be done. Go ahead and start your practice runs. (Do 3 practice runs.) Are you done with your practice runs? We will now start recording.

Remember the goal is to stay on the surface and keep the line green. I will count to 3 and say Go. Any questions? Ready? 1, 2, 3, Go. (Backspace to clear timer before each run. Spacebar to enter results after each run. Repeat three times.) We are now done with this part of the experiment. You can take the headset off now. Please give me a moment to setup the next experiment.

**Conclusion**

We are now done with the experiment. I have a questionnaire for you to fill it out. If you would like, I will notify you about my results and thesis defense. Thank you again for participating! Please take it to the front cubicle and leave it in the folder. You are free to leave when you're done. Thanks!

# Appendix C. Post-Experiment Questionnaire

## General

1. Subject Number:
2. Email (if interested in the results/defense):
3. Age:
4. Class Level: Freshman Sophomore Junior Senior Graduate
   N/A
5. Major:
6. Gender (Circle One): Male Female
7. Was the task of drawing a straight line with a remote unfamiliar to you?
   Yes No
8. Which environment would you recommend for training somebody else in the same task? Physical Virtual
9. Which environment did you prefer overall? Physical Virtual

## Physical Environment

10. On a scale from 1 to 5 (with 1=difficult and 5=easy), how easy was it to complete the task in the physical environment?
    1 2 3 4 5
    Why?
11. On a scale from 1 to 5 (with 1=uncomfortable and 5=very comfortable), how comfortable was the equipment while performing the task in the physical environment?
    1 2 3 4 5
    Why?
12. On a scale from 1 to 5 (with 1=least strain and 5=most strain), how much muscle strain did you feel while performing the task in the physical environment?
    1 2 3 4 5
    If yes, please describe what you felt and where (hand, arm, face, etc.):
13. On a scale from 1 to 5 (with 1=least reliable and 5=most reliable), how reliable was the equipment in completing this task in the physical experiment? (Did you perceive that the equipment worked the way it was intended to?)
    1 2 3 4 5
    Why?

14. On a scale from 1 to 5 (with 1=least effective and 5=most effective), how effective is the physical environment for training people for this task?
    1 2 3 4 5
    Why?

15. What did you like about the physical experiment?

16. What did you dislike? What are your suggestions for improvement?

**Virtual Reality Environment**

17. On a scale from 1 to 5 (with 1=difficult and 5=easy), how easy was it to complete the task in the virtual environment?
    1 2 3 4 5
    Why?
    Do you have suggestions to improve ease of use?

18. On a scale from 1 to 5 (with 1=uncomfortable and 5=very comfortable), how comfortable was the equipment while performing the task in the virtual environment?
    1 2 3 4 5
    Why?

    Do you have suggestions to improve comfort?

19. On a scale from 1 to 5 (with 1=least strain and 5=most strain), how much muscle strain did you feel while performing the task in the virtual environment?
    1 2 3 4 5
    If yes, please describe what you felt and where (hand, arm, face, etc.):

20. On a scale from 1 to 5 (with 1=least reliable and 5=most reliable), how reliable was the equipment in completing this task in the virtual experiment? (Did you perceive that the equipment worked the way it was intended to?)
    1 2 3 4 5
    Why?

21. On a scale from 1 to 5 (with 1=least effective and 5=most effective), how effective is the virtual environment for training people for this task?
    1 2 3 4 5
    Why?

22. What did you like about the virtual reality experiment?

23. What did you dislike? What are your suggestions for improvement?

24. Did you notice any delay between your motion and the visual feedback in virtual reality? Did it affect your ability to complete the task? How so?

25. Did you experience headache, disorientation, confusion, dizziness, or any other physical discomfort?

## Appendix D. Statistics and ANOVAs

Table 7. ANOVA for Data: Average Error Ratio. Effects: Gap, Gender, Environment, Block(Gap, Gender). Interaction between Environment and Gap is significant.

```
Analysis of Variance

Source                    DF   Adj SS   Adj MS  F-Value  P-Value
  Gap                      3  0.53005  0.17668    17.77    0.000
  Gender                   1  0.00108  0.00108     0.11    0.743
  Environment              1  3.23299  3.23299   325.19    0.000
  Gap*Gender               3  0.00692  0.00231     0.23    0.873
  Environment*Gap          3  0.36600  0.12200    12.27    0.000
  Environment*Gender       1  0.01605  0.01605     1.61    0.211
  Environment*Gap*Gender   3  0.00770  0.00257     0.26    0.855
  Block(Gap, Gender)      40  0.34773  0.00869     0.87    0.663
Error                     40  0.39767  0.00994
Total                     95  4.90620

Model Summary
        S    R-sq  R-sq(adj)  R-sq(pred)
0.0997085  91.89%     80.75%      53.31%
```

Table 8. ANOVA for Data: Average Error Ratio. Effects: Gap, Gender, Environment, Direction, Block(Gap, Gender, Direction). Direction does not have a significant effect.

```
Analysis of Variance

Source                         DF   Adj SS   Adj MS  F-Value  P-Value
  Gap                           3  0.53005  0.17668    18.30    0.000
  Gender                        1  0.00108  0.00108     0.11    0.740
  Envmt                         1  3.23299  3.23299   334.93    0.000
  Directn                       1  0.00136  0.00136     0.14    0.710
  Gap*Gender                    3  0.00692  0.00231     0.24    0.868
  Envmt*Gap                     3  0.36600  0.12200    12.64    0.000
  Gap*Directn                   3  0.05338  0.01779     1.84    0.159
  Gender*Directn                1  0.00051  0.00051     0.05    0.820
  Gap*Gender*Directn            3  0.03151  0.01050     1.09    0.368
  Block(Gap, Gender, Directn)  32  0.26097  0.00816     0.84    0.682
  Envmt*Gender                  1  0.01605  0.01605     1.66    0.207
  Envmt*Directn                 1  0.00533  0.00533     0.55    0.463
  Envmt*Gap*Gender              3  0.00770  0.00257     0.27    0.849
  Envmt*Gap*Directn             3  0.07105  0.02368     2.45    0.081
  Envmt*Gender*Directn          1  0.00000  0.00000     0.00    0.985
  Envmt*Gap*Gender*Directn      3  0.01240  0.00413     0.43    0.734
Error                          32  0.30889  0.00965
Total                          95  4.90620

Model Summary
        S    R-sq  R-sq(adj)  R-sq(pred)
0.0982489  93.70%     81.31%      43.34%
```

Table 9. ANOVA for Data: Change in Average Error Ratio. Effects: Gender, Gap, Direction of Randomized Test Sequence.

```
Analysis of Variance

Source                  DF   Adj SS    Adj MS  F-Value  P-Value
  Gender                 1  0.03209  0.032091     1.66    0.207
  Gap                    3  0.73201  0.244002    12.64    0.000
  Direction              1  0.01065  0.010654     0.55    0.463
  Gap*Gender             3  0.01540  0.005133     0.27    0.849
  Gap*Direction          3  0.14210  0.047366     2.45    0.081
  Gender*Direction       1  0.00001  0.000007     0.00    0.985
  Gap*Gender*Direction   3  0.02480  0.008267     0.43    0.734
Error                   32  0.61778  0.019306
Total                   47  1.57484

Model Summary
       S    R-sq  R-sq(adj)  R-sq(pred)
0.138945  60.77%     42.38%      11.74%
```

Table 10. ANOVA for Data: Change in Average Error Ratio. Effects: Overall Environmental Preference, Gender, Gap.

```
Analysis of Variance

Source                    DF   Adj SS    Adj MS  F-Value  P-Value
  OverallPref              1  0.09058  0.090578     5.39    0.027
  Gender                   1  0.02826  0.028261     1.68    0.204
  Gap                      3  0.70148  0.233825    13.92    0.000
  OverallPref*Gender       1  0.00909  0.009087     0.54    0.468
  OverallPref*Gap          3  0.06150  0.020500     1.22    0.319
  Gender*Gap               3  0.02207  0.007356     0.44    0.727
  OverallPref*Gender*Gap   3  0.05131  0.017102     1.02    0.398
Error                     31  0.52060  0.016793
Total                     46  1.52637

Model Summary
       S    R-sq  R-sq(adj)  R-sq(pred)
0.129590  65.89%     49.39%           *
```

Table 11. ANOVA for Data: Change in Average Error Ratio. Effects: Environmental Preference for Training, Gender, Gap.

```
Analysis of Variance

Source              DF   Adj SS    Adj MS  F-Value  P-Value
  TrainPref          1  0.00072  0.000718     0.03    0.855
  Gender             1  0.06960  0.069603     3.27    0.079
  Gap                3  0.58754  0.195848     9.21    0.000
  TrainPref*Gender   1  0.03543  0.035435     1.67    0.205
  TrainPref*Gap      3  0.02744  0.009147     0.43    0.733
  Gender*Gap         3  0.02551  0.008504     0.40    0.754
Error               35  0.74439  0.021268
  Lack-of-Fit        2  0.11315  0.056573     2.96    0.066
  Pure Error        33  0.63124  0.019129
Total               47  1.57484

Model Summary
       S    R-sq  R-sq(adj)  R-sq(pred)
0.145836  52.73%     36.53%           *
```