Contents

Chapter 1

Introduction

1.1 Overview

1.1.1 Motivation

The design of a cockpit is a complex and lengthy process. The engineering team must constantly manage a variety of complex design constraints, requirements, human factors, mission context, and technical limitations. To manage the complexity of the project, the design is reviewed throughout the process to ensure requirements are being met. One important part of these reviews is an evaluation from the future users of the cockpit – i.e. the pilots or astronauts. Feedback from pilots can validate the current design, highlight potential issues, or even discover major design flaws.

The tools used for the pilot evaluations evolve as the design matures, becoming more and more realistic to the eventual flight mission. Initially, evaluations may be done from paper studies of the design, progressing to engineering mockups, then flight simulators, and finally flight tests. These are the typical tools used throughout the design of a cockpit, though the specifics of each can depend on the project and resources available. For example, a flight simulator can range in fidelity from a desktop simulator to a motion-base simulator. Just as with the simulator, the mockup can range in fidelity from as simple as a foamboard structure with printouts of cockpit instruments to a machined, full scale construction of the size and layout of the craft in design. The distinguishing feature between a mockup and a simulator is that a mockup is non-functional. The mockup is typically built early in the design process, before a full simulator can be developed. Often, the main purpose of a mockup is for ergonomic studies and it is not as effective of a tool for pilot evaluations as a simulator.

The simulator can provide a mission context to the evaluation pilots, and with a functioning simulation they can determine the workload of their flight tasks and identify any potential issues. However, this comes at a cost of significant time and money to develop the flight simulator. Additionally, by the time the cockpit design is ready for the development of a simulator it can be no longer feasible to undertake large design changes. If the pilots are unsatisfied with the design at this stage, the deficiencies risk being ignored under pressure of schedule and cost. Instead, the remaining deficiencies of the cockpit are expected to be overcome through additional crew training. This represents a cost in training throughout the lifetime of the vehicle, typically measured in decades, and "unknown" deficiencies of the human-machine system are often discovered too late, and are described in accident reports.

As opposed to the simulator, mockups are more likely to be developed during the early stages of the design process where large scale changes are more frequent and exact sizing and placement of components may not be complete. Mockups are built to provide designers a physical representation of the cockpit or flight deck in design instead of committing to building a fully functioning simulator. The mockup as a pilot evaluation tool is limited, and the reason for this is the same reason it is adaptable to large changes: the components are non-functional. Mockups are built without functioning instruments or displays so that changes can be made more rapidly.

In this work, we describe a tool named the "Rapidly Reconfigurable Research Cockpit" (R3C) created to bridge the gap between mockup and simulator. To achieve this, we have designed and built a prototype of a virtual environment that adds the immersion and interactivity benefits of a more functional simulator to a non-functional geometric mockup. Evaluation pilots, while seated in the mockup, wear an immersive head-mounted display that provides a high fidelity and dynamic virtual view of the cockpit panel, interior, and exterior window views. The head-mounted display is tracked and gaze-registered, enabling the cockpit visuals to be overlaid on the low fidelity physical mockup. The motion and gestures of the cockpit evaluator are tracked using an unobtrusive hand tracker to read the users' inputs on the non-functioning cockpit instruments. This provides the essential interactivity to turn the mockup from a non-functioning device into a functioning simulator. With the R3C system, an evaluator can wear a head-mounted display to provide an immersive virtual-visual layer registered with the physical world in the mockup and still retain the ability to touch and feel the cockpit. As opposed to a full simulator, this approach does not require a time and cost expensive hardware integration; instead it can be defined in software. Similarly, as opposed to using a purely virtual reality simulator without any physical representation, the combined mockup with the virtual environment provides a more realistic simulation since it becomes coupled with the tactile sense. This virtual environment combines the roles of mockup and simulator, bringing the benefits of a full simulator to the benefits of early stage design mockups: faster and more drastic redesign iterations.

Our work does not aim to supersede any existing process or design philosophy, but rather provides a tool that human systems integration engineers can use to enhance their process. Specifically, we aim to improve the iterative design process of cockpit design by reducing cost and time for the redesign cycles by bringing a more functional tool to the early stages of design. The importance of the iterative process has been highlighted in many design approaches. For example, quoting from NASA (2015):

> "Human-centered design [...] [i]s characterized by early and frequent user involvement, performance assessment, and an iterative design-test-redesign process."

Another goal of the R3C system is to not only increase the number of iterations, but to improve the feedback from the evaluations. The use of a mockup for design evaluation is standard in the industry. Zea et al. (2012) reported on design of a cockpit architecture for the then in-development DreamChaser spacecraft. Three mockup versions were constructed, and they found that "a significant amount of knowledge acquired from the mockups was not attainable through computer models". A physical mockup that also acts as a functioning simulator will help to improve the feedback of the cockpit evaluations. Quoting from Sexton (1988):

> "[The test pilots] must consider the design in a mission context, as if they are using the systems while flying – the more realistic the testing environment, the more meaningful the findings."

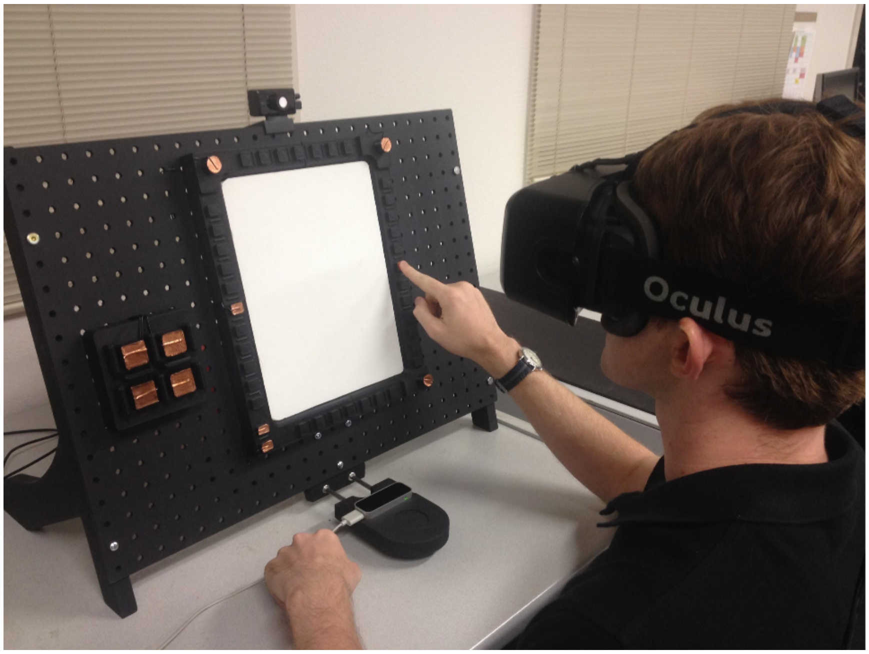
By allowing the cockpit evaluators to use the mockup as a functioning simulator, the feedback will be more effective with the context of the mission being provided with the virtual layer. The R3C system will allow for more iterative cycles with more meaningful feedback from pilots early in the design process. This combination will lead to a more optimized cockpit design. Eliminating the design deficiencies early will not only lead to cost savings in the design process, but the eventual design will reach a more optimized state. This will reduce crew training and errors throughout the life of the vehicle. The cost savings and safety benefits of a more optimized design can be enormous, both during the design stage and beyond.

The overall objective is to provide an unprecedented ability to optimize a new spacecraft cockpit layout via rapid, low-cost design iterations by integrating state-of-the-art technologies in virtual reality, hand tracking and 3D printing. The work presented in this dissertation reports on the design and integration of the R3C prototype, experimental results both to understand how humans perform with the prototype and to evaluate the performance of the prototype as a design tool.

1.1.2 Technical Approach

The prototype system of the “Rapidly Reconfigurable Research Cockpit” (R3C) that has been developed consists of a virtual reality (VR) headset, an optical hand tracking device, and 3D printed cockpit surfaces (Figure 1.1). The VR head-mounted display (HMD) provides a virtual-visual view of the cockpit environment. A challenging aspect of immersive VR simulations is to provide a sense of touch. We solve this by using 3D printed instruments co-located with the virtual environment, a technique commonly called “passive haptics”. The instruments remain inert and non-functioning in the real world, but appear fully functional in the virtual world presented to the user. They can be modeled to the exact shape of a cockpit instrument being designed, and printed within a matter of hours. Instead of requiring the mockup instruments to have functioning buttons, the pilots’ touch of a button is read using the hand tracker to detect and display the interaction with the cockpit. The optical based hand tracker measures the real-time position of the pilots’ hand to determine if they are touching a button of the cockpit. The hand tracker also displays their hand in the virtual world for the user, as the use of an HMD blocks the view of one’s own limbs. Additionally, the use of the hand tracker allows for a more thorough examination of the hand movement of evaluation pilots in a new cockpit design than is typically possible.

The combination of these technologies allows an existing mockup to be used as a functioning simulator with visuals overlaid on top of screens and/or out-the-window views, and with functioning buttons on the instru-



(a) Photograph of setup.

(b) Virtual view of user.

Figure 1.1: The Rapidly Reconfigurable Research Cockpit prototype.

ment. With merged mockup and simulator functions, adjustment of the developing cockpit can be made simpler, cheaper, and more quickly turned around. The result of employing this virtually-enhanced mockup should be a human interface that is truly optimized, which can result in increased safety and efficiency, as well as reduced crew workload/fatigue.

1.1.3 Experimental Approach

The R3C prototype has been the technical base for the research performed in this work. The three main tracks of the research can be summarized in these aims:

1. Create a virtual reality system with passive haptics and hand tracking that allows for evaluation of early design cockpits (Chapter 2).

2. Validate the technical approach and evaluate the performance of subjects in the virtual reality environment with passive haptics (Chapters 3 and 4).

3. Evaluate the use of the prototype in a mock design evaluation study to determine differences in standard performance metrics and subject feedback from using the R3C system (Chapter 5)

The technical approach and development of the prototype is described in Chapter 2. Three experiments with human subjects have been performed and are reported on to support these aims in Chapters 3, 4, 5. The first experiment (Chapter 3) asked subjects to repeatedly target one of four buttons on the panel under different conditions, testing differences between real versus virtual world as well as passive haptics versus no passive haptics (mid-air reaching). We recorded time to completion and accuracy to the button center. The second experiment (Chapter 4) had subjects perform a Fitts' Law task in different conditions to further characterize the difference between reaching movements when using passive haptics versus without the passive haptics. Additionally, this experiment had a more extensive subject survey to collect the opinions of the subjects, especially on the use of passive haptics. The final experiment (Chapter 5) performed the same design evaluation study using the R3C system and a traditional simulator. This was done to compare the results to determine where, if any, differences were obtained from the evaluation results. Subjects were split into two groups, one using the R3C and the other a touchscreen simulator. Two instrument designs were presented to each subject and they were asked to perform the same task on both. Afterwards, their comments about the two designs were collected. Quantitative performance measures of the task were also recorded and analyzed. The goal of the experiment was to determine where subject feedback and performance may differ between an evaluation study using R3C and one using a touchscreen. The differences found were described and discussed.

1.2 Background

To frame this research, this section will summarize relevant previous work as well as background on the dependent measures used in the experimental work. The motivation for our virtual environment is to be used for the design of a cockpit — a safety-critical human-machine interface. The most important performance measures in the human-cockpit interaction are the time and accuracy of the pilot, as well as their workload. A major focus of the experimental work is on characterizing the changes in these measures due to the use of the R3C system.

The first few sections summarize different tracks of virtual reality research: it's use in design and prototyping, different haptic environments, and the evaluation of virtual environments. Next, input device evaluation research such as Fitts' Law is discussed, highlighting work that has been performed in virtual environments. Finally, a summary of various human reaching trajectory research is given to understand how the human formulates trajectories in a virtual environment.

1.2.1 Virtual Environments for Design

The use of virtual environments in design processes is well established. Many design and engineering teams adopted virtual reality early in order to enhance their prototyping process. Brooks Jr (1999) reviewed the (at the time) current use of virtual reality applications in industry. A number of engineering design firms had already seen the potential for using virtual reality set ups in design reviews, though display applications were typically wall-projections instead of a head-mounted display. John Deere claimed $80K of savings from preliminary mockup costs from using a VR system and Chrysler performed ergonomic reviews using a head-mounted system, though the HMDs at the time were very heavy and low resolution (Brooks Jr, 1999). Recent advancements in virtual reality technology has renewed interest in the benefits of virtual environments for prototyping, and its use has been increasing across industries - manufacturing (Choi et al., 2015), assembly (Pontonnier et al., 2014), automotive (Bordegoni and Caruso, 2012; Lawson et al., 2016), and medicine (Nagendran et al., 2013).

Aromaa and Väänänen (2016) reviewed the use of virtual prototypes for human factors/ergonomics (HFE) evaluation. They found that the fidelity of a virtual prototype may not affect subjective results, but quantitative task performance can be affected, hindering the utility of the evaluation itself. This is supported by findings that decreased presence (the feeling that the simulation is real) can negatively affect task performance (Youngblut and Huie, 2003). Lawson et al. (2016) provided recommendations for incorporating virtual prototypes in the automotive industry for HFE evaluation, including using haptic feedback and marker-less body tracking. They found that the use of markers can encumber the participant, affecting evaluation results. Our technical approach uses a markerless hand tracker which will help evaluation results be as realistic as a non-virtual simulator. Abate et al. (2009) investigated the use of haptic interactions for virtual reality training of assembly and maintenance task, and performed a usability assessment which found technicians preferred the use of haptics in a virtual maintenance task.

Flight simulation was one of the first applications of virtual environments, and the use of virtual environments for training and development of new cockpits has grown throughout the years (Hancock et al., 2008). Virtual reality simulators for aviation have recently become a more active track of research. Wan et al. (2011) used an augmented reality approach to register a virtual-visual layer on top of a mockup of a commercial aircraft cockpit for design and evaluation but did not describe any interaction methods. Yavrucuk et al. (2011) combined an immersive head-mounted display with hand tracking gloves to provide a virtual simulator of a helicopter.

They did not use any haptic feedback and users interacted with the instruments using the hand tracking, with no physical mockup. Aslandere et al. (2015) had a similar technical setup to our system with a head-mounted display and a hand tracker for inputs on cockpit instruments, except using a hand tracking system requiring the user to wear special gloves. Additionally, their virtual flight simulator prototype did not have a physical mockup or any tactile inputs. Their report focused on the virtual button interaction using a hand tracker, . They found that the collision volume for a virtual button interaction did not impact the interaction, but a more realistic hand avatar increased efficiency. The first appearance of work similar to ours was by Schiefele et al. (1998) in Germany, which also used a head-mounted display (HMD) to replace the visuals of a flight simulator (though technology has drastically changed in HMDs since their work). They recreated only the positions of the panels with a flat plate and then used a magnetic hand tracker for reading the pilot input. A short usability study was presented, and they found that users could activate switches, knobs and dials in less time with the physical panel than without (i.e. reaching into mid-air). Their tactile feedback was limited to the flat panel instead of having geometrically accurate instruments, which we provide in our system using 3D printing.

Although many past research studies and industrial processes have used virtual environments in the design process, to our knowledge, a technical setup with a similar approach to ours has not been attempted. In this work, we aim to create a unique prototype of a new virtual environment for design. As well as the quantative measures of the performance of our system we investigate, we also aim to test our virtual environment as a design tool (Chapter 5).

1.2.2 Virtual Tactile Environments

A unique aspect of our virtual environment is the use of the mockup to provide a strong tactile sense to the user. In this section, other methods for providing the tactile sense is investigated, as well as research that uses similar methods to ours.

One of the challenges of virtual environments has been presenting the user with a tactile sensory input that matches the visual sensory inputs they are experiencing. Many haptic glove products have been created that attempt to simulate this with motor vibrations. Since they are attached to your body, they cannot provide any external stopping force, so the simulation can break down when interacting with rigid objects. Another technique is using a stylus mounted to a 6-DOF actuator that can simulate external forces. These have been used for many studies with the application of surgical tasks, when a tool would be used in the real scenario as well.

The focus of this work is the idea of *passive haptics*, whereby the tactile environment is recreated to provide the sense of touch, including external forces. The disadvantage is that the tactile environment cannot dynamically change during the simulation. Some researchers have attempted to create dynamic passive haptics (McNeely, 1993; Tachi et al., 1994), which use a robotic arm to deliver the appropriate tactile experience to the user as they reach into the virtual world. This approach quickly becomes complicated and expensive. Others have tried approaches to morph the virtual world to adapt to a generic passive haptic device (Kohli, 2009; Kohli et al., 2012), which can work only for small discrepancies. Fortunately, the adaptability during a simulation of the passive haptic environment is not a requirement for the use case of a pilot evaluating a new cockpit interface. We still retain the ability to quickly change the environment in between simulation sessions via 3D printing new parts.

Previous work with passive haptics, though limited, have shown it to provide benefits. Insko (2001) investigated the impact of passive haptics on generic virtual environments. It was found that the passive haptics increased the sense of presence felt by users. Presence is discussed in a later section. An experiment had two groups complete a maze in a virtual environment, one group with passive haptics on boundaries, and one without. After training, the passive haptic group bumped into far fewer obstacles than the group without.

Borst and Volz (2005) created a “mixed” haptic virtual environment, whereby they combined an active haptic glove with a passive haptic panel. The panel provided the stopping force and depth location of the interface, while the glove haptics provided the finer details of the buttons, switches and dials. They compared subjects performing tasks on the panel using the glove by itself (no panel), the passive haptics by itself (no glove), as well as their mixed condition (panel and glove). They found the glove by itself consistently performed worse than the other two, but overall no significant difference existed between passive haptics and their mixed haptic condition.

An important reason for the use of haptic feedback is that co-located haptic feedback has been shown to increase performance of tasks in a virtual environment (Swapp et al., 2006). Bouguila et al. (2000) found that haptic feedback improved the perception of depth in a stereoscopic virtual environment. The ability to perceive depth would be an important factor for the use of a virtual environment for human factors evaluation.

Passive haptics have seen limited use in virtual environment research, possibly due to the limitations of the simulation with a static haptic environment. Studies that have utilized passive haptics have found that they benefit task performance and presence. There has not been a study which fully characterizes the difference between no haptics and passive haptics in an immersive virtual reality environment. Since this is a major component of our approach, we aim to study the difference with the use of performance measures described in the next few sections.

1.2.3 Virtual Environment Evaluation

Since we are building a unique virtual environment, it is important to evaluate the effectiveness of our prototype. In addition to the quantitative metrics we discuss later, in this section we investigate the various qualitative measures used to evaluate virtual environments.

Presence

It is notoriously hard to quantify the effectiveness of a virtual environment (VE). One can report on quantitative measures of the virtual reality (VR) setup such as latency, field of view, resolution and others, but in the end the only important measure is whether the user finds the virtual environment convincing. This can also change depending on the user – one user could find a VE convincing while another finds it intolerable. The effectiveness of a VE is often linked to the sense of *presence*, defined by Witmer and Singer (1998) as "the subjective experience of being in one place or environment, even when one is physically situated in another." Sheridan (1992) argued that presence is a subjective sensation that necessitates a subjective measure. In general, it is important to measure presence as it has been shown to correlate with performance in the virtual environment (Youngblut and Huie, 2003). A strong sense of presence increases the feeling that the objects shown in the virtual environment are real and the interactions are the same as they would be in the real world. Presence also provides the user with a better memory of the environment (Dinh et al., 1999).

The most common method for measuring presence is through the use of subject questionnaires. Two of the more popular post-questionnaires that have been developed are the Presence Questionnaire (PQ) (Witmer and Singer, 1998) and the Slater-Usoh-Steed (SUS) questionnaire (Slater et al., 1994). The former (PQ) consists of 32 questions that cover various factors of presence such as sensory, control, distraction and realism. The SUS questionnaire consists of only 6 questions and is focused on the feeling of presence. Nystad and Sebok (2004) found the PQ to be more sensitive to technology and interaction methods, while the SUS fits better to the definition of presence. For this reason, we chose to work with the PQ as it should be more sensitive to changes in technology between conditions. In the second experiment (Chapter 4) we use the PQ to compare between the use of no haptics and passive haptics. The possibility of increased presence with passive haptics found by Insko (2001) previously mentioned is an important finding we aim to replicate for our virtual environment.

Arm Fatigue

We also hypothesize that our system will reduce the amount of arm fatigue experienced by the operator compared to a mid-air pointing virtual environment (i.e. no tactile workstation present). This common fatigue phenomenon in virtual environments is often called "gorilla arms", as people feel that their arms are too heavy to support after prolonged use reaching out into the void. A measure of this arm fatigue is not included in the standard presence questionnaire. The literature also does not contain a large volume of experiments that measure arm fatigue in virtual environments. Qualitative measures include a Likert scale, the ISO-9241-9 "Assessment of Comfort and Fatigue" and the Borg scales (RPE and CR10) (Borg, 1998). The Borg scale performs favorably to the other scales for measures of exertion (Grant et al., 1999). Recently, a quantitative approach has been proposed that uses optical tracking of the arm and wrist which validated well against a Borg scale (Hincapié-Ramos et al., 2014). Adapting their software to our system would be out of scope (and requires more upper-limb measurement than we are currently capable of) but the validation against the Borg scale is promising. We use the Borg scale to measure the arm fatigue in the second experiment (Chapter 4).

1.2.4 Workload

Workload is an important measure for cockpit design as a higher workload can lead to degraded performance or increased errors. During the design of a cockpit, it is important to measure the workload of a pilot to ensure that they are given managable tasks to perform. In our work, we investigate workload to understand how the use of the R3C system might affect the measurement of workload. The third experiment (Chapter 5) compares the workload of subjects using the R3C system to subjects doing the same task in a more traditional simulator.

Workload is a term that describes the *amount of work* an individual has to do to accomplish a task. It can refer to the actual amount of work, but typically in the context of aviation and spaceflight it refers to the pilot's perception of the workload. Hart and Staveland (1988) describes workload as a "construct that represents the cost incurred by a human operator to achieve a particular level of performance". The subject perceived workload is measured in our work using the NASA Task Load Index (NASA-TLX) survey (Hart and Staveland, 1988). The TLX survey asks for a rating of the perceived workload on a number of subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. The NASA-TLX survey has been widely adopted as a standard for measuring workload (Hart, 2006).

1.2.5 Input Device Evaluation

Our use of the hand tracker to record inputs of the pilot on the cockpit panel is not an entirely new concept, but given the differences in hand tracking technology and virtual reality devices, it presents a significant difference to past research to warrant an investigation into how well the input method works. For decades one of the most important empirical models in the Human-Computer Interface (HCI) community has been Fitts' Law. The basis of the model is a method to predict the movement time of a user towards a target. The average movement time can be accurately predicted using just the size and distance to the target after empirically determining the model parameters. It has been used to evaluate new input devices to determine how fast a user can move to a target (e.g. mice, trackballs, joysticks), as well as a guideline for interface designs (ensuring that target areas are the right size). A summary of the field is presented here, with notable contributions and similar work following. Fitts' Law used in the context of virtual environments and differences with our work is also discussed.

Fitts' Law

Fitts (1954) originally reported on experiments of targeted reaching movements between varying target sizes, and compared the discovered correlation to information theory models. He postulated that the entire control system of the human, from motor control to visual and proprioceptive feedback, can process information at a certain maximum rate, and the bandwidth of the channel must be derivable from the physical parameters of the targeting task. Fitts adapted Shannons' Theorem (Shannon, 1949) from information theory for the bandwidth of a channel in the presence of a signal and noise. The distance to the target (D) is considered the signal, and the width of the target (W) is considered to be the noise (Figure 1.2). This analogy allowed Fitts' to create a formula for the *index of difficulty* (ID) of a target:

$$\mathrm{ID} = \log_2\left(\frac{2D}{W}\right) \tag{1.1}$$

The units of index of difficulty are *bits*.

It was found experimentally that the index of difficulty is linearly correlated to *movement time* (MT), the time taken to move between the start and the target. This leads to the equation that is now known as Fitts' Law:

$$\mathrm{MT} = a + b \cdot \mathrm{ID} \tag{1.2}$$

where a and b are the parameters of the linear regression, which are determined experimentally. A typical experiment consists of subjects performing movements to a range of targets with varying indices of difficulty, and

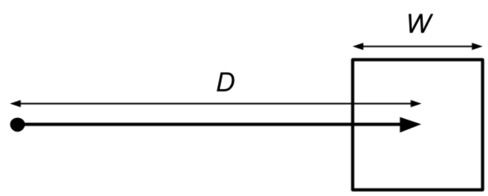


Figure 1.2: The dimensions of a Fitts' Task (D = distance, W = width)

recording their movement time. The data is then fit to a linear regression to find a and b.

Since its original formulation, refinements to the model have focused on the formula for index of difficulty, ID, and the definition of its dependencies, width and distance. Crossman (1957) suggested a very important refinement called the adjustment for accuracy. This provides a correction to the ID to compensate for the task which is actually performed by the subject as opposed to the task that was asked to be performed. For example, if a subject is presented with two large targets close together ($D << W$), then they tend to aim towards the inner edges of the targets instead of the center, effectively using a much smaller target width. The adjustment for accuracy replaces the width, W, with the *effective width*, W_e. The basis for this adjustment is that the information theory basis for Fitts' Law expects that the "noise" of the target width induces a 4% error rate (MacKenzie, 1992). If more or less than 96% of the target hits are successfully within the target, then the target width is adjusted with the adjustment for accuracy to fit the target error rate of 4%. Motor tasks have been shown to result in a Gaussian distribution of target accuracy (Crossman and Goodeve, 1983; Woodworth, 1899), thus 96% of target endpoints will be within 4.1 standard deviations (σ). The effective width is thus calculated with the standard deviation from the endpoint data:

$$W_e = 4.1\sigma \tag{1.3}$$

It is recommended to calculate the effective width for each subject in each condition as it is sensitive to both between and within subject variability (Soukoreff and MacKenzie, 2004). The effective width can then be used to calculate an *effective index of difficulty*, ID_e. The effective width described here is for one dimensional targets, but there are extensions to 2D targets as well (Murata and Iwase, 2001).

The form of the logarithmic term in the index of difficulty equation (Eq. 1.1) has encountered some debate over the years. Fitts did not describe the reasoning for including the factor of two in his original formulation, and Welford (1968) suggested a modification that provided a better fit for low values of index of difficulty:

$$\text{ID} = \left(\frac{D}{W} + 0.5\right). \tag{1.4}$$

A more recent variation proposed by MacKenzie (1989) suggests that by comparing to Shannons' Theorem, as Fitts originally proposed as the theoretical basis for his work, then the form should be,

$$\mathrm{ID} = \left(\frac{D+W}{W}\right) = \left(\frac{D}{W} + 1\right) \tag{1.5}$$

which has been widely adopted in literature as the Shannon's Formulation. As well as being based in the original theoretical underpinnings of Fitts' Law, the mathematical structure of this formulation will not produce a negative index of difficulty with real, positive dimensions, unlike the original equation.

The first use of Fitts' Law in the human-computer interface (HCI) community was with Card et al. (1978), comparing the use of a mouse, joystick and keys for text selection. This led to an extensive use of Fitts' Law in the HCI field, both as a way to evaluate input devices and a design law for user interfaces. Best practices of using Fitts' Law have been developed, and the standard ISO-9241-9:2000 provided guidelines for the use of Fitts' Law in evaluation of input devices (International Organization for Standardization, 2000). Most notably, it was suggested to use the ISO-9241 "Fitts' circle" for evaluations of 2D tasks (Figure 1.3). This input pattern has been widely adopted in literature, and provides a standard for comparing across studies.

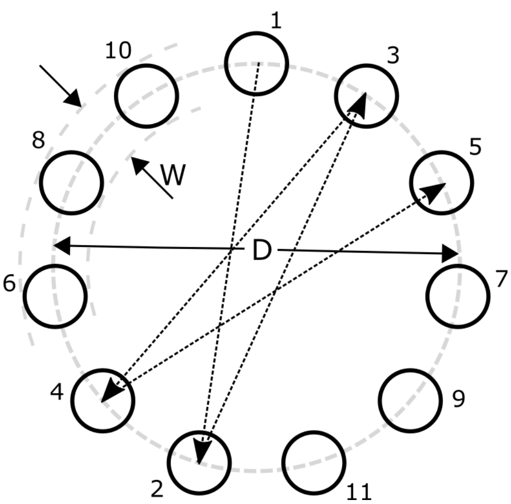


Figure 1.3: ISO-9241 Fitts' Circle or multi-directional tapping task.

Soukoreff and MacKenzie (2004) present a review of the best practices for designing and performing a Fitts' law experiment, including the use of the Shannon Formulation and the ISO-9241 "Fitts' circle". They also advise the use a large range of ID values, ideally from 2 bits to 8 bits.

When Fitts' Law is used to evaluate input devices, the metric of comparison is often the *throughput*. Throughput (TP) is defined as the index of difficulty (ID) over the movement time (MT).

$$\mathrm{TP} = \frac{\mathrm{ID}}{\mathrm{MT}} \tag{1.6}$$

The units of throughput are bits per second, and it represents the information capacity of the input device. A higher throughput indicates that the human is able to convey more information (*bits*) through one input device versus another in a given time (*per second*). Soukoreff and MacKenzie (2004) recommend the use of throughput as the measure for comparing experimental conditions as well. One of the advantages of using throughput is that it combines speed and accuracy requirements through the use of the index of difficulty. Throughput is often cited as independent of speed and accuracy tradeoffs due to the use of index of difficulty. MacKenzie and Isokoski (2008) performed a brief validation study investigating this hypothesis, and found that subjects recorded the same throughput whether they had been instructed to prioritize speed or prioritize accuracy.

The debate on which form and refinement of Fitts' Law is the best to use continues to this day (Drewes, 2010; Hoffmann, 2013), and this work will not attempt to settle this debate. Since no form of the equations have been rejected outright in the literature, it is recommended to perform analysis with all versions and note if any significant differences in major conclusions result (Soukoreff and MacKenzie, 2004).

Dimensionality of Fitts' Law

Throughout the Fitts' Law literature, one of the main differences between experiments is how many dimensions the subjects can move their targeting device (hand, stylus, cursor). There are two dimensionalities to consider when discussing Fitts' Law experiments: the dimensionality of the task, and the dimensionality of the movement. The dimensionality of the task refers to the number of dimensions the targets are placed in. A 1D task means the targets are aligned along a single axis, whereas a 2D task has them located on a plane (such as the Fitts' circle in Figure 1.3). The dimensionality of the movement refers to how many dimensions the subjects need to move when moving between the targets. For example, the original Fitts' Law task was a 1D task as the targets were varied in width along the movement direction, but sufficiently long in the perpendicular dimension, removing that dimension as a variable. The movement, however, was effectively 2D, as the subjects could (and needed to) lift up from the table to move between targets.

The popularity of Fitts' Law in the HCI field has led many researchers to focus on the use of Fitts' Law on a 2D computer screen controlled by a mouse. This is a 2D task, with a constrained 2D movement. For our application, the movement we are interested in is a 2D task with 3D movement since the buttons on a cockpit panel are configured in approximately a single plane, with pilots reaching towards them in unconstrained 3D movement. There are some newer studies that present this task and movement dimensionality, as it is analogous to the use of a touchscreen device. Some have proposed minor variations to the formulation of Fitts' law to adjust for touchscreens (Bi et al., 2013; Sears and Shneiderman, 1991), but others have found the original formulation to perform well (MacKenzie, 2015). There are also studies discussed in the next section that perform this task in a virtual environment.

Refinements have been proposed to account for the possible effects of extending the task dimensionality to 2D, such as the bivariate target model (Accot and Zhai, 2003; MacKenzie and Buxton, 1992; Wobbrock et al., 2008). Some work has also explored this for 3D tasks, introducing a trivariate model (Grossman and Balakrishnan, 2004) and corrections for 3D placement of targets (Cha and Myung, 2013; Murata and Iwase, 2001). These experiments place targets in a truly 3D space, with of course 3D motion required of the user. The use of a Fitts' circle on the 2D plane effectively removes the need for these accommodations, as the targets are symmetric and have the same width when approached from any direction.

Fitts' Law in Virtual Environments

Fitts' Law has been used in evaluating virtual environments and their input devices. Chun et al. (2004) evaluated a set of 3D stereo displays with a single haptic-enabled stylus using a Fitts' tapping task. They did not have well-fit regressions, but this could have been due to a small range of ID values used (2-3), targets being placed in 3D space (a 3D task with 3D movement), with random order (making the sequence of the task unlearnable), and averaging across subjects before completing the regression. Teather et al. (2010) performed a similar task and also found a poor fit of the regression in the fully 3D task. One condition performed a mid-air 2D planar task with 3D movement, and no significant difference in throughput was found compared to the same task with constrained 2D motion. Liu et al. (2009) performed a planar multi-directional Fitts' task with a stereoscopic display, and compared virtual world to real world results, finding movement time twice as long in the virtual condition. The trajectories were analyzed to determine where the extra movement time was spent, which is discussed in the next section (Human Targeting Motion). Teather and Stuerzlinger (2011) performed an ISO-9241 Fitts' circle task in a CAVE virtual environment (stereoscopic projection on walls) with a motion tracked stylus, comparing performance with and without a passive haptic on the plane of the targets. They found increased throughput with the passive haptics, but throughput performance was lower than they had expected.

One of the few experiments that utilize Fitts' Law in an immersive HMD was performed by Kohli et al. (2012) who focused on evaluating "discrepant" (warped virtual space) virtual environments to fool the user of the nature of the physical world. They had the subjects perform an ISO-9241 Fitts' circle on a panel placed in front of them, and in some conditions the panel was rotated in the virtual world but not in the real world or vice-versa to create the discrepancy. The subjects' hand movement was warped to compensate for this. Their non-discrepant condition (when virtual and real were at the same angle) should compare to our results with passive haptics.

Coelho and Verbeek (2014) evaluated the LeapMotion hand tracker (the same technology used in our work) in a 3D Fitts' task (with visual feedback on a computer screen), but did not perform a Fitts' model analysis, comparing only time. Seixas et al. (2015) performed an ISO 9241-9 Fitts' circle task with the LeapMotion. Their goal was to compare a gesture based selection technique, and they provide a Fitts' throughput comparison between that and direct mid-air selection. To our knowledge, ours is the first work to perform a Fitts' Law evaluation with a passive haptic, immersive VR, and the LeapMotion hand tracking technology.

1.2.6 Human Targeting Motion

The use of the hand tracker also opens up another quantitative analysis – the hand tracker provides the dataset of the complete trajectory of motion of the user. In the past, in order to obtain this data an expensive and/or intrusive hand tracker would be required. With the unique combination of technology we use in our system, the complete hand motion of the subject can be recorded without any extra equipment or considerations from the subject. In this section, we look at previous research experiments that have used trajectory information to analyze subjects moving in a targeting motion, like the movement of pilots' targeting buttons in a cockpit.

In one of the earliest experiments to measure trajectories of human targeting motions, Woodworth (1899) discovered that there are two distinct phases of the targeting motion: a ballistic phase and a corrective phase. Subjects typically perform the bulk of their movement in the ballistic phase, often considered more of an open-loop or feed-forward mechanism, and as they near the target switch to a feedback, closed-loop, corrective phase (Elliott et al., 1999).

Experiments in virtual environments have shown that users typically spend more time in the corrective movement phase than when performing the same task in a non-virtual environment (Liu et al., 2009). However, this work was completed for a mid-air interaction, so it is an open question whether the introduction of passive haptics would reduce the corrective phase 'penalty' for targeting motions in a virtual environment. MacKenzie et al. (1987) showed that there was correlation between the target size and the time of the deceleration phase for a Fitts' task, suggesting that subjects will adapt their trajectory for accuracy requirements. Our application to a cockpit environment is one where accuracy demands are often more important than time demands for manipulation of instruments.

Other metrics have been developed in the literature to evaluate the trajectory of a targeting motion. Meyer et al. (1988) developed a parsing criterion to divide the trajectory into submovements and defined three types: return from overshoot, accelerate due to undershoot, and slowing towards target. MacKenzie et al. (2001) presented a set of accuracy measures for targeting motions based on trajectory information. They were developed to compliment a Fitts' input device evaluation. The measures are named target re-entry, task axis crossing, movement direction change, orthogonal direction change, movement variability, movement error, and movement offset. They quantify types of movement that are not along the ideal straight-line path from start to goal.

In our experimental work, we focus on the subdivision of trajectories into the ballistic and corrective phase. We expect the corrective phase to be significantly affected by the use of passive haptics. Past experiments have shown that virtual environments can increase the amount of time subjects spend in the corrective phase, but with the use of the passive haptics the subjects have a backstop to aid with locating the virtual button. We investigate the differences of trajectory phases due to the effect of passive haptics in the second experiment (Chapter 4).

1.2.7 Summary

The combination of immersive virtual reality simulator with an inert passive haptic has not been extensively reported on in the virtual reality literature. Other approaches at immersive haptics have focused on creating a dynamic environment (changing in real time), but our application of a static cockpit does not need to change geometrically during the simulation (not to be confused with our re-configurability between simulations). Unobtrusive hand tracking (requiring no gloves on the user) like the one we are using has been highlighted as important for human factors evaluations. The integration of these new hand tracking technologies is a further innovation, allowing us to provide interactivity with the passive haptics.

To evaluate this new virtual environment, we use two validated measures – presence and Fitts' Law. Fitts' Law is a well validated and actively used model for evaluating both input devices and interfaces. To date there have been only a few uses of Fitts' Law in immersive head-mounted display virtual environments. As of our knowledge, no one has performed a Fitts' Law experiment in a virtual environment using the new generation optical hand trackers (i.e. the LeapMotion we are using). A Fitts' Law characterization is an important step to take to evaluate the system we have created, and provides a standard means of comparison across previous and future literature. Additionally, the trajectories of the targeting motions can be used to further analyze the motion. Past research has found that the use of a virtual environment affects the trajectory but no work has investigated the effect of passive haptics.

The use of virtual environments in design and prototyping is well established, but we present a unique approach. Some research has started to investigate the use of various forms of augmented and virtual reality for design tasks, but our combination of technology has not been tested in the context of a design evaluation tool.

1.3 Research Questions

The research presented in this work consists of three experiments with human subjects to answer our fundamental research questions. As is natural with research work, our research questions and motivations evolved with the findings of the previous experiment. Presented here are a summary of the major research questions that were investigated with the experimental work. The questions are further detailed in their respective experiment chapters.

1. Can the technical approach of a mockup providing passive haptics with a virtual-visual overlay using a head-mounted display be used successfully?

2. Can a user select a button in the R3C prototype with the passive haptics and hand tracker as quickly and accurately as one in the real world?

3. Does the use of passive haptics change the performance of subjects targeting a button in the virtual environment?

4. Does the use of passive haptics increase the presence of subjects using the virtual environment?

5. Does the use of passive haptics change the trajectory formation of subjects targeting a button in the virtual environment?

6. Can the R3C prototype be used effectively as a design evaluation tool?

7. How does the R3C prototype compare to other design evaluation tools?

1.4 Summary

We have introduced our goals and motivation for the creation of the Rapidly Reconfigurable Research Cockpit prototype, a system for providing a higher fidelity evaluation of new cockpit designs while still retaining a low iteration cost. The background for the experimental work has been described and the open research questions that we explore in this work have been summarized. In the following chapters, we first describe the technical work to create the Rapidly Reconfigurable Research Cockpit (Chapter 2), before reporting on the three experiments and their results (Chapters 3-5).

Chapter 2

Prototype Design

2.1 Overview

Throughout the course of this dissertation work, the Rapidly Reconfigurable Research Cockpit (R3C) prototype underwent many changes and improvements to the technology used. Our underlying motivation and goals of the technical approach of the prototype remained constant, but as technology evolved and our experience and feedback grew, a number of improvements and revisions were developed. In this chapter, three major versions of the prototype are outlined, and the evolution of the technical approach is described. The first prototype was not used in a formal experiment, but two versions based on the second prototype were used for the first two research studies (Chapters 3 and 4). The final prototype was used in the third experiment (Chapter 5).

2.2 Requirements

The following high level requirements for the Rapidly Reconfigurable Research Cockpit (R3C) prototype originated from the motivation described in Section 1.1.1. These requirements guided the development of the prototypes, based on the technology available.

Rapid prototyping of physical mockup. Based on the motivation of enabling more rapid redesign iterations, the R3C system should allow for rapid changes in the physical representation of the cockpit (the mockup) without needing to undertake large software or hardware reconfigurations.

Immersive head-mounted display. The need for an immersive HMD with a large field of view is based on the need to use the system as a design evaluation tool. Without the ability to place visuals on the screens or create out-the-window views, design evaluations may not be as useful. Current augmented reality or mixed reality goggles do not provide the same level of immersion that a fully-virtual HMD provides.

Unobtrusive hand tracking. For a realistic design evaluation, it is important that the hand tracker does not modify the users behavior or inhibit their tactile sense. Many hand trackers use a glove or other marker worn on the hand to provide the tracking ability. Wearing a glove to achieve hand tracking would remove one of the benefits of the R3C system – providing a realistic tactile sensory input. Newer optical-based hand trackers allow this requirement to be fulfilled.

Minimal setup and calibration. Our goal is to provide a system that can be used "on top" of existing mockup systems without extensive modifications. One consequence of this means that as much as possible, the hardware used must be easy to set up and able to use in constrained spaces. This requirement eliminates some of the high-end head-mounted displays that use complicated head tracker systems requiring precise calibration. Our solution also proposes the use of hand tracking to recognize the pilots' input without the need for outfitting the mockup with working input actuators (i.e. buttons). Past approaches to hand tracking which use multiple precisely calibrated cameras, or magnetic sensors in a well-defined magnetic field, will not satisfy this requirement.

2.3 Prototype Development

2.3.1 First Prototype

The first prototype was a proof of concept system, providing a platform for the future prototypes. The major components of the first prototype were:

- Oculus Rift Development Kit (VR HMD)
- LeapMotion (hand tracker)
- 3D Printed Cockpit Instrument
- EDGE Rendering Engine with custom software for integration

This prototype was quickly outdated with new technology and software updates before it could be used for a formal research study. However, many informal feedback sessions did help highlight the usability and technical issues that led to the next prototype. The software developed during this prototype also provided a significant base for the software used throughout all the prototypes. The hardware used for this prototype was the first generation Oculus Rift Development Kit and the first generation of software for the LeapMotion hand tracker.

Rendering Engine

We are using a NASA developed rendering engine named EDGE to provide the visuals for the virtual scene rendered in the head-mounted display (HMD). EDGE is highly customizable and extendable through C/C++ plugins, Tcl scripts, or networking functions. Many of the integrations described in the prototypes were created as EDGE plugins. This rendering engine was used in all subsequent prototypes and research studies.

Virtual Reality Headset

The head-mounted display (HMD) we used for the first prototype was an Oculus Rift Development Kit (Figure 2.1). The lightweight headset provides an immersive virtual reality experience by combining a wide field-of-view scene with accurate head tracking, giving a stable virtual world. The small display (cell phone sized) is viewed through a single set of lenses. A unique image is presented to each eye, providing stereoscopic 3D vision. The orientation of the head is tracked using internal sensors (accelerometer, gyroscope and magnetometer) and exposed through the software developer's kit (SDK). The first generation Oculus Rift Development Kit has a 1280x800 LCD screen (with 640x800 for each eye) with a refresh rate of 60 Hz. The field of view is approximately 110°.

Figure 2.1: The original Oculus Rift Development Kit.

3D Printed Instruments

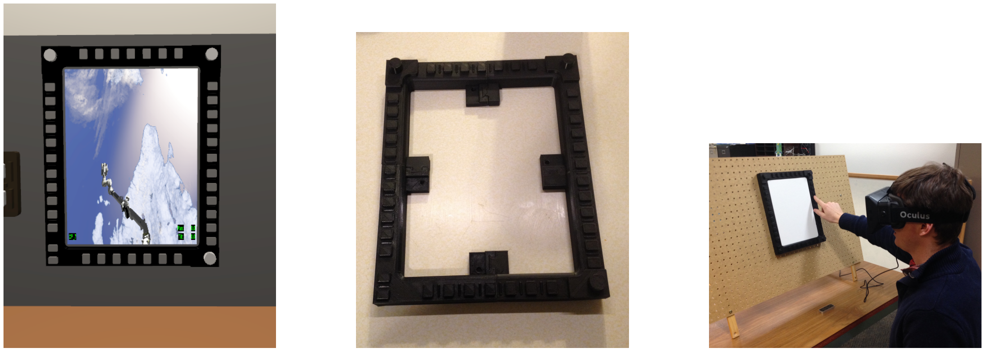
A central tenant of our technical approach is to use physical, geometrically accurate instrument shapes that provide tactile feedback to the fingertips, but no functionality (i.e. no screens, no working buttons, etc.). In order to achieve this, we have produced 3D printed "instruments" that can be easily rearranged on a panel mount (pegboard). An instrument in a cockpit can be defined as any device that provides information about or provides control of the airplane and its flight situation. In our case, the instruments we prototype consist of a device with buttons and screens in a variety of geometrical configurations and layouts. Since the devices are rapidly prototyped, they can be redesigned much more rapidly than functioning simulator instruments. By placing the geometrically accurate 3D printed instruments at the correct cockpit locations, the user is provided with accurate tactile feedback without the need for stimulating true touch with virtual haptic feedback.

To use an instrument in the R3C system, it requires two versions of the 3D model. The first is the version for the rendering engine, which requires textures and possible work to reduce the number of polygons so that rendering is not slowed down by model complexity. The second model is for the 3D printing, which depending on the 3D printer used, will often require small changes to allow for a successful print. Both of these formats require a surface mesh model. Often, the original version exists as a CAD (Computer-Aided Design) model which typically does not describe the surface mesh directly but instead describes the geometric models that define the part. The difficulty of the conversion from CAD to a surface mesh depends on the model and the requirements of the rendering engine.

For this initial prototype, a demonstration instrument was developed based on a large multi-function display with edge keys. This design was chosen based on their standard use in current aircraft and spacecraft cockpits (e.g. see US Department of Defense (1999)) and was based off the current design for the NASA Orion spacecraft cockpit (Figure 2.2). The instrument model as viewed in the rendering engine is shown in Figure 2.3a. The 3D printed version is shown pictured in Figure 2.3b. Since it is only used for the tactile feel, it has a plain aesthetic. The color and visual detail of the instrument are only on the rendered model in the virtual world,

Figure 2.2: NASA JSC simulator of Orion cockpit instrument. The instrument design was the inspiration for our demo multifunction display instrument. Image by NASA JSC.

simplifying the physical prototyping process. This instrument was also 3D printed in segments due to the limitations of the printer used, but this did not affect the virtual representation. Figure 2.3c shows the instrument mounted on a panel and being used. This instrument provided a base for the demos. Although this specific design was not used in a research study, the workflow learned to create this demo instrument was used for future instruments utilized in the research studies.



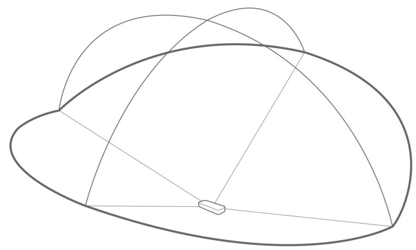
(a) The MFD rendered in the virtual world. **(b)** The 3D printed version of the MFD. **(c)** MFD mounted on a panel.

Figure 2.3: The demo multifunction display (MFD) instrument in various configurations.

Hand Tracking

The LeapMotion hand tracker was selected as it provided an unobtrusive method for hand tracking. The device itself is small (3.0 in × 1.2 in × 0.5 in), allowing it to be used in constrained environments (pictured in Figure 2.4a). It uses two infrared cameras to feed its proprietary tracking algorithm, and requires no gloves to be worn by the user. The tracking volume extends rougly 2 feet above the controller, and 2 feet about the center on the other two axes. This large, roughly hemispherical, field of view is shown in Figure 2.4b and can easily cover the working area of our research cockpit panel. A larger cockpit could be supported with multiple devices. The original version of the software, used in this prototype, provided real-time location and orientation of each fingertip. It also gave the position and orientation of the palm, but did not provide any details on the joints or which finger

(a) LeapMotion device.



(b) LeapMotion field of view.

Figure 2.4: The LeapMotion device and field of view. Images by LeapMotion. Interaction area is 2 feet above the controller, by 2 feet wide on each side (150° angle), by 2 feet deep on each side (110° angle).

of the hand the fingertip belonged to. The tracking rate is approximately 120 Hz, with an advertised latency of under 10 milliseconds.

Button Input Recognition

Since the 3D printed buttons on the instruments are non-functioning by design, the hand tracker data needs to be used to determine when a user is "pushing" a button. The 3D printed instruments do not have any movement when the buttons are pushed, so the button or key is considered activated as soon as it is touched by the users' fingertip. The users feel the tactile feedback of the button location as they are "pushing" the button, and with a working button press recognition method the button would appear to work when the user touches it.

The next step was to develop the method to determine when users were touching a button using the hand tracking data from the LeapMotion. The button recognition system was developed within the EDGE engine independent of the specific hand tracker, so that future prototypes could use a different hand tracking technology if a new device became available.

The button detection algorithm is a simple collision detection model. A rectangular box is defined in software that extends outside the button, including a tolerance zone to account for misalignment and poor tracking. When a fingertip enters and stays in the box for approximately 150 ms[1] then a button event is triggered. The purpose of the delay is to account for false positives when a user might accidentally enter the box without intention to press the button. The advantage to using the optical tracking to determine when a user has selected a button is that it has the potential to significantly reduce the complexity of the system. If the user interactions with the panel can be determined solely by tracking his/her hands from the external sensor, then the cockpit panel needs only to provide physical feedback, and does not require any wiring.

[1] This time is configurable and was often tweaked for each experiment.

Challenges

The hand tracking caused two major usability issues with the first prototype. The first was that the reliability of the tracking caused many instances of dropped tracking, causing virtual fingers as displayed to the user to disappear unexpectedly. The second was the conversion between hand tracker coordinates and instrument coordinates, a concept called registration. Registration refers to the alignment between a virtual world and the real world, a concept called registration. A well-registered virtual world will place the virutal image in a stable and precise location relative the real world. In our application, since the virtual world is colocated with the physical componenets, having accurate registration is important to the success of the prototype. This configuration of the prototype presented many challenges with achieving good registration. Even with precise alignment of the hand tracker, the fingertip positions did not always align properly with the buttons in the virtual world. Often, when a subject had their finger on a physical button their virtual finger was misaligned. Since the hand tracking was not very reliable, many users would almost completely ignore its input and find the buttons by tactile feedback. This then led to confusion when the button would not register a button press despite the strong tactile feedback. When this occured, users would not leave the physical button to hunt for the virtual button location.

Another challenging portion of the registration was that the original version of the Oculus Rift could not measure the position of the users head relative to the real world. This meant that the head position of the user in the virtual world was set manually. Depending on the actual position of the user, the panel would not appear with the correct scale and at the correct distance in the virtual world. This caused users to initially reach too short or too far for the panel, and decreased the realism as the panel seemed too small or large compared to the real world object. Since the head tracking sensors were all internal to the headset, it also experienced drift after a few minutes. Most noticeable in the yaw, it would cause the panel to become misaligned and require a manual reset to realign.

2.3.2 Second Prototype

A number of improvements were made in the second generation of the prototype. Due to vendor provided software upgrades, the hand tracking became more robust and provided more complete information about the entire hand position. The visuals were upgraded with the second Development Kit of the Oculus Rift, which provided improved visual resolution, but, more importantly, an external head tracker (using an infrared camera) which enabled accurate detection of head movement and virtually eliminated drift of the internal intertial sensors. A major focus of this version

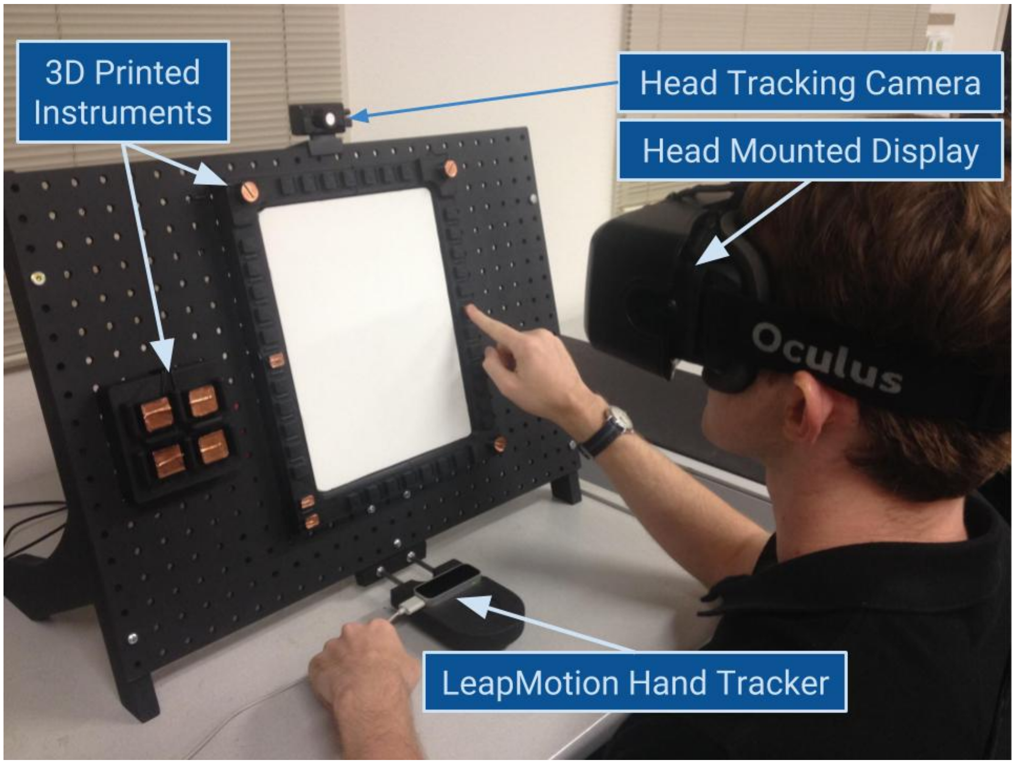


Figure 2.5: A version of the second prototype with major components noted.

of the prototype was to improve reliability and registration. The reliability was partially improved with the new hand tracking software, but other countermeasures developed are described as well. Registration (the alignment between the physical world and the virtual world) was improved with the use of a new calibration mechanism described in this section.

Virtual Reality Headset

The head tracking provided in the second Oculus Rift development kit version gave a significant improvement over the original version in the registration between the real world and the virtual world. Now with the head fully tracked with positional updates, the user can more freely explore the virtual world. With the original Oculus Rift, the location of the head was fixed in the virtual world and only orientation was provided. The head positional tracking follows the movement of the head from side to side or back and forth, providing a more immersive experience to the user. This provides the user with additional parallax depth cues to better gauge the size and distance of objects.

The position and orientation of the user's head is reported relative to the tracking camera which is placed facing the user. The camera can be mounted in a well-known location relative to the panel which provides a head position in the virtual world precisely measured relative to the panel. This means the virtual world is rendered with a more correct eyepoint of the user compared to the first version. The virtual components are now always rendered at the correct distance and with proper scale. The head-tracking camera is shown mounted on the instrument panel in Figure 2.5.

Hand Tracking

The upgrade to the LeapMotion hand tracking software provided two integrated features. The first feature was an improvement in the fidelity of the tracking, including full information on the joints of the hand. The

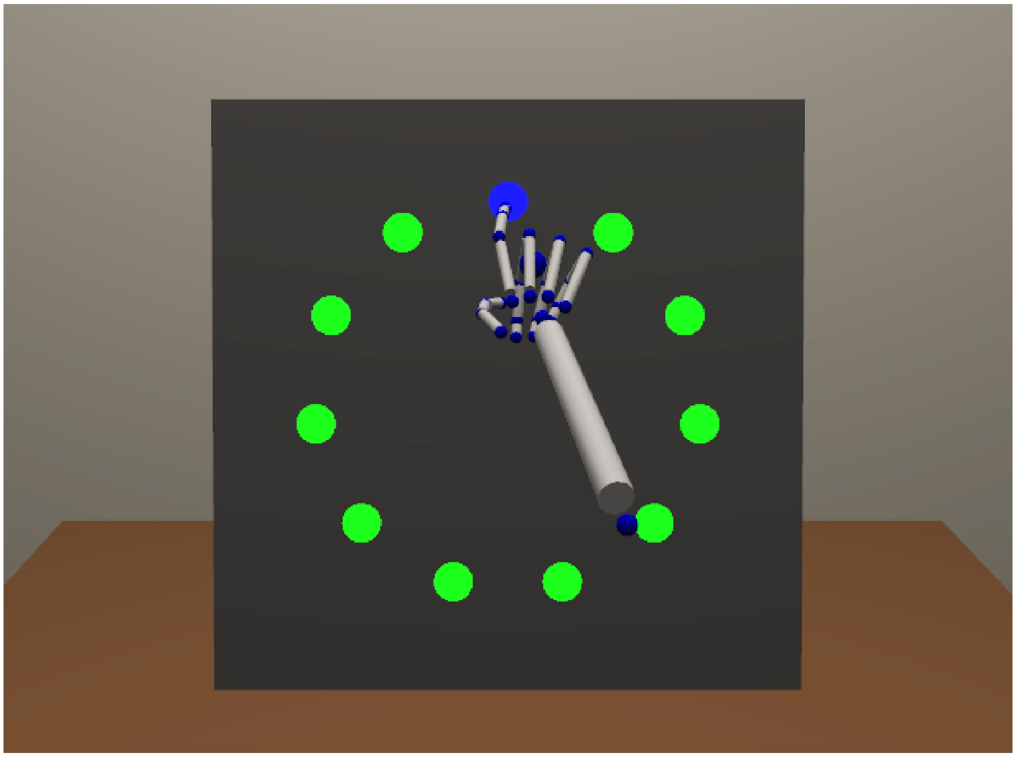


Figure 2.6: The complete hand tracking of the LeapMotion v2 software.

second was the introduction of a "head-mounted" mode, which allowed for tracking to be optimized for looking down at a hand.

The tracking upgrade added the location and orientation of the entire hand at a skeletal level. This means all the joints and the bones between them are tracked. The upgraded view is shown in Figure 2.6. This provided a more immersive feeling than the floating fingers of the original LeapMotion software used in the first prototype.

At the same time of the tracking upgrade, a "VR Mode" was introduced to the LeapMotion software to provide hand tracking in a virtual environment by way of mounting it on the head-mounted display itself. Upon initial testing of this feature, two observations were made. The tracking was improved with a perspective looking down at the hands, instead of being below looking up at the palms of the hands. This configuration seemed to provide more robust and reliable tracking data. Using it mounted to the head-mounted display, however, caused a large issue with the registration between the real and physical world. The registration between the hand tracker and the location of the instruments relies on a known, rigid connection. When the tracker was mounted on the head-mounted display, the transformation between LeapMotion coordinates and panel coordinates relied on knowing the location of the head. Although this information could be obtained from the newer model HMD which included head tracking, it was not precise or stable enough to use as the base for the hand tracking coordinates. For example, if a hand was placed resting on the panel and kept still, the virtual hand would appear to move relative to the virtual panel if the head were moved at even a moderate pace. The additional latency introduced by relying on the head tracking sensors simply would not keep accurate registration. This led us to develop a mount that would hold the LeapMotion upside down, looking down over the panel and instruments (Figure 2.7). The software could still be configured to the "VR Mode", optimizing for face-down tracking, but with a rigid mount. This

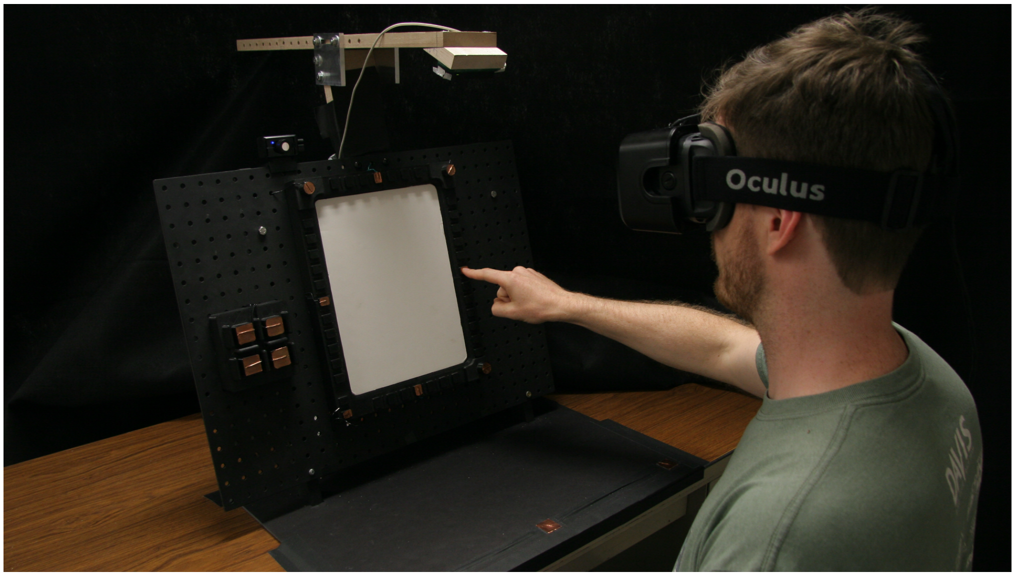


Figure 2.7: The LeapMotion mount, holding the sensor above the panel and looking down.

provided the best of both worlds: the reliability of the camera down tracking, and a stable transformation between the tracker coordinates and the panel coordinates.

Button Recognition Countermeasures

Two countermeasures were introduced in this prototype that aided new users in learning the button recognition algorithms of the hand tracker. A simple visual indication was added when a user entered the detection box in front of the button. This was typically implemented with the button itself being highlighted or changing color. With this crucial feedback, new users were able to learn the tolerances of the button zone and understand when the button was going to be activated. For experienced users, the feedback provided confirmation that they were in the zone and were not waiting for it to register incorrectly. The second countermeasure was the addition of an aural feedback upon the button press being registered. Instead of relying solely on visual feedback, the addition of a "click" noise when the button press was registered allowed the user an additional way to confirm the action of the button press had properly registered. This was especially helpful when the user became familiar enough to register buttons with their peripheral vision as they could use the audio feedback to know the button had been pressed. These two countermeasures for the difficultly of working with the button recognition were observed as being very helpful to new users, and were included in all versions of the prototype used for the research studies.

Capacitive Touch Sensors

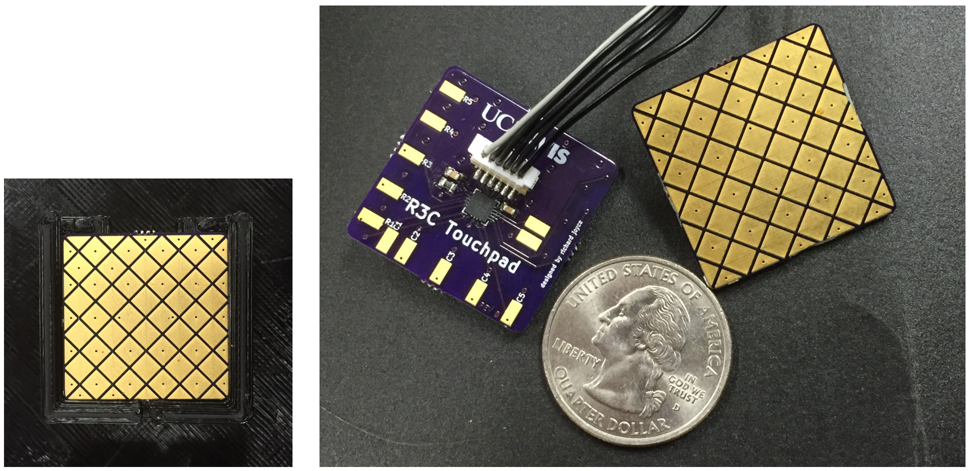
In order to be able to activate the buttons reliably when the hand tracking was degraded or dropped out, the use of capacitive touch sensors was investigated. The capacitive touch sensors were initially developed as a countermeasure to the problems encountered with the hand tracking from the original prototype. As the hand tracking became more robust with the new software from the manufacturer, however the countermeasure was not as important, but the sensors played a new role in the prototype: validating the accuracy, and providing calibration for the hand tracker.

With the requirement of minimal setup, the original capacitive touch sensors were developed with copper tape electrodes placed on the top of the 3D printed instrument buttons. These electrodes are shown on the demo multifunction instrument in Figure 2.8. A Freescale MPR121 capacitive touch sensor was used to read the electrodes and communicate with an Arduino which sent touch events over a serial communication line to the computer. These serial events were read by the rendering engine to trigger events when a copper pad was touched. These simple, single electrode per button sensors provided a reliable method for determining when the user had actually touched a button. The main use of these sensors evolved to be for the calibration mechanism described in the next section. However, as the registration improved, a research question was developed around how accurately the user could place a fingertip on the physical button while immersed visually in the virtual world. This led to the development of a second generation of capacitive touch buttons, used in the Pointing Experiment (Chapter 3).

Since touch accuracy can be important in safety-critical applications, we developed the ability to record where on the button the finger press was located using a capacitive touch sensor. To accomplish this, a custom

Figure 2.8: The copper tape electrodes for capacitive touch sensors on the demo multifunction instrument.

printed circuit board was developed that provides an electrode array of 5 rows and 5 columns over a 1-inch by 1-inch square. This board is shown in Figure 2.9. The capacitive state of each row and column can provide a measurement of the center of the finger press on the grid created by the rows and columns. The center of the finger press location can be found with an accuracy of under 0.1-inch using this configuration. The location of the finger press can help provide a measure of the accuracy of the registration between the optical sensors and the real world location, as well as any bias introduced from using the VR headset and hand tracker.



(a) Capacitive array mounted in a 3D printed button.

(b) Bare PCB showing back and front.

Figure 2.9: Capacitive touch array button. Each row and column of diamond pads are connected as one electrode each and together can provide location information of where the user presses.

Calibration

The hand tracking of the LeapMotion was observed in pilot studies to provide precise and repeatable measurement of the hand positions throughout its tracking volume. However, the accuracy was observed to be insufficient. Put another way, the position of a finger on the button in the physical world was shown offset from the button position in the virtual world, yet the offset was consistent between movements. This error also increased as the finger moved from buttons close to the sensor to buttons further from the sensor. This led us to develop a calibration to provide a more accurate registration between the virtual and physical hand positions. With the addition of the capacitive touch sensors, a calibration mechanism was developed with the use of the known positions in the physical world (the capacitive touch sensor locations). This means that when a finger was placed on a physical button with a capacitive sensor, the known position can be recorded along with the "measured" position of the hand tracker. After collecting enough points, the calibration can correct for the measured offset and provide an accurate registration between the real and virtual worlds. The mathematical basis for this calibration is described here.

The calibration works on the assumption that there exists a transformation matrix, **T**, that can solve the following equation:

$$\vec{x}_{known} = \mathbf{T}\vec{x}_{measured} \tag{2.1}$$

The $\vec{x}_{known}$ and $\vec{x}_{measured}$ correspond to the known location of a calibration point (button) and the location measured by hand tracker, respectively. Since we know those two vectors, it is only required to solve for the transformation matrix itself.

A simple least squares approach is used to find the coefficients of the matrix. The transformation matrix is not constrained to a simple rotation (i.e. not assumed orthogonality or other special properties) so the solution is found by expanding and solving the general homogeneous coordinates transformation matrix.

$$\begin{bmatrix} x_{known} \\ y_{known} \\ z_{known} \\ 1 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{measured} \\ y_{measured} \\ z_{measured} \\ 1 \end{bmatrix} \tag{2.2}$$

Typical least squares approaches would attempt to find $\vec{x}_{measured}$ in Eq. 2.1, however we desire to find the matrix $\mathbf{T}$ itself. It can be shown that expanding the matrix equation (Eq. 2.2) for the multiple calibration points (i.e. $\vec{x}_{known,1}, \ldots, \vec{x}_{known,n}$ and $\vec{x}_{measured,1}, \ldots, \vec{x}_{measured,n}$) and then collecting like terms will convert the problem into three different least squares problems. They are shown here with subscripts k and m for 'known' and 'measured'.

$$\begin{bmatrix} x_{k1} \\ x_{k2} \\ \cdots \\ x_{kn} \end{bmatrix} = \mathbf{X}_M \begin{bmatrix} T_{11} \\ T_{12} \\ T_{13} \\ T_{14} \end{bmatrix}, \quad \begin{bmatrix} y_{k1} \\ y_{k2} \\ \cdots \\ y_{kn} \end{bmatrix} = \mathbf{X}_M \begin{bmatrix} T_{21} \\ T_{22} \\ T_{23} \\ T_{24} \end{bmatrix}, \quad \begin{bmatrix} z_{k1} \\ z_{k2} \\ \cdots \\ z_{kn} \end{bmatrix} = \mathbf{X}_M \begin{bmatrix} T_{31} \\ T_{32} \\ T_{33} \\ T_{34} \end{bmatrix},$$

$$\text{where } \mathbf{X}_M = \begin{bmatrix} x_{m1} & y_{m1} & z_{m1} & 1 \\ x_{m2} & y_{m2} & z_{m2} & 1 \\ & \cdots & & \\ x_{mn} & y_{mn} & z_{mn} & 1 \end{bmatrix}$$

Now the problem is stated with a known matrix and one unknown vector (per each of the three equations). Once we have collected more than 4 points, it becomes an over-determined linear system, and a least squares calculation is used to find the solution. From the solution of the three separate equations, the original $\mathbf{T}$ matrix is reconstructed.

At least 4 points are needed to solve this system, and it has been found that a calibration with small least squares residuals can be achieved with 10-20 well chosen points. Provided no changes to the lighting or the position of components, the calibration setup can be performed once for the setup and works across users.

Instead of using the calibration matrix, $\mathbf{T}$, to transform all of the points in the hand, the index finger was used as the datum for the calibration. The offset of the index finger between the measured and calibrated positions was calculated and then the entire hand was offset by this vector. This was done as it was quickly observed that significant and unrealistic warping occurs with the virtual hand if all points are transformed with the calibration matrix. This kept the relative position of each joint in the hand as reported by the LeapMotion while calibrating against the point that was most frequently used for button targeting[2].

Lessons Learned

The capacitive touch sensors used to activate the buttons were found to be effective. However, they added additional complexity to the hardware setup of the R3C system. The calibration mechanism developed created a more accurate registration that enabled button activation without the capacitive touch sensors. The difference in using the capacitive touch sensors and the hand tracking only to activate buttons was investigated in the first research study, detailed in Chapter 3. Before the calibration working to improve registration, new users would rely on the physical feedback and wait

[2] Subjects in all research studies were instructed to use their index finger.

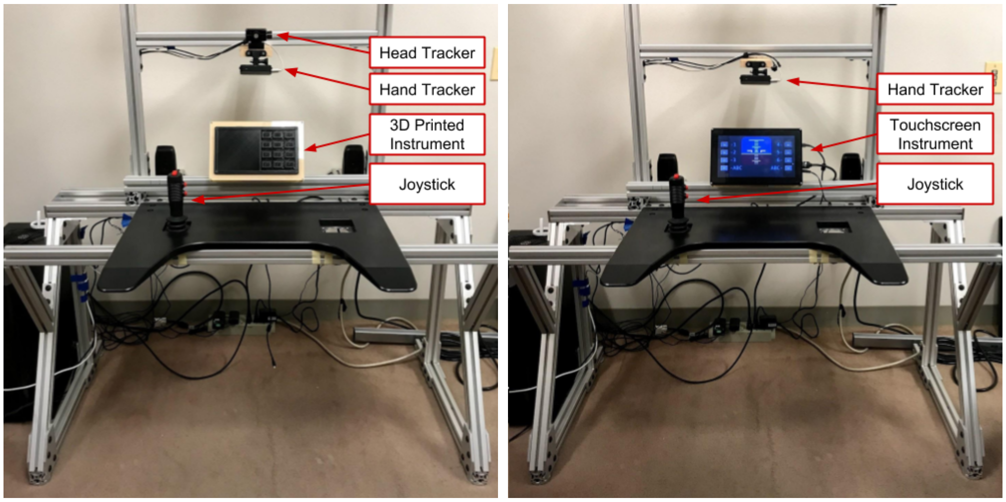
for the button registration with their finger on the physical button, despite the visuals showing them their finger misaligned in the virtual world. A more experienced user, knowledgeable about the method of button detection, would know to ignore the physical feed and move to find the detection zone by aligning the finger with the virtual button and not the physical button. Additionally, we discovered that hand pose and speed of movement can influence the performance of the hand tracker, and that more experienced users tended to learn techniques that gave the hand tracker a reliable measurement. For example, when the tracker was first acquiring the hand (from outside the field of view), having the fingers spread out led to a quicker acquisition time for the tracker. Providing the hand tracker with a good angle to view the hand at all times was something that experienced users learned as well. We often found that a small amount of training time would greatly improve performance.

Not all of the input interactions with a cockpit may be measurable with just optical tracking. Some interactions which require a fast reaction time or dynamic input (i.e. flight control manipulation) may not be suitable with the current technology. These observations are investigated throughout the research studies, and help to quantify which tasks may be appropriate for this type of prototyping system.

With this prototype it was initially noticed that the tracking performance became less reliable and would often drop out as the hand approached the panel and instruments. This was particularly frustrating for users as it was precisely where the hand tracking was needed the most as it was needed to activate the button. After isolating the problem to the presence of the panel and instruments, it was discovered that we had inadvertently provided optical interference to the hand tracker. When looking at the raw image captured by the LeapMotion infrared cameras we found that our glossy black 3D printed instruments were highly reflective in infrared, causing little to no contrast between the hand and the instruments. Applying a matte paint finish improved this, and the reflections in infrared were monitored for future prototypes and configurations. We have also found better results when covering the entire backdrop of the LeapMotion field of view with a dark matte material, as this helps provide a greater contrast between the hand and the background.

2.3.3 Third Prototype

The third prototype was developed for the design evaluation experiment. The motivation for the changes of the third prototype were guided more from research goals than a need for technical improvements of the system. These research goals and the prototype is explained in Chapter 5, a summary is given here to describe the technical changes to the prototype.



(a) R3C Setup. **(b)** Touchscreen Setup.

Figure 2.10: Simulator workstation for the third experiment.

The third experiment had subjects provide feedback on the design of two different instrument designs. For this reason, two instruments were designed, modeled, and 3D printed. To compare the feedback received from using the R3C prototype compared to a traditional simulator, the experiment had two groups of subjects. One group used the R3C prototype to evaluate the two designs, and the other used a touchscreen with no virtual reality. The touchscreen was mounted at the same location as the 3D printed instruments (they used the same mounting plate). The two setups (touchscreen and 3D printed) were interchangeable, and are pictured in Figure 2.10.

With the improvements developed from the previous prototypes, the third prototype was designed to use only the hand tracker (without any capacitive touch sensors) to perform input on the 3D printed cockpit instrument buttons. The hand tracker was mounted above the instrument and pointed towards the area in front of the instrument. For a more realistic flight simulation, the task also involved using a joystick, which was mounted to the left of subjects at desk height. To maintain realism, the joystick was modeled in the virtual environment, so that the movement of the joystick in the real world was visible to the subject in the virtual world.

Calibration

The virtual/real position registration calibration logic was modified for this setup to calibrate based off the position of the touchscreen. This provided a very accurate and easy way to calibrate the hand tracker. Instead of using the buttons on the instrument, requiring capacitive touch sensors on the buttons, the calibration was performed by switching to the touchscreen and pressing points on the screen.

We experienced some initial trouble with the calibration using a touchscreen due to the calibration points being contained in a single plane. The problem this introduces is that the calibration least squares solution becomes over-fit to the plane, and causes any movement perpendicular to be scaled down near to the plane, causing the hand to appear to be 'stuck' to the calibration plane. To solve this, an artificial point was added at 1 inch outward from the middle of the sensor datum. The same point was added as a known and measured point, forcing the calibration matrix to fit the entire tracking volume instead of just the plane. This technique came from previous observations which showed that the accuracy of the sensor was improved as the hand neared the tracker. After solving this issue, the calibration became very accurate with the touchscreen as the data source for known positions.

2.4 Summary

An overview of the technical approach of the evolving Rapidly Reconfigurable Research Cockpit was presented. The final prototype achieved many of the goals set out by the motivations and requirements. Developments that were outside the scope of this research work and could be undertaken as future work are discussed in Chapter 6. The specific configurations used in each experiment are described in detail within their respective chapter. Some of the experiments required additional technology development that was not described here as the focus was on the R3C prototype itself.

Chapter 3

Pointing Experiment

Portions of this chapter were originally published in the conference proceedings for AIAA Modeling and Simulation Technologies 2015 (Joyce and Robinson, 2015). This initial experiment aimed to provide our first quantitative look at how the use of passive haptics and virtual reality affected reaching movements.

3.1 Introduction

We have developed a proof of concept of the Rapidly Reconfigurable Research Cockpit (R3C) system, which was used to perform an initial targeting study to validate our approach, and provide a basis for future work. This study aims to validate the technical approach by having untrained human subjects attempt to perform a button targeting task in the new virtual environment. We also record their performance to understand the effect of the visuals, haptics, and hand tracking on their time and accuracy of button targeting.

The R3C prototype used in this experiment contains several components. The user wears an immersive virtual reality head-mounted display (Oculus Rift Development Kit 2), which presents a virtual scene that is spatially stabilized with respect to a physical instrument panel. The physical instrument panel is made of non-functioning 3D printed instruments. As the user reaches out toward the panel of instruments, a LeapMotion hand tracker reads the position of each finger and the pose (attitude and configuration) of the hand. A simple collision detection algorithm uses this information to determine when a user has touched a button. Since it is important for the user to visually track their own hand during pointing and actuation tasks, we also use the hand tracker to render the dynamic position of the hand in the virtual scene. To aid with the validation and performance measures, the buttons are outfitted with a capacitive touch sensor as a secondary sensor for true button touch by the subject.

In this experiment, the subjects were asked to target a set of buttons multiple times in various levels of virtual environment fidelity to understand the effects on targeting performance. Specifically, the three independent variables are virtual reality (using VR or the real world), passive haptics (use of the panel or not), and button detection (using the optical trackers or capacitive touch sensors). We recorded the time to activate the button using the detection sensor, the accuracy on the button, and the success rate of activating buttons.

Recent work investigates virtual reality tools in a simulator (Wan et al., 2011; Yavrucuk et al., 2011; Aslandere et al., 2015). However, our concept of merging an accurate but low-cost tactile environment with the high-fidelity virtual view is so far untested. Accurate haptic feedback has been a goal of virtual reality (VR) researchers since the emergence of VR. Providing dynamic haptic feedback, however, still proves challenging (Stone, 2001; Lécuyer, 2009). Our approach allows accurate haptic feedback for the case of a static workstation, such as found in a real cockpit. Combined with 3D printing and our virtual simulator overlay, this provides an inexpensive platform to create a functioning simulator, able to adapt quickly to large-scale design changes.

3.2 Experimental Methods

3.2.1 Experiment Goals and Motivation

We performed a study to validate our technical approach and measure the effects of using the R3C system on subject reaching motions. Specifically, we were interested whether an untrained user could accurately target the correct button in an instrument panel, and how the different technologies used affected the targeting task in time and accuracy. There were three main independent variables under study in this experiment:

- The effect of the use of the VR HMD versus no VR HMD (i.e. virtual vs. real world).
- The effect of the physical panel versus no physical panel present (i.e. having the tactile feedback vs. no tactile feedback)
- The effect of the optical tracking button detection versus the capacitive touch button detection

The use of the virtual reality headset also implies that subjects have to rely on the visual feedback from the hand tracker virtual hand to target the button.

We were concerned about the effect of both touch-selection accuracy and movement time, but the context for our design is an aerospace cockpit where accuracy in selecting the intended button is typically prioritized over movement time. Success rates of targeting tasks in virtual environment have not been consistently reported in previous research studies. Often, the researchers explain that incorrect trials were discarded or redone without noting the frequency of errors. We are specifically interested in the success rate so incorrect trials were recorded.

Previous work has investigated various 3D pointing tasks in the real and virtual world without tactile feedback (Liu et al., 2009), or with tactile feedback but no immersive virtual reality (Teather et al., 2010), or with virtual haptic feedback (Chun et al., 2004), or pure virtual worlds (Bruder et al., 2013; Grossman and Balakrishnan, 2004). Our work differs in the use of the tactile feedback provided by the panel combined with an immersive virtual world.

3.2.2 Experiment Setup

Figure 3.2 shows a diagram of the experimental setup. A picture of the setup and the view of the virtual world is show in Figure 3.3. The panel was configured with a single four-button keypad, arranged in a 2x2 grid. Three different starting pads were placed on the desk near the edge.

To measure the accuracy of the subject on the keypad buttons, each of the buttons were equipped with the capacitive touch sensor previously

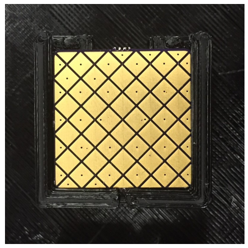


Figure 3.1: Capacitive touch array button. Each row and column of diamond pads are connected as one electrode each and together can provide location information of where the user presses.

described in Chapter 2.3.2. These sensors have an electrode array of 5 rows and columns over the 1 in × 1 in square. This provides a position accuracy of under 0.1 in when a user contacts the surface. Figure 3.1 shows a photograph of this circuit board mounted on a button. The start pads were capacitive touch copper tape electrodes, also 1 in × 1 in, but had no position detection capability.

As previously described (Chapter 2.3.1), the hand tracker mounted above the panel enables detection of when the user touches a button, ideally without need for a touch sensor. For this experiment, both sensors were utilized to compare the performance of subjects with both. The optical tracking button detection uses a simple collision detection box to determine when a subject has pressed a button. When the finger of the user enters the detection box and remains inside for 50 ms, the button press event is activated and recorded. The purpose of the small delay is to reject false positives if a user moves quickly through the detection box of a button they are not pressing. The color of the button changes in the virtual world when they enter this detection box to indicate to the user that they are about to register a button press. For this experiment, the collision detection box was set to extend 0.5 inches inward and outward from the button, including a 0.1-inch tolerance on all dimensions. Aslandere et al. (2015) found no significant effect with button selection ability based on changing this volume in a purely virtual environment, so we determined an appropriate size in pilot studies and did not include the size of the detection box as a variable.

3.2.3 Experiment Task

Audio prompts were given through headphones, which indicated the starting pad and panel button goal before each task. The participants were asked to start with their index finger on the start pad and then target the stated button on the panel. The use of audio prompts and a delay before starting removes any bias of time to process where they were targeting. The movement times also do not include the reaction time as they started when subjects lifted their finger from the start pad.

Each participant performed 48 targeting tasks under each of the four different conditions:

1. Wearing the HMD, button selection registered by capacitive sensors,

panel physically present

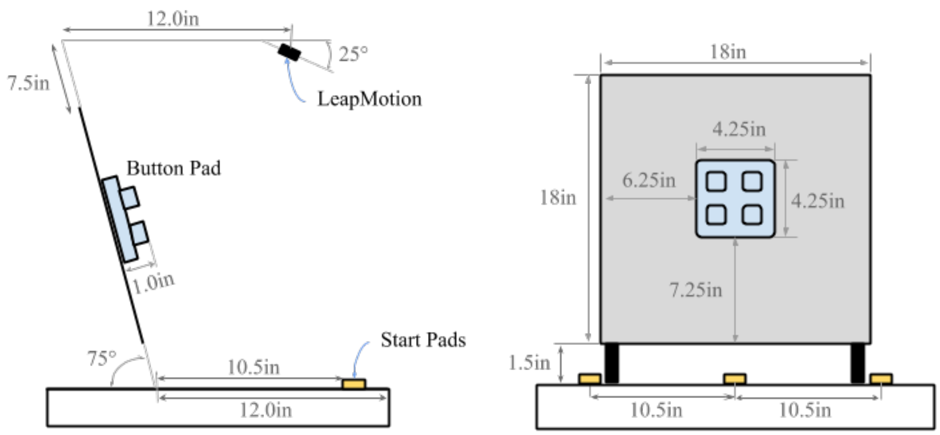
2. Wearing the HMD, button selection registered by hand tracking sensors, panel physically present

3. Wearing the HMD, button selection registered by hand tracking sensors, panel not physically present

4. Not wearing the HMD, button selection registered by capacitive sensors, panel physically present

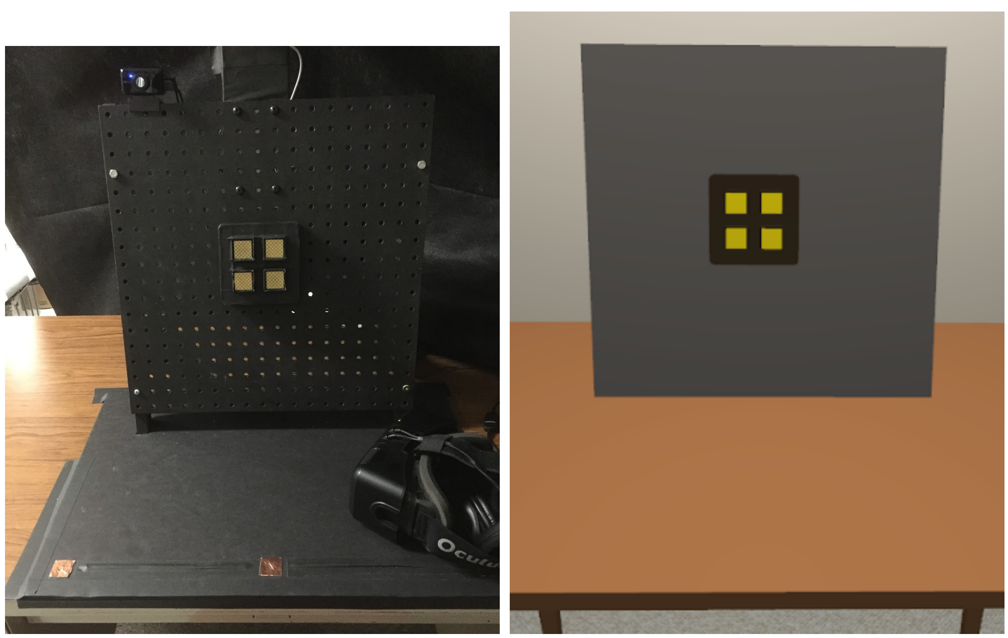
Before each condition the subjects performed one set of 12 trials for familiarization of that particular condition. This training session was not included in the results. During the “no panel” condition, the hand tracker remained mounted at the same point, but the panel was moved out of the way, such that the subjects were targeting the buttons in the virtual world only. The button selection in every condition was indicated by a sound played in the headphones when the appropriate sensor measured a button press. The tasks would time out after 10 seconds. The task was not repeated for an incorrect button selection or timing out, instead the subject moved on to the next task. The condition order was randomized for each subject, and the tasks were performed in a pseudo-random order.

The subjects were all briefed on the technology from a prewritten speech explaining how to activate each of the sensors. Subjects were instructed



(a) Side View (b) Front View

Figure 3.2: Diagram of Experiment Setup



(a) Photo of setup in real world. **(b)** View of setup in virtual environment.

Figure 3.3: Experimental setup and view of virtual environment.

to target the center of the correct button, but were given no instructions on speed. Subjects targeted the button using their index finger of their dominant hand, which was the right hand for all subjects. No information was given on how to best perform the task or advice on working with the various sensors.

The dependent variables measured were time from starting pad to trial goal button, accuracy as measured on the capacitive touch sensors (for conditions with the panel present), and success rate (selected the correct button within 10 seconds).

3.2.4 Subjects

The experiment was performed with 8 subjects, recruited from the university student population, all with no or limited previous exposure to the R3C setup. All subjects indicated they had either never used a virtual reality headset or briefly tried them once or twice. Ages ranged from 18-31 with 6 male, 2 female. They all had correctable vision to 20/20 and none indicated difficulty seeing the image in the virtual reality HMD. After the experiment, none of the subjects reported symptoms of motion sickness, and only mild eyestrain was reported. The subjects spent under one hour performing the experiment, with about thirty minutes of virtual reality use.

3.3 Results

The time was recorded for each trial from release of starting pad to the successful button press using the correct button registration. Trials which timed out or where the incorrect button was targeted were discarded for the time analysis (but not the success rate reported later). The time for each trial was corrected for the varying distances of the movements by multiplying by the average distance (18.4 in) over the straight-line distance between the start and goal button for that trial (which varied from 15.5in to 20.5in). The results for each condition are summarized in Table 3.1, and Figure 3.4 illustrates the three main effects with respect to corrected time.

The significance of the effects were determined using a two sample t-test on the corrected time measure between the appropriate conditions. The effect of the use of the HMD on corrected movement time was significant ($t(7) = -5.1, p < 0.01$) between condition 1 (HMD, Panel, Capacitive: $M = 3.15$ seconds) and 4 (No HMD, Panel, Capacitive: $M = 1.57$ seconds). On average, using the HMD caused the corrected movement time to be 1.58 seconds longer. The button detection sensor also has a significant effect on corrected time ($t(7) = 3.6, p < 0.01$), comparing conditions 1 (HMD, Panel, Capacitive: $M = 3.15$ seconds) and 2 (HMD, Panel, Optical: $M = 3.78$), but the drop in performance was only 0.63 seconds. The effect of the panel

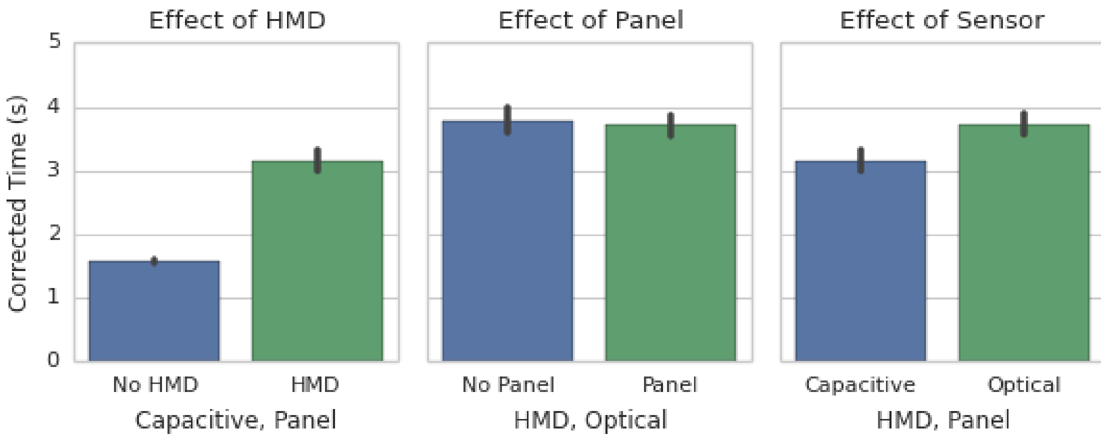


Figure 3.4: Major effects studied for corrected time of movement. Effect of HMD and Effect of Sensor are significant effects ($p < 0.01$), while Effect of Panel is insignificant.

	Condition	Mean Corrected Time (s)	Distance from Button Center (in)
1	HMD, Capacitive, Panel	3.15 ($\sigma = 1.70$)	0.23 ($\sigma = 0.09$)
2	HMD, Optical, Panel	3.78 ($\sigma = 1.90$)	0.25 ($\sigma = 0.09$)
3	HMD, Optical, No Panel	3.70 ($\sigma = 1.62$)	N/A
4	No HMD, Capacitive, Panel	1.57 ($\sigma = 0.43$)	0.15 ($\sigma = 0.08$)

Table 3.1: Mean results across subjects. Distance from center is recorded by capacitive touch sensor. Standard deviations are reported as σ.

was found to be not significant ($t(7) = -0.5, p = 0.66$) between condition 2 (HMD, Panel, Optical: $M = 3.78$) and condition 3 (HMD, No Panel, Optical: $M = 3.70$).

The selection of the correct button was performed almost without error, which is a promising result for new users of the system. There was no significant effect between the different conditions on success rate. Over 8 subjects and 1517 trials, only 12 trials selected the wrong button and only 12 trials timed out, giving an overall 98.4% success rate. Observations made during the experiment suggest that at least some of few incorrect button selections were due to misheard prompts or loss of attention and not due to the system itself.

The accuracy of the button press itself was also measured, and Figure 3.5 shows the distribution of the locations registered by the capacitive touch sensors. In the plots, (X,Y) = (0,0) corresponds to the bottom left corner of the button. The two conditions shown (1 and 4) were the two conditions where capacitive touch registered the button selection. There is a smaller distribution for the no HMD condition (Figure 3.5(b)) than with the HMD (Figure 3.5(a)). The mean location of the two conditions were 0.06 in apart, which is lower than the accuracy of the capacitive touch sensors (0.1 in). In the HMD conditions, the only visual feedback of the hand was the rendered model of the hand that used the hand tracker as the source data. The no HMD condition had a large benefit in this regard, as the subjects were operating with the normal visual feedback of their own hand in the real world – using their normal hand-eye coordination they have trained since birth. These results suggest that even with the visual feedback only provided by the hand tracker data, subjects are able to select the buttons as accurately as they could for the same task in the

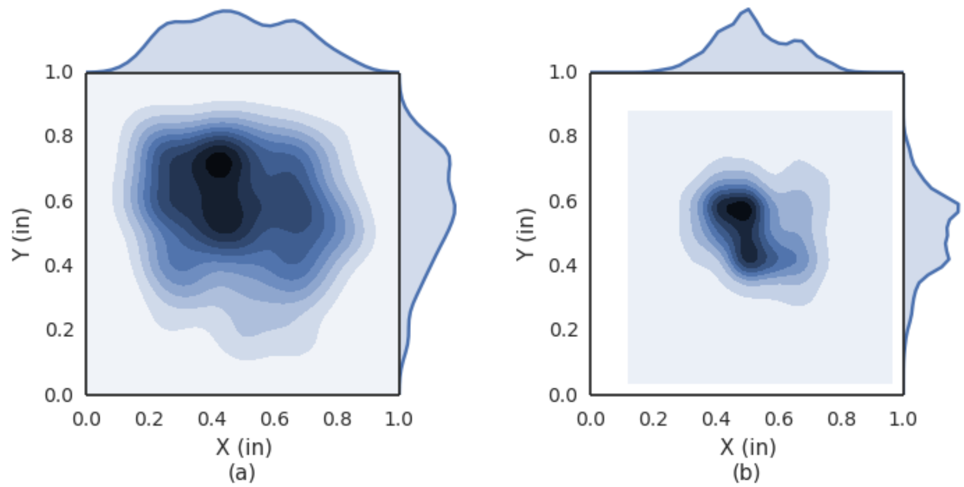


Figure 3.5: Finger press location 2D distribution for (a) HMD, Capacitive, Panel and (b) No HMD, Capacitive, Panel. The distribution is calculated as a Gaussian kernel density estimate. The mean (x, y) location and standard deviation for each is (a) $(0.48 \pm 0.18, 0.56 \pm 0.17)$ and (b) $(0.52 \pm 0.12, 0.51 \pm 0.12)$.

non-virtual world.

A Fitts' law analysis of the data was not the goal with the experiment design, and since we did not vary the button width a Fitts' analysis could only be performed with varying distance. The index of difficulty would only range from 4.0 to 4.5, providing an insufficient range for a proper Fitts' analysis. Including extra parameters for the 3D nature of the task (Murata and Iwase, 2001; Cha and Myung, 2013) did not improve the fit, especially considering the symmetry of the setup. Furthermore, the time recorded of the tasks were typically more than in a Fitts' task, since many subjects spent time honing based on the accuracy instructions.

3.4 Discussion

The start pads and hand tracker were located relative to each other in such a way that the hand would not be in view of the tracker at the beginning of the task, thus the user would have to wait for the hand tracker to acquire the hand when in view. This meant for the trials using the HMD the subject would typically have an open loop (blind) ballistic phase followed by the closed loop homing phase once the virtual hand appeared in the virtual world. Across subjects for all HMD conditions, this hand acquisition by the tracker took on average 660 ms ($\sigma = 649$ ms) from leaving the start pad. An expert user who performed the experiment was able to get the hand to appear in the tracker in 330 ms ($\sigma = 133$ ms) on average. This could indicate that training on how to get the best hand tracker performance could improve results. The same expert also only experienced a 416 ms slowdown between using the HMD and not using the HMD (conditions 1 and 4), compared to the 1580 ms slowdown measured across the 8 new subjects.

We had originally hypothesized that having the panel would improve the selection performance of the button, but for this task it had no effect. There are two possible contributing factors that reduced the effect of the panel on the performance of the subjects. First, the subjects were instructed to target the center of the button. Some subjects spent time ensuring they were centered over the button before they even entered the zone in front of the button, thus eliminating the need for the panel to provide location. Secondly, this task was a discrete pointing task, where the subject would 'reset' to the starting position on the desk between each movement. It is possible that the panel being in position would help the subject build a better sense of the relative locations of the buttons when moving between buttons on the panel (a serial task). In this experiment, subjects only targetted one button on the panel before returning to the start pads. The passive haptics may not provide as strong of an effect in this type of movement. It is also possible that while the panel did not improve targeting performance in this task, it may provide a benefit to the subjective feeling of presence in the virtual world, which may be beneficial in a more demanding situation with multiple tasks.

Another variable that was not measured was fatigue. Having the panel as a backstop may decrease the amount of fatigue for a repeated targeting task, as floating the arm in free space to target a button in a purely virtual environment with no panel can cause a buildup of arm fatigue. If the virtual world and physical world are correctly aligned, then the finger can momentarily rest on the button while the hand tracker works to detect a button press.

Overall, the lack of an effect of the panel could also be interpreted as a positive result. As discussed before, a significant problem we've been working on is the improvement of tracking performance when the hand nears the panel. The results indicate that this problem has been mitigated at least as much to provide performance on par with the purely virtual world.

3.5 Conclusion

We have shown that new users can accurately select buttons in a simple targeting task in our system. All of our independent variables showed no significance on correct button selection. The use of the head-mounted display to provide visuals of hand position provides a small time penalty in button selection. The use of the physical panel provided no significant effect in time compared to having the subjects target purely virtual buttons. Our optical tracking algorithm had a slight negative effect in time compared to using the capacitive touch sensors. Overall, subjects spend more time to select buttons in the virtual reality conditions, but just as accurately and successfully.

Chapter 4

Passive Haptics Experiment

The first experiment, described in the previous chapter, provided validation that the R3C technology could be used for interacting with a non-functioning mockup. We found that subjects had a near-perfect success rate using the hand tracking sensor to activate buttons, with accuracy comparable to their performance in a non-virtual environment. There was a difference in time, with the subjects performing faster in the real, non-virtual environment, and slower in all the virtual environment conditions. One of the unanswered questions was the effect of the panel as passive haptics on the movement time and accuracy. The difference in movement time between the subjects using the passive haptics and those without were insignificant. The main focus of this experiment is the independent variable of passive haptics. The experimental conditions are reduced from the last experiment to just two: one with passive haptics and one without. Both conditions are performed with the use of the head-mounted display and the hand tracker. A more thorough examination of movement time and accuracy is provided by the use of Fitts' Law, and additional dependent measures are also investigated, including presence, learning rate, and trajectory metrics. This experiment aims to provide a more complete understanding on how the use of passive haptics affects the interaction of a subject with a virtual cockpit panel. A version of this chapter was presented at the 2017 AIAA Modeling and Simulation Technologies (Joyce and Robinson, 2017).

4.1 Introduction

Passive haptics is a term that has been used to describe a variety of technologies or techniques to provide the sense of touch to a user of a virtual environment. It is often defined by its distinction from active haptics, which simulate the sense of touch with energy exchange, typically electromechanical. A common active haptic technology used in immersive virtual environments is a haptic glove, which utilizes small motors at the fingertips. In contrast, passive haptics often utilize proxy objects placed in the physical world to co-incide with the virtual environment experience.

The proxy objects can be simple or complex. They can be colocated and accurate with the virtual world or purposefully designed to trick the user. In our paper, we utilize a simple colocated passive haptic device and measure its effect on the presence and performance of subjects using a 2D panel in a 3D immersive virtual environment.

The advantages to using a simple passive haptic can be easily understood: less cost and complexity compared to most active haptic solutions. However, the disadvantage comes with its inflexibility. Due to their nature, passive haptics often have to be purpose built for a single or limited experience. While past research has aimed to address this, by either actively positioning a proxy object or simplifying the proxy object to fool the user, our application suffers only minimally from this limitation. The motivation for our research comes from the application of designing aerospace cockpits, complex human-machine interfaces where the user is stationed at their workspace. For the purpose of evaluating a cockpit design, the user does not need a dynamic tactile environment. Furthermore, many cockpit design processes already create a physical mockup which can provide the passive haptics for this evaluation.

We present our findings in testing passive haptics versus no haptics in an immersive virtual reality environment. Using a head-mounted display and a hand tracker, the subjects performed the same Fitts' Law style task under these two haptic conditions. The passive haptics was a flat surface placed at an angle on a desk in front of their seating area. Their performance on the Fitts' task was recorded as well as their responses to a presence survey, a self reported arm fatigue score and a general questionnaire.

4.2 Background

In this section, a brief summary of the relevant topics is given. A more thorough treatment of the background is in Chapter 1.2.

4.2.1 Passive Haptics

Passive haptics have seen limited, but promising results in research. Insko (2001) had subjects navigate a maze in a virtual environment. They found that subjects reported increased presence using passive haptics, and also found that subjects trained with the passive haptics performed better after they were removed than a control group that never used passive haptics. Borst and Volz (2005) created a 'mixed haptics' system, combining active haptic gloves with a physical panel for passive haptics. Again, performance was increased with the haptics, but minimal differences were found between using the mixed haptics and the passive haptics alone. Kohli et al. (2012) performed a Fitts' Law evaluation of passive haptics, but did not compare to a no haptics condition. Similar to our motivation and work, Schiefele et al. (1998) replaced a cockpit panel with a flat panel in an immersive head-mounted virtual environment, and found that users could activate buttons and switches in less time with the panel present than without. We build on this previous work by investigating the effects of passive haptics with the latest virtual environment technology, as well as performing a complete Fitts' Law characterization between no haptics and passive haptics.

4.2.2 Fitts' Law

Fitts' Law is a well established model of human movement that has been used extensively to measure the performance of input devices. We use the Fitts' throughput in this experiment to compare our experimental conditions. Throughput (TP) is defined as:

$$\mathrm{TP} = \frac{\mathrm{ID}}{\mathrm{MT}} \tag{4.1}$$

where MT is the movement time and ID is the index of difficulty. Index of difficulty is a property of the geometry of the targeting movement, combining the distance to the target from the starting location (D) and the width of the target (W). We use the Shannons' formulation of index of difficulty:

$$\mathrm{ID} = \log_2\left(\frac{D}{W} + 1\right) \tag{4.2}$$

The index of difficulty has units of bits, and the throughput has units of "bits per second" (bps). Throughput can be thought of theoretically as the amount of information ('bits') the human can input with the particular input device or method per time ('per second'). A higher throughput indicates that humans are able to use the input method more accurately and more quickly.

In this experiment, we use the ISO 9241-9 "Fitts' circle" standard (Figure 4.1). The adjustment for accuracy (Welford, 1968) is used to calculate the effective width. The effective width, W_e, performs a correction for the performance of the subject based on the end point positions on the target for all their movements. The effective width is calculated as $W_d = 4.1\sigma$, where σ is the standard deviation of the endpoint positions.

Previous work of evaluating virtual environments and input devices with Fitts' Law mostly consist of evaluating 3D tasks (Chun et al., 2004; Liu et al., 2009; Teather et al., 2010). Kohli et al. (2012) used a passive haptic environment with a head-mounted display, but evaluation focused on their warped virtual space technique. Recently, some reports have evaluated the Fitts' performance of the LeapMotion hand tracker (Coelho and

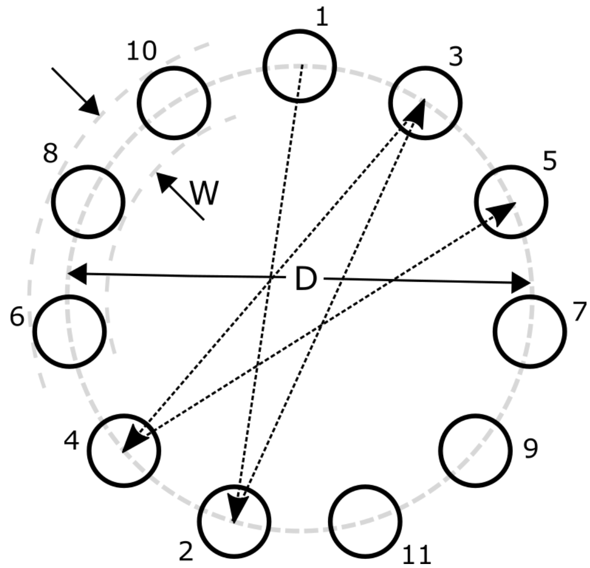


Figure 4.1: Fitts' Circle diagram. D is the distance between targets and W is the width of the targets.

Verbeek, 2014; Seixas et al., 2015), but none have combined hand-tracking with the use of a head-mounted display.

4.2.3 Presence

The feeling of presence is often specified as a goal of a virtual environment. A definition from Witmer and Singer (1998) reads:

> "*Presence* is defined as the subjective experience of being in one place or environment, even when one is physically situated in another."

Increased presence has been shown to lead to increased performance in a virtual environment task (Youngblut and Huie, 2003). We use a version of the Presence Questionnaire (PQ) proposed by Witmer and Singer (1998).

The post-questionnaire consists of 16 questions with a seven-point Likert scale. Nystad and Sebok (2004) found the PQ to be sensitive to technology and interaction methods, so it was chosen to investigate the differences between the two conditions.

4.2.4 Arm Fatigue

Despite the concern of arm fatigue in virtual environments (Burdea and Coiffet, 2003), it was surprising that most results in literature were anecdotal or for mitigation without quantification of the fatigue. The arm fatigue scale used within this experiment is the Borg Rating of Perceived Exertion (RPE) scale that ranges from 6 to 20 (Borg, 1998). Hincapié-Ramos et al. (2014) proposed a model for quantifying and predicting the amount of arm fatigue that correlated well with a Borg scale. Although there does not exist a standard for arm fatigue measurements in virtual environments, we hope that future researchers will consider including arm fatigue ratings in their experiments.

4.3 Methods

The purpose of the experiment described within this paper is to answer the following research questions:

1. Will Fitts' throughput be higher with passive haptics?
2. Do subjects learn the task more quickly with passive haptics?
3. What are the differences between the formation of reaching motion trajectories with passive haptics?
4. Do subjects report less arm fatigue with passive haptics?
5. Do subjects feel more presence with passive haptics?

4.3.1 Experimental Setup

For the experimental setup, subjects were seated at a desk with a blank panel mounted on an angle in front of them. The plywood panel (square with length of 18 in) was used to provide only the backstop of the virtual buttons for the "Passive Haptics" condition. The button selection is registered by the subject moving their index finger into a detection zone (cylinder for the circle buttons) in front of the button that extends outward 0.5 in. Their entrance into the hover zone is indicated to them by the button changing color. A successful button press is registered after 150 ms, and is indicated by the color turning off and a button click noise being played over speakers.

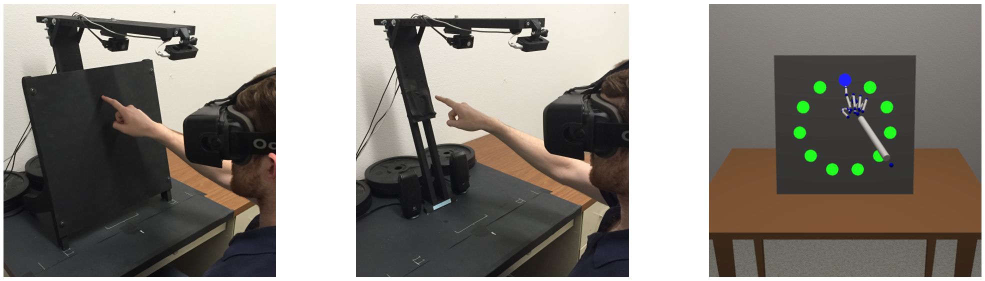
The equipment used consists of an Oculus Rift DK2 (Development Kit

2) head-mounted display (HMD) and a LeapMotion hand tracker. The low-persistence OLED display of the HMD has a resolution of 1920x1080, with a refresh rate of 75 Hz. The field of view is approximately 100°. It utilizes internal sensors (accelerometers and gyros) and an external infrared camera for head tracking. The configuration and calibration of the LeapMotion hand tracker is described in Chapter 2.3.2.

4.3.2 Experimental Task

The experimental task was a Fitts' circle in the virtual environment, performed by subjects in two haptic conditions. The subjects were seated at a desk for the experimental task, and the circle was located on a panel mounted on the desk. The two haptic conditions were "No Haptics (NH)" and "Passive Haptics (PH)." These conditions are pictured in Figure 4.2. For the "Passive Haptics" condition a physical panel was co-located with the panel in the virtual world, and it was removed for the "No Haptics" condition.

With the two haptic conditions, there were no differences in the task itself or in the method of button activation. The only difference between the conditions was the removal of the physical panel. The hand tracker remained in the same location, preserving the location of the buttons in the



(a) Passive Haptics condition (PH) **(b)** No Haptics condition (NH) **(c)** View of virtual world (same for both)

Figure 4.2: Experimental conditions and view of virtual environment.

virtual environment. The dimensions of the virtual world were no different for either condition. In fact, there was no change to the software between the conditions.

Subjects performed the Fitts' circle for three different distances (20 cm, 30 cm and 40 cm) and five different button widths (5 mm, 10 mm, 15 mm, 20 mm and 25 mm). These configurations were chosen to span a wide range of indices of difficulty (3.2 to 6.4). For each configuration of distance and width, subjects had to complete the full pattern of 11 buttons three times consecutively. This set of 30 movements[1] for a single configuration is referred to as a single trial for a subject. The distance was kept constant for the consecutive trials until all five button widths were complete. The distances were presented in either smallest to largest or vice versa, which was counterbalanced among subjects.

This set of 15 trials was repeated for each haptics condition and the

[1] 10 movements between the 11 targets.

order was kept the same within subjects. The sequence that the two conditions were presented to each subject was also counterbalanced. The dependent measures will be tested for interaction with sequence to determine if the results were affected by condition order.

4.3.3 Experiment Design

As described in the previous section, the experiment was performed with a within-subjects design. Subjects were asked to complete the same experimental task for both conditions of haptics: Passive Haptics (PH) and No Haptics (NH). The two different haptic conditions are the main independent variables.

We expected significant skill transfer between the two conditions, so the order in which subjects performed the two conditions was counterbalanced. This created a second independent variable that is between subjects. The subjects who performed the conditions with the PH being their first condition were one group, and the subjects who performed NH as their first condition were a second group. We call this grouping "sequence" and refer to the two groups as "PH First" and "NH First".

For a Fitts' Law evaluation, it is often recommended that the data collection only begins when the subject is fully trained on the task. However, one of the goals of the experiment is to investigate the learning rate of the subjects. For that reason, the subjects were given no separate training time for the task or virtual environment. In lieu of collecting the data from fully trained subjects, the throughput analysis will be carried out on movements that are determined to be composed of mostly ballistic motion. The filtering parameters were determined post-hoc from the trajectory recordings. Their development and parameter selection are discussed in the results.

4.3.4 Dependent Measures

The main dependent measure is the Fitts' throughput, measured through the movement time between button presses. Additionally, the trajectory of each movement is recorded from the hand tracker for analysis. To determine the arm fatigue, subjects were asked to rate their arm fatigue on the Borg scale from 6 to 20. The scale was presented with anchors as shown in Appendix A.1 (Table A.13). The arm fatigue rating was collected at the beginning of each condition, and then after every other configuration of distance and width combination (and after the final trial due to an odd number of trials). At the completion of each condition the subject was given a presence questionnaire. At the end of the experiment, an additional condition comparison survey was given to ask for opinions on the two haptic conditions.

4.3.5 Trajectory Phases

Human reaching movements have long been known to consist of two distinct phases (Woodworth, 1899): a ballistic phase and a corrective phase. The ballistic phase is an open-loop, gross adjustment towards the target, while the corrective phase is more refined movement honing in on the target. To separate the trajectories into the various phases, a simple algorithm was developed. First, the local minima are found throughout the velocity profile of the movement to separate the movement into various submovements. The "ballistic phase" is then classified as the submovement which contains the peak velocity of the entire movement. The various submovements after the ballistic phase are classified as the "corrective phase". Any movement before the ballistic phase is classified as a "reaction time". An example of the results of this classification is shown in Figure 4.3. This mathematical definition does break down for certain cases where a subject might have two submovements in the ballistic phase due to a mid-course correction or similar, however one of the main purposes of this classification for this experiment was to find movements which are appropriate to use for the Fitts' Law calculations. Therefore, movements with multiple submovements are likely to not be ballistic, well-learned movements.

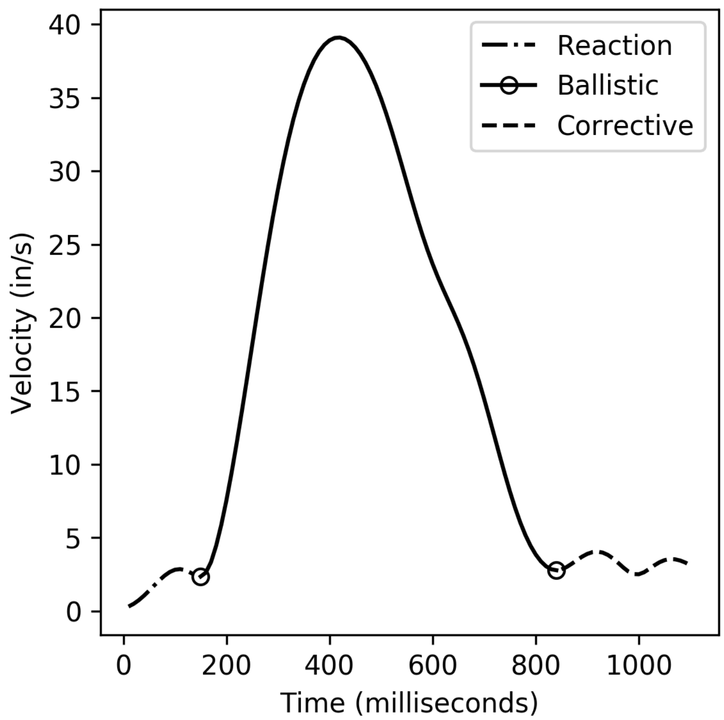


Figure 4.3: Example trajectory with three phases indicated.

4.3.6 Trajectory Filtering

We used a low-pass filter on the trajectory recordings to reduce the amount of noise. The LeapMotion processes data at a variable frequency, thus creating a variable rate for recording. The frequency typically varies from about 100 Hz to 120 Hz. To perform the filtering, the data was first resampled to a fixed rate of 100 Hz. The filter used is a fourth-order Butterworth filter with a cut-off frequency of 5 Hz. The cut-off frequency was chosen as voluntary hand movements have been shown to be below such a rate (Riviere et al., 2003).

4.4 Results

4.4.1 Participants

Twenty (20) subjects were recruited from the UC Davis engineering student population, both undergraduate and graduate students. The age range was 19 to 29 ($M = 23.0, \sigma = 3.0$) with 16 males and 4 females. The genders were balanced among the counterbalanced sequence groups. All subjects indicated either less than one hour or no prior experience with virtual reality. The total time spent in the experiment was under one hour, with a range of approximately ten to twenty minutes performing each of the two conditions in virtual reality.

4.4.2 Throughput

The throughput is calculated per movement using Equation 4.1 and averaged by trial before being averaged by subject and condition. Throughput is used to investigate the first two research questions: do the subjects learn more quickly and does their throughput performance improve with passive haptics? We first present the learning rate results, but before the effect of throughput performance on passive haptics is reported, our data reduction methods are discussed.

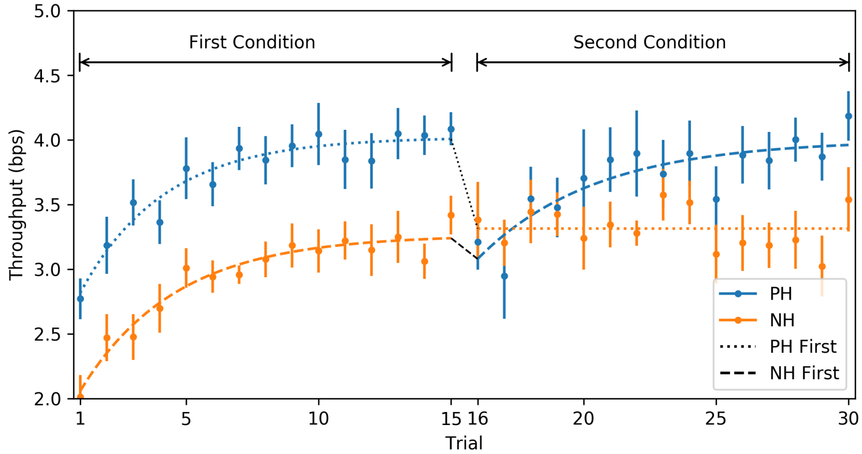


Figure 4.4: Throughput per trial. The learning curve exponential fit is given by Eq. 4.3 with parameters from Table 4.1.

Rate of Learning

The average throughput for each trial, separated by haptics condition, is shown in Figure 4.4. An exponential rise to a learned state is fit to each haptics condition to model the learning curve of the subjects. The equation used is given as:

$$\mathrm{TP}(T) = \mathrm{TP}_\infty - (\mathrm{TP}_\infty - \mathrm{TP}_0)e^{(-T/\tau)} \tag{4.3}$$

where T is the trial number, TP_∞ is the asymptotic learned value of throughput, TP_0 is the initial value at $T = 0$ and τ is the time constant. The time constant is defined as the number of trials for the throughput to rise 36% of the difference between the fully learned throughput (TP_∞) and the initial throughput (TP_0). The parameters of the fit for each condition is shown in Table 4.1, as well as the standard error of the estimate (SEE). The value of throughput from the regression fit as a percentage of the fully learned value (TP_∞) for representative trials are listed in Table 4.2.

The rate of learning is very similar among both groups in the first condition. For the first condition it can appear that the subjects performing NH learned more quickly than their counterparts performing PH by looking at the time constants. However, as the values in Table 4.2 show, the NH First group started their first trial at a lower percentage of their learned state. The shape of the learning curves are very similar and both groups reached approximately 90% of their fully learned state by the 5^{th} trial.

The learning curves are quite different for the second condition. The NH condition does not have a learning curve at all, with a straight line being a better fit than the exponential function. This indicates the transfer of training from the PH condition allowed the group who did PH first to immediately perform in NH at the same level as the fully learned state of the subjects who learned NH in their first condition. This transfer of training to the second condition did not occur as strongly for the group who did NH first. Their initial performance of PH did start out at a slightly higher level than the subjects who did PH first (2.8 bps vs 2.4 bps), but after 5 trials this difference had converged (3.6 bps vs 3.7 bps).

Sequence	Condition	Haptics	TP_∞	TP_0	τ	SEE
PH First	First	PH	4.0	2.4	3.2	0.58
PH First	Second	NH	3.3	3.3	0.0	0.65
NH First	First	NH	3.3	1.7	3.6	0.49
NH First	Second	PH	4.0	2.8	4.5	0.78

Table 4.1: Exponential fit parameters of Eq. 4.3. Curves are shown in Figure 4.4. TP_∞ is the asymptotic learned value of throughput. TP_0 is the initial value at trial 0. τ is the time constant of the exponential function. SEE is the standard error of the estimate for the fit.

Sequence	Condition	Haptics	TP_1	TP_5	TP_{10}	TP_{15}
PH First	First	PH	70.0%	91.5%	98.3%	99.6%
PH First	Second	NH	100.0%	100.0%	100.0%	100.0%
NH First	First	NH	63.0%	87.8%	97.0%	99.2%
NH First	Second	PH	77.0%	90.6%	97.0%	99.0%

Table 4.2: Percentage of fully learned state for various trials for each group and condition from learning model fit. TP_i is $TP(i)/TP_\infty$ using Eq. 4.3. Trial 15 is the final trial for each condition.

These results indicate that subjects did not learn faster with Passive Haptics. The only differences between learning rates is the positive transfer of training from performing Passive Haptics first followed by No Haptics second. The answer to the research question *Do subjects learn the task more quickly with passive haptics?* appears to be that the passive haptics does not make subjects learn faster, but they are able to learn the task more quickly without passive haptics afterward. It does appear that the fully learned state is different between the haptic conditions, which is investigated through the use of throughput after discussing data filtering.

Ballistic Movement Filtering

Before they could be used for the Fitts' Law analysis, the trajectories were filtered so that only movements which were direct to target were included. A well-learned movement appropriate for Fitts' Law is one which moves directly towards the target and does not have much extraneous movement or idle time beyond what is required to complete the task. As a result of the experiment design and research questions requiring subjects to learn the task during the data collection, we expected many of the movements to deviate from a well-learned movement. A number of common reasons for deviation were observed during the experiment and post-hoc examination of the dataset. The goal of this filtering is to produce a dataset that only includes the direct to target movements appropriate for a Fitts' Law analysis for the final throughput calculations. We describe in this section the three metrics developed based on these observed deviations that were used to determine whether a movement was direct to target.

The first metric was the ratio of path distance traveled in the ballistic phase to the distance between the targets for that movement (i.e. 20 cm, 30 cm and 40 cm). Ideally, the ballistic portion would cover the majority of the distance between the targets. A movement which covered too little of the target distance could mean the subject slowed or stopped in the middle of movement, and one that covered more could mean the subject overshot or had an indirect trajectory. A sample movement that gets flagged by this filter is shown in Figure 4.5, which has a ratio of 1.49. This is an example of a movement that does seemingly have a direct movement toward the target, but includes a large deviation perpendicular to the movement axis. This deviation could mean the subject initially aimed their ballistic portion in the wrong direction but performed a correction during the movement. The limits for this filter were chosen as having a ratio between 0.90 and 1.10, i.e. within 10% of the target distance. This filtered out 4577 of the 17970 movements (25.5%).

The second filtering metric was also based on the ballistic phase path distance. For this metric the ballistic phase path distance was compared to the total path distance the subject traveled for their entire movement. This filter targeted movements where, after the ballistic phase, the subject moved away from the target, or had a smaller but significant movement before the main ballistic movement. An example is shown in Figure 4.6 which shows a movement which passed the first metric (the ballistic phase distance ratio to the target distance was 1.03), but the ballistic phase was only 56% of the total distance traveled during that movement. This was a common problem where a subject would have a false start for the next target and move away from the current target before their button press was activated. The threshold for the filter was set at 0.80, which rejects 4264

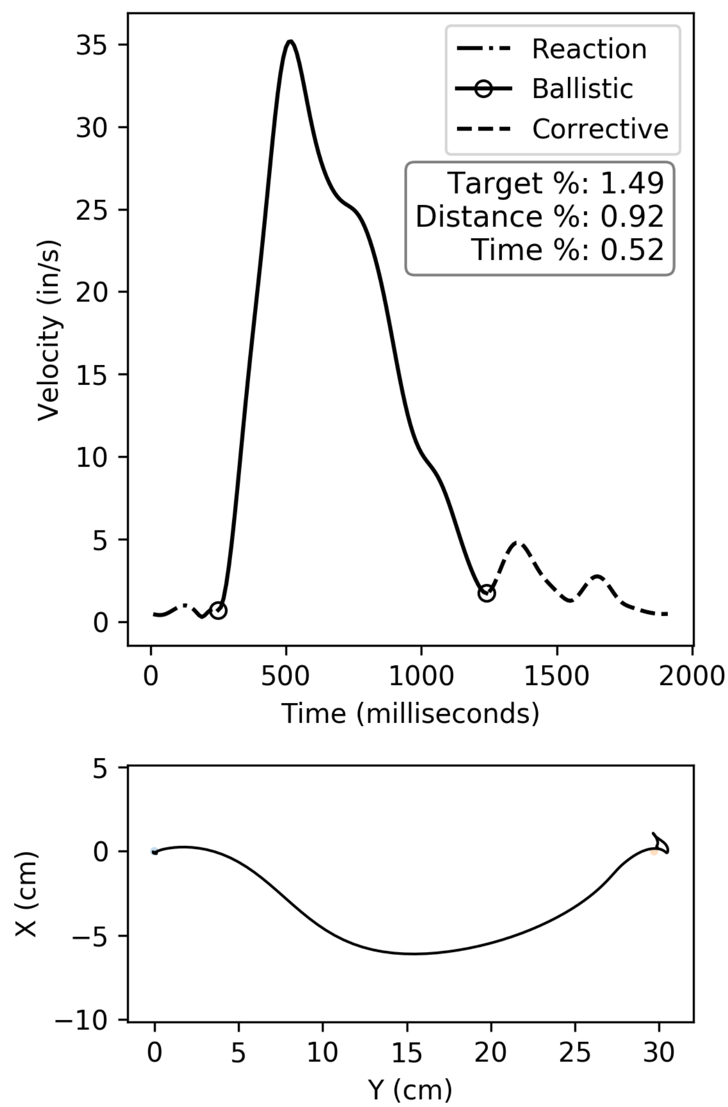


Figure 4.5: An example of a movement with a large ballistic distance to target distance ratio. The bottom plot shows the projection of the trajectory on the plane of the Fitts' circle.

movements that are lower than the threshold. Of the rejected movements, only 1675 were unique from the rejected movements of the previous target distance filter. Between the two filters so far, there have been 6252 of the 17970 movements filtered out (34.8%).

The last filter took a time based approach, and looked at the ratio of the time spent in the ballistic phase over the total movement time. This filter removed movements where the subject spent an inordinate amount of time either before or after the ballistic phase. If they were not moving during the non-ballistic phases, it would not have been caught by the distance-based filters either. The threshold of 0.40 meant that 4496 movements were filtered, however only 873 of those are unique of the other two filters. Figure 4.7 illustrates a movement where the subject waited before initiating the ballistic movement. Since the subject did not move during this idle time, the other two filters did not flag this movement.

The combination of these three filters led to 7125 of 18000 movements being filtered out, leaving 60.4% of the movements. For each of the metrics, the threshold was determined by investigating the distribution and looking at sample movements on either end of the threshold to determine if it was an appropriate value. The histograms of these metrics as well as more sample movements are shown in Appendix **??**. The major conclusions in the following sections do not change with different choices in the thresholds

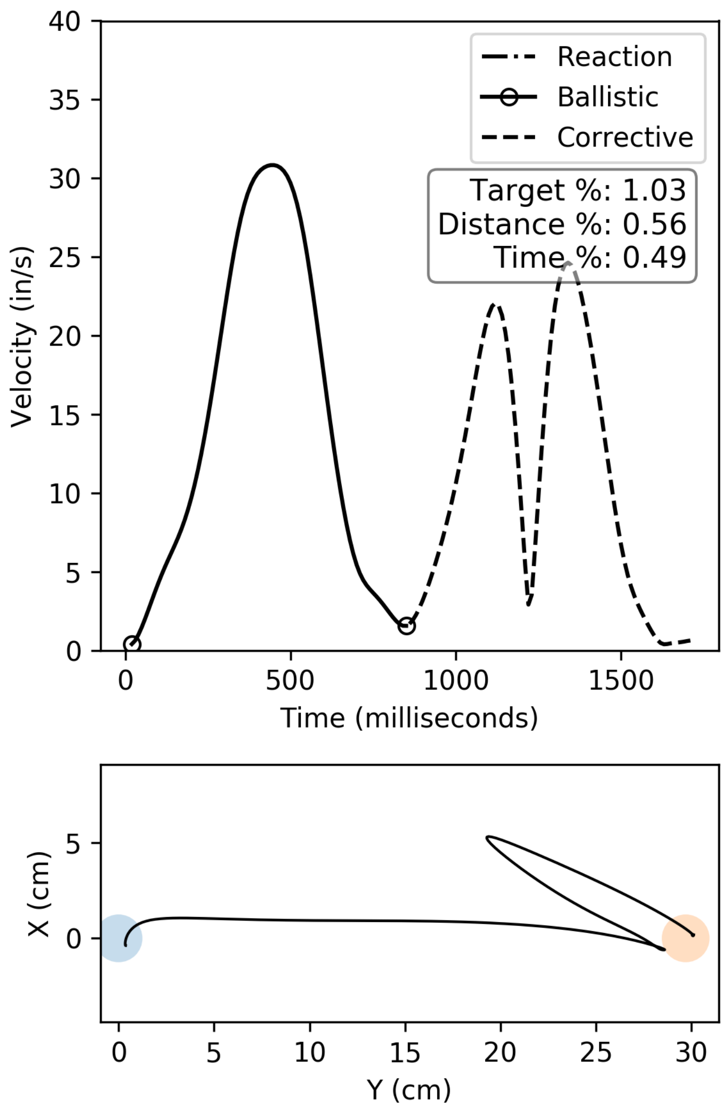


Figure 4.6: An example of a movement with a small ballistic distance to total path distance ratio. The bottom plot shows the projection of the trajectory on the plane of the Fitts' circle.

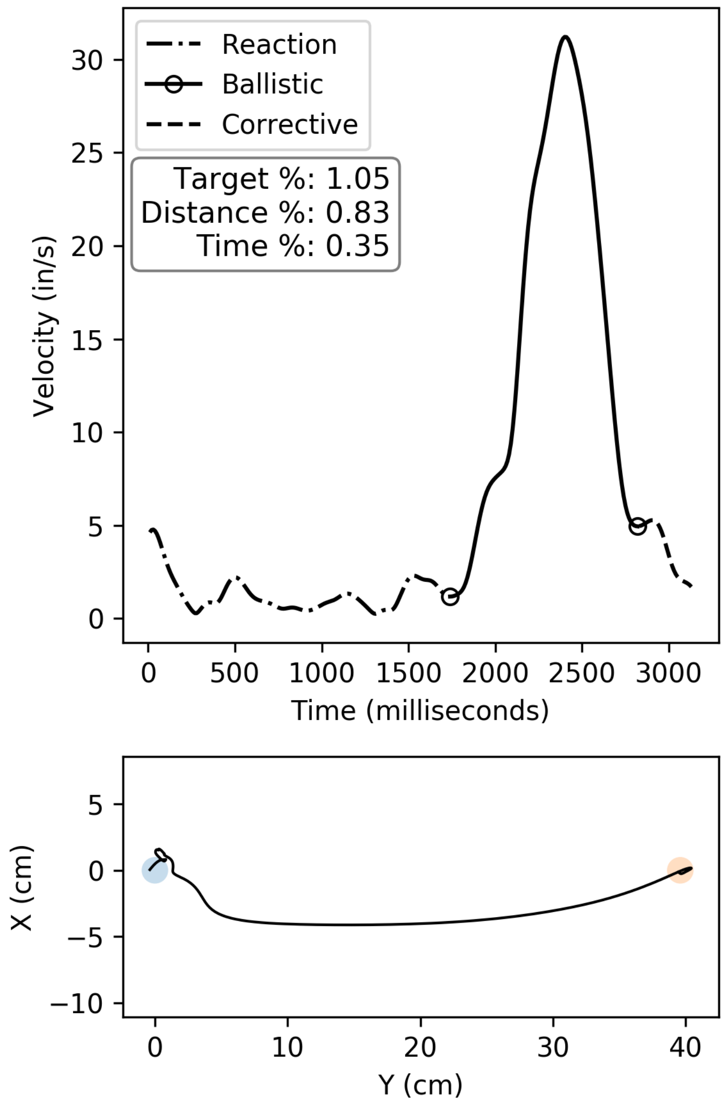


Figure 4.7: An example of a movement with a small ballistic time to total time ratio. The bottom plot shows the projection of the trajectory on the plane of the Fitts' circle.

(i.e. the statistical effects are not sensitive to the threshold choices).

The final check before performing the Fitts' calculation was to ensure that each trial had enough data for the adjustment for accuracy calculation. The adjustment for accuracy is based on the distribution of endpoint data from a single trial (which is one distance and width configuration). One trial consists of 30 consecutive movements, but the ballistic filtering could diminish the amount remaining in each, so a trial was only included if at least half (15 of 30) of the movements were considered to be good movements by the ballistic filters. This means that movements from a trial that did not have enough good movements were also filtered out from the Fitts' calculation. Not only is this important to make sure the adjusted width is valid, it also removes trials where the subject likely did not reach a fully learned state, as most of their movements were not primarily ballistic movements direct to target. On average, 11 of 15 trials per condition from each subject ($M = 10.98, \sigma = 3.25$) had enough good movements to be included. This left 9218 movements for the Fitts' calculation, just over half of the total movements (51.3%). Slightly more movements were filtered from the No Haptics condition, with 47.1% of NH movements left after the filtering, compared to 55.5% of the PH movements. For the remaining sections, unless otherwise noted, the results were based on using only the movements after filtering.

	Filtered		Unfiltered	
Haptics	Mean	SD	Mean	SD
PH	4.25	0.44	3.86	0.50
NH	3.76	0.38	3.25	0.37

Table 4.3: Throughput scores by haptics condition. Results are shown for the dataset before (Unfiltered) and after (Filtered) the ballistic filtering.

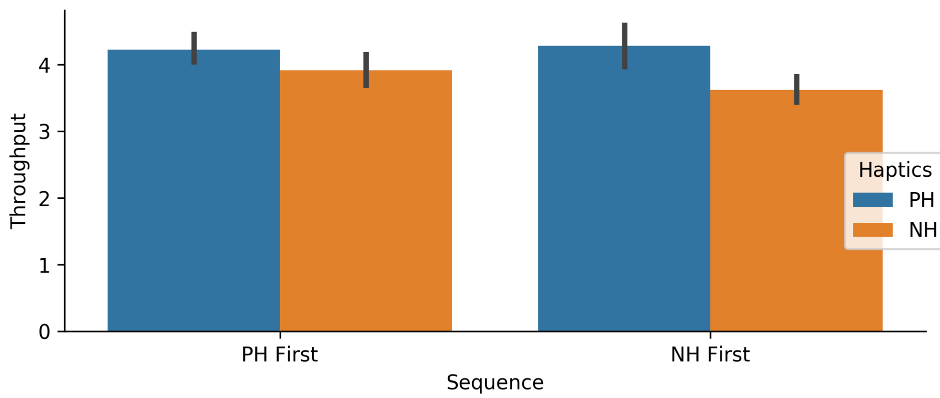


Figure 4.8: Throughput results by haptics and sequence.

Throughput

In this section, we investigate the first research question, *Will Fitts' throughput be higher with passive haptics?*. The throughput results are listed by haptic condition in Table 4.3. Figure 4.8 shows the throughput separated by haptics and sequence (the order the subjects performed the haptic conditions).

Throughput was found to be higher in the PH condition, at 4.25 bps compared to the 3.76 bps of the NH condition. A two-way mixed ANOVA was performed to determine the effect of haptics condition. Since we expected order effects, and have seen them with the transfer of training seen in the Rate of Learning section, the sequence the subjects performed the haptic conditions was included as a between subjects factor. The effect of haptics was found to have a significant effect on the throughput ($F(1, 18) = 35.59, p < 0.001$) between the PH condition ($M = 4.25\,\text{bps}, \sigma = 0.44\,\text{bps}$) and the NH condition ($M = 3.76\,\text{bps}, \sigma = 0.38\,\text{bps}$). There was no effect on throughput based solely on sequence group ($F(1, 18) = 0.53, p = 0.47$), but there was a marginally significant interaction effect between the sequence and haptics ($F(1, 18) = 4.48, p = 0.048$).

As can be seen in Figure 4.8, this marginal interaction effect appears to indicate that both groups had improved performance but that the PH First group performed better at the NH condition. A post-hoc repeated measures t-test between haptic conditions for the subjects who performed PH First was significant ($t(9) = 4.62, p < 0.001$), with the PH condition ($M = 4.23\,\text{bps}, \sigma = 0.34\,\text{bps}$) outperforming the NH condition ($M = 3.91\,\text{bps}, \sigma = 0.40\,\text{bps}$). The mean of the differences between subjects was 0.32 bps. The group of subjects who performed NH first also had a significant effect in the t-test ($t(9) = 3.96, p < 0.001$) between the PH condition ($M = 4.28\,\text{bps}, \sigma = 0.54\,\text{bps}$) and the NH condition ($M = 3.62\,\text{bps}, \sigma = 0.32\,\text{bps}$). There was a higher mean of differences (0.66 bps) with the NH First group than the PH First group. These post-hoc tests confirm that both groups had a significant effect of haptics, though the NH First group had a larger difference between the conditions.

It is worth noting that without using the ballistic filter, the major conclusions found do not change. The only major difference is the magnitude of the throughput and size of the differences. The statistical tests have the same results (these are shown in Appendix A.1: Tables A.4-A.6). The results by haptics for both filtered and unfiltered are shown in Table 4.3. These results indicate that subjects do have higher throughput with passive haptics, answering our first research question.

4.4.3 Trajectory Phases

As described in Section 4.3.5, each movement of the subjects can be dissected into three distinct phases: reaction time, ballistic phase, and corrective phase. We have already seen that, overall, the subjects took more time to complete a movement without the passive haptics in place. In this section we investigate the differences in time spent in the three phases. We report here the means of time spent in each of the three phases. Times reported are all milliseconds. The results are shown in Table 4.4 for the filtered and unfiltered results, where the filtered results only include movements that were deemed primarily ballistic by the filtering methods in Section 4.4.2. The unfiltered results include all movements. The phases

		Filtered			Unfiltered		
		Mean	SD	p	Mean	SD	p
Reaction Time	PH	118.58	45.32	0.57	241.57	205.39	0.33
	NH	121.44	34.80		291.97	169.19	
Ballistic Phase	PH	719.54	59.99	< 0.001	679.48	44.79	< 0.001
	NH	772.57	67.62		727.53	60.01	
Corrective Phase	PH	199.13	48.10	< 0.001	564.73	253.56	< 0.001
	NH	256.58	42.08		837.22	293.15	

Table 4.4: Time in each movement phase by haptics conditions.

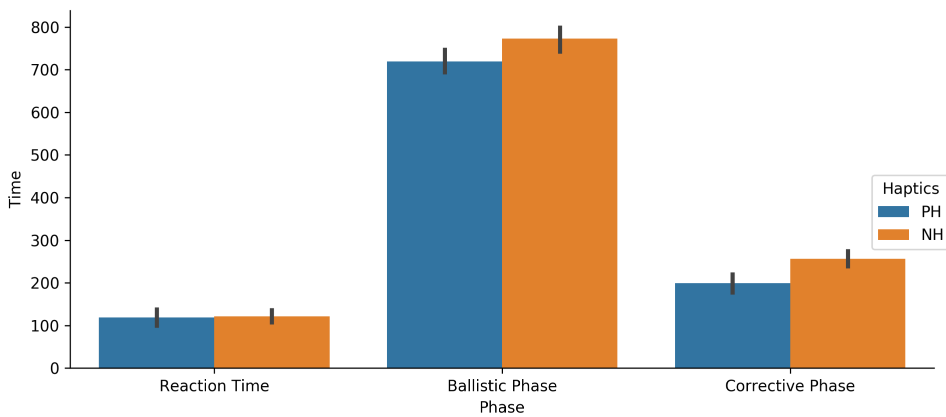


Figure 4.9: Time spent in each trajectory phase by haptics.

were averaged by subject first, and then by condition. The time spent in each phase by haptic condition is shown in Figure 4.9. Each phase was tested for the effect of haptics and sequence through a mixed two-way ANOVA. The statistical tests reported here are for the filtered results, but the significance findings do not change between filtered and unfiltered.

The reaction time was not significantly affected by haptics ($F(1, 18) =$

$0.32, p = 0.58$) or sequence ($F(1,18) = 0.001, p = 0.98$). However, for the interaction between the two a significant effect was found ($F(1,18) = 18.56, p < 0.001$). The interpretation of this interaction effect without main effect significance is that the first and second condition had a different reaction time, without dependence on the haptics condition or group. The mean reaction time in the first condition for both groups was 130.9 ms ($\sigma = 41.0$ ms), but in the second condition it was just over 20 ms faster, with an average of 109.1 ms ($\sigma = 36.6$ ms). The reaction time for both conditions was lower than generally accepted values for reaction time to a visual or aural stimulus (Teichner, 1954). This is not surprising, as the task was a serial task which the subjects would learn the pacing of throughout the experiment. It would be expected that they were able to learn to anticipate the activation of a button (which was also the start of the next movement) as it would activate 150 ms after the subject entered the zone of the previous button. This interaction effect indicates that subjects did learn how to anticipate the activation event independent of the haptics or the order they performed the conditions.

The ballistic phase time had a significant effect of haptics ($F(1,18) = 24.14, p < 0.001$) between PH ($M = 772.6$ ms, $\sigma = 67.7$ ms) and NH ($M = 719.5$ ms, $\sigma = 60.0$ ms). There was no effect of sequence or the interaction effect between haptics and sequence. Since the ballistic phase should be mostly independent of the use of passive haptics, it was not expected to see the ballistic phase have an effect of haptics. It is unclear the exact mechanism that led to this, but it could be an artifact of the passive haptics causing the subjects to learn the movement, and thus allowing them to move more quickly. The difference between the two haptic conditions was small in magnitude (53 ms or 7%), so this may not be a practical significance. There is little difference between the filtered movements and unfiltered movements results, the difference of the means were within a few milliseconds. Although the ballistic phase had lower movement times with the unfiltered movements, this is due to including movements where the ballistic was not the complete movement, reducing the mean.

The corrective phase time also had a significant effect of haptics ($F(1, 18) = 22.46, p < 0.001$). Subjects spent an additional 57.5 ms in the corrective phase with NH ($M = 256.6\,\text{ms}, \sigma = 42.1\,\text{ms}$) than they did with PH ($M = 199.1\,\text{ms}, \sigma = 48.1\,\text{ms}$). There was no effect of sequence or the interaction effect between haptics and sequence. This result was expected as one of the main benefits of the passive haptics is that the subject does not have to 'find' the target along one dimension. With the Passive Haptics condition, the subjects can rely on the backstop of the passive haptics to stop their movement normal to the target. This allows them to focus their corrective phase on finding the button in the plane of the target. In the

No Haptics conditions, the subject must slow to a stop at the button and find the button in all three dimensions without the backstop to aid them.

There was a more noticeable difference between the results of the unfiltered and filtered movements for the corrective phase. The corrective phase time was much higher in both conditions for the unfiltered, with PH having a mean of 564.7 ms ($\sigma = 253.6$ ms) and NH having a mean of 837.2 ms ($\sigma = 293.2$ ms). This was expected as many of the non-ballistic movements had to spend more time in the corrective phase to recover.

These results provide insight into our third research question *What are the differences between the formation of reaching motion trajectories with passive haptics?*. The analysis finds that the most notable difference is that subjects spend significantly less time in the corrective phase with the passive haptics.

4.4.4 Arm Fatigue

The subjects were asked for a rating of their arm fatigue every other trial, as well as before the first trial and after the last trial of each condition. One trial lasted for 30 movements and consisted of a single distance and width configuration. The scale ranged from 6 to 20, and subjects were allowed to record decimal ratings. The full scale with anchors are shown in Appendix A.1 (Table A.13). The average rating at each trial, separated

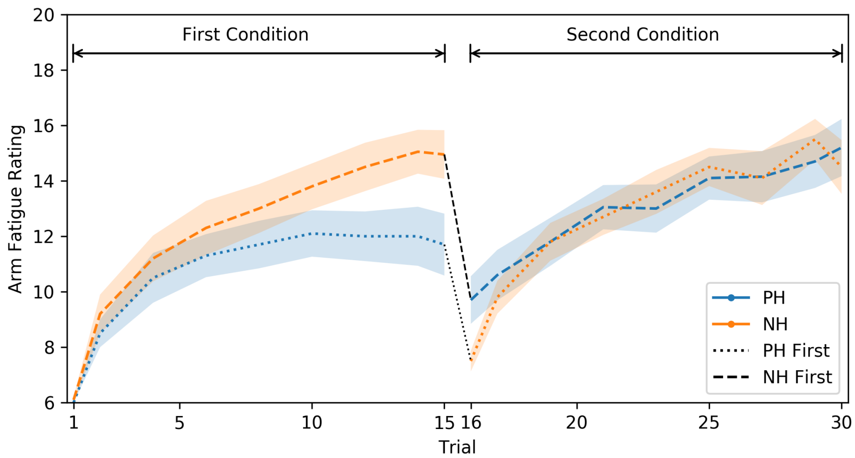


Figure 4.10: Arm Fatigue by Trial.

by haptics condition, is shown in Figure 4.10.

There is an evident difference between the two haptic conditions for the first condition performed, with the NH condition subjects accumulating more fatigue throughout the trials. At the end of the first condition, the subjects who performed the PH condition rated their arm fatigue 3.25 points lower on average than the NH condition subjects. Both groups have a similar rate of recovery, but the second condition quickly converges and shows no apparent difference between the two haptics conditions.

A within-subjects repeated measure (haptics) with two between-subjects measures (sequence and trial) ANOVA was performed to test the significance of haptics and the interaction effect of haptics and sequence. The interaction effect of haptics and sequence was found to be significant ($F(1, 18) = 22.6, p < 0.001$) as well as the main effect of haptics

($F(1, 18) = 5.47, p = 0.03$). As a result of the significant interaction effect, a post-hoc ANOVA with haptics and trial as the two within subjects repeated measures was run on both sequence groups. The NH First group had no effect due to haptics ($F(1, 9) = 2.08, p = 0.18$), consistent with the observations from Figure 4.10. The subjects rated the same trial between conditions an average of only 0.69 points ($\sigma = 2.1$) higher for the PH condition, which was their second condition. The PH First group did have a significant effect due to haptics ($F(1, 9) = 42.37, p < 0.001$). This group rated the PH condition an average of 2.0 points ($\sigma = 2.0$) lower within trials. There are two factors that could contribute to the subjects experiencing less fatigue in the first condition with passive haptics. The first is that with the help of the passive haptics, the subjects spent less time with their hand extended unsupported. The second factor is that the passive haptics condition took less time than the no haptics conditions for the subjects to complete (this can be seen with the higher throughput result). In the second condition, both NH and PH experienced similar levels of fatigue. This suggests that by this point in the experiment, the subjects had accumulated too much fatigue to see any benefit from the passive haptics.

These results show that the subjects only had reduced arm fatigue using the passive haptics for the first condition. For all other conditions the arm fatigue ratings reached the same level by the end of the condition. This provides an answer to our fourth research question, *Do subjects have lower arm fatigue with passive haptics?*. Subjects reported lower arm fatigue using passive haptics for the first condition, but the cumulative fatigue negated the effect by the second condition.

4.4.5 Presence

The presence survey was administered after each haptics condition. The questions had a 7-point Likert scale response with anchors at either end and the middle. The score given in this section is a sum of the responses, on a scale of 1 to 7, where a higher score indicates higher presence. A few questions were asked with an inverted scale (i.e. a score of 1 indicated higher presence) and were reversed before the score was calculated. The internal consistency of the presence questionnaire was tested per condition using Cronbachs' alpha, and was found to be consistent in both conditions ($\alpha = 0.72$ and $\alpha = 0.71$ for PH and NH, respectively). The full survey questions and average responses per condition are listed in Table 4.6. The item total correlation (the correlation between the questions' score and the total score) is also listed for each question.

The average scores per condition are given in Table 4.5. The presence scores were tested with a mixed within subjects repeated measures (haptics) and between subjects measures (sequence) ANOVA. The score

Haptics	Score	Std. Dev	Cronbach's alpha
PH	77.7	9.56	0.718
NH	71.0	9.70	0.711

Table 4.5: Presence Score Summary

had a marginally significant effect between haptics conditions ($F(1, 18) = 6.08, p = 0.024$), with Passive Haptics having a slightly higher mean ($M = 77.7, \sigma = 9.56$) than No Haptics ($M = 71.0, \sigma = 9.70$). There was no significant effect of sequence ($F(1, 18) = 4.01, p = 0.58$) nor for the interaction effect between sequence and haptics ($F(1, 18) = 0.71, p = 0.41$).

The questions that correlated most strongly with the presence score were questions 1, 2 and 7, which asked about control mechanisms and how natural they felt. The PH condition found these to be significant ($p < 0.001$). The question with the biggest difference between the two conditions was question 5 which asked directly about the tactile aspects of the environment. The PH condition scored 3 points higher on average on this question. After that, the questions with the largest difference were questions 1, 2, 10 and 15: all questions asking about control mechanisms again, and all providing a larger score with the PH condition.

The fifth research question, *Do subjects feel more presence with passive haptics?*, can be answered with these results. Subjects reported a marginally significant increase in presence for the passive haptics condi-

Question	NH		PH	
	Score	ITCorr	Score	ITCorr
1. How much were you able to control events?	4.8	0.79	5.3	0.74*
2. How natural did your interactions with the environment seem?	4.3	0.45	5.0	0.81*
3. How much did the visual aspects of the environment engage you?	5.4	0.56	5.3	0.65
4. How much did the auditory aspects of the environment engage you?	6.2	0.52	6.0	0.19
5. How much did the tactile (sense of touch) aspects of the environment engage you?	2.6	0.51	5.6	0.73
6. To what extent did you associate the computer generated arm and hand with being “your body” while in the virtual environment?	4.5	0.55	5.0	0.65
7. How natural was the mechanism which controlled movement through the environment?	4.0	0.48	4.5	0.72*
8. How much did your experiences in the virtual environment seem consistent with your real-world experiences?	4.0	0.49	4.5	0.66
9. How involved were you in the virtual environment experience and the task you were performing?	5.8	0.70	5.8	0.22
10. How distracting was the control mechanism?†	4.0	0.72	4.7	0.29
11. How much delay did you experience between your actions and expected outcomes?†	4.7	0.58	4.9	0.04
12. How quickly did you adjust to the virtual environment experience?†	2.9	−0.08	2.1	−0.51
13. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?	4.9	0.65	5.2	0.25
14. How much did the control devices interfere with the performance of assigned tasks?	4.0	−0.40	4.0	−0.45
15. How well could you concentrate on the assigned tasks rather than on the mechanisms used to perform those tasks?	4.7	0.42	5.5	0.17
16. Were you involved in the experimental task to the extent that you lost track of time?	4.3	0.22	4.2	0.53

Table 4.6: Presence questions and scores for each condition. ITCorr is the item total correlation, where * indicates a significant correlation ($p < 0.001$). † indicates a question which where a lower score indicated higher presence and were inverted before reporting.

Question	NH	Neither	PH
1. In which condition did you feel you performed faster?	1	1	18
2. In which condition did you feel you performed more accurately?	3	2	15
3. Did you feel that your arm fatigued more or quicker in one condition over the other?	13	3	4
4. Did you feel that you were actually in the virtual room more in one condition over the other?	3	4	13
5. Which condition did you prefer?	0	2	18

Table 4.7: Condition comparison survey summary of results.

tion.

4.4.6 Condition Comparison

The condition comparison survey asked the subjects five questions directly comparing the two conditions. The questions and a summary of the responses are shown in Table 4.7. The subjects overwhelmingly responded that they preferred the Passive Haptics (PH) condition (Q5), with all but two subjects choosing it. In fact, no subject preferred the No Haptics (NH) condition. The two subjects who did not choose PH responded that neither was more preferred.

The other questions had responses similar to the results from the other sections. The majority of subjects responded that they were more accurate and faster in the PH condition (Q1 and Q2), which agrees with the throughput results. The subjects who chose Neither or NH were usually not actually faster or more accurate in the NH condition. In fact, only two subjects had a throughput that was higher in the NH condition, and neither subject chose NH as the condition they performed faster in, though one did say they performed more accurately in the NH condition.

Question 4 asked subjects directly about their feeling of presence, and 13 subjects said that they felt more present in the virtual room with PH, with the remaining split between 4 saying neither and 3 saying NH. The results of the presence questionnaire suggested that subjects felt more present with the passive haptics, which this agrees with. The arm fatigue question (Q3) was worded to ask which condition they felt provided either more or quicker arm fatigue, and 13 subjects felt this was the case with the NH condition. The remaining were split between neither (3) and PH (4). The results of the arm fatigue questionnaire during the experiment found a similar result as the condition comparison survey.

4.5 Discussion

After the subjects performed their first condition, the experimenter would re-position the panel and explain their next condition. This meant that subjects who started with the panel (PH) would then see the panel be removed and have the no panel (NH) condition explained to them, and vice-versa. Subjects often made comments during this change about how they were happy for the panel to be added, or concerned about losing it. For example, one subject upon the placement of the panel commented "this is definitely going to be better." Another who was transitioning from PH to NH simply remarked, "Oh, no!" One of the subjects who felt that they performed faster and more accurately without the panel (NH condition) in the final questionnaire explained afterward that this was due to often needing to press the button twice in the PH condition, which we believe was likely due to their difficulty learning the button detection algorithm.

The throughput found in both conditions compares to the range found for a mouse input. A review of nine ISO 9241-9 style Fitts' Law studies by Soukoreff and MacKenzie (2004) found that the range for a typical computer mouse was between 3.7 bps and 4.9 bps. The value of throughput for a touchscreen device (direct input) has been found to be much higher, with one study measuring 6.95 bps (MacKenzie, 2015).

The PH condition is a similar setup to one condition tested in Kohli et al. (2012). They found a throughput value of ~ 6 bps for this condition, compared to our result of 4.25 bps, which was significantly lower. Their setup used a marked fingertip tracker instead of our marker-less approach, and the target activation occurred with low latency upon contacting their passive haptics. Both of these could be the reason for their higher throughput. Seixas et al. (2015) performed a similar condition to our NH condition, using a LeapMotion to perform a Fitts' circle. The major difference was the visual feedback was given on a computer monitor, and offset from the physical location of the hand. Their results found a throughput of 2.9 bps, much lower than our 3.76 bps. This performance improvement in our experiment is likely due to the additional immersion of having a head-mounted display and the colocation of the visual feedback with the hand movement.

Insko (2001) found that passive haptics provided a transfer of training. Their experiment found that subjects who performed the task with passive haptics before they performed it with no haptics could perform better than the control group who were never exposed to passive haptics. We did find that subjects who performed no haptics after passive haptics did better in throughput than those who started with no haptics. This is shown in Figure 4.8 comparing the NH results in PH First and NH First. However, a group that performed NH twice would provide a better control group as there may have been training unrelated to the haptics that caused better performance in the second condition. This interpretation relies on comparing performance from the second condition of one sequence group to the first condition of the other.

4.6 Conclusion

We found that the use of a passive haptic for a 2D targeting task caused a significant increase in Fitts' throughput. The subjects did not learn the task any quicker with passive haptics, but a positive transfer of training existed for subjects who used passive haptics and then later performed the same task without passive haptics. Upon investigating the trajectories of the movements, it was found that the passive haptics reduced the amount of time subjects spent in the corrective phase. Subjects reported lower arm fatigue when using the passive haptics, though only for the first condition performed. The Presence Questionnaire (PQ) score was marginally higher for the passive haptics condition. Subjects did overwhelmingly report that they preferred the passive haptics condition.

This experiment provided insight into the benefits of using the passive haptics. The results indicate that the use of passive haptics changes the targeting time and accuracy of reaching movments. Up to this point, our research has focused on the effect of the R3C system on fundamental measures of human performance (such as the time and accuracy). An unanswered question is how the human performs in the R3C system for a task which requires more cognitive workload than targeting a button. This will be investigated in the next experiment.

Chapter 5

Design Evaluation Experiment

5.1 Introduction

The final experiment combines the lessons of the previous experiment to investigate the use of the Rapidly Reconfigurable Research Cockpit (R3C) in a design evaluation study. The goal of this experiment is to determine if the R3C system can be used in the place of a more traditional evaluation tool. As previous chapters have discussed, there are a number of self-evident advantages to using the R3C system. However, there remain some technical limitations to the technology that could hinder adoption. We found that a button targeting task took more time in our virtual environment than in the real world (Chapter 3). The following experiment (Chapter 4) found that a Fitts' Law task produced a higher throughput using a passive haptics layer, mitigating some of the time increase of targeting buttons in a virtual environment. In the experiment described in this chapter, we used the R3C system as the simulation tool for a design evaluation study of a cockpit instrument. The purpose in undergoing this evaluation study is to understand if these limitations interfere with the metrics used in evaluating a new cockpit design.

We designed an experiment which asks subjects for feedback on the design of a cockpit instrument. The subjects were divided into two groups. One group used an R3C setup to operate the instrument, while the other used a more traditional touchscreen simulator of the instrument without a head-mounted display. This separation of groups will allow a comparison of the feedback from subjects between groups. Both groups evaluated the same two instrument designs, and subjects were asked to provide feedback using the same questionnaires. We hypothesize that the R3C system could be used in place of a traditional simulator if the two groups provide similar responses to the designs. Additionally, we utilized common quantitative metrics to evaluate performance in each group. The quantitative measures may change in magnitude between groups, but we need to determine if changes between designs are the same between groups.

5.2 Methods

For the mock design study we needed to define a task that the subjects could perform, be analogous to flight tasks, and could be presented in two different designs for the comparison. We chose to have subjects fly a one dimensional flight simulator using a joystick and use their other hand to input text on an instrument. This is analogous to a common cockpit task where a pilot may have to enter information on an instrument (e.g. navigational waypoints or autopilots commands) while maintaining a flight path. The difference in the instrument design will primarily affect the text input task, as will be described below in the description of the designs. However, the purpose of including the flight task is to provide an increase in workload to the subjects. This is in marked contrast to our previous experiments, which had subjects perform button pressing as the sole task.

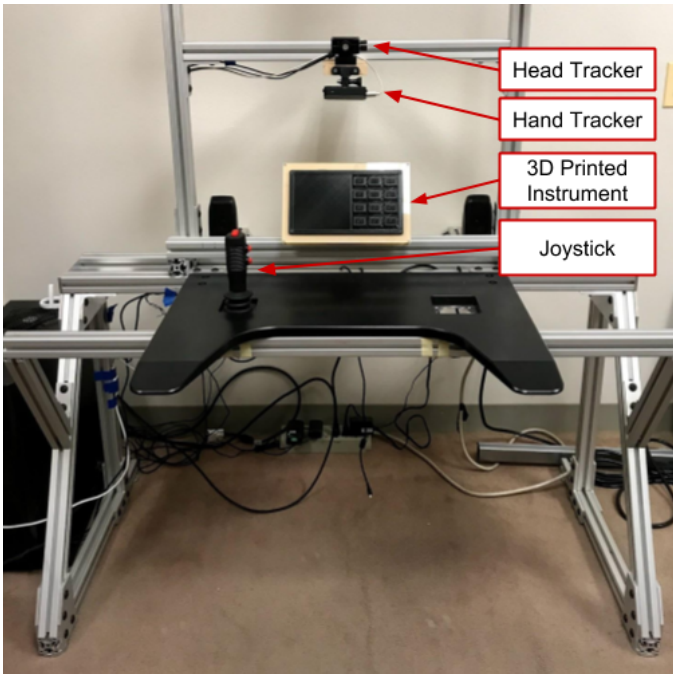
Subjects were divided into two groups based on which simulator setup they used: Touchscreen (TS) or Virtual Reality (VR). The TS group would use a touchscreen to input the text on the instruments, while the VR group used the R3C system (passive haptics, hand tracking and virtual reality). In this section, the hardware used and instrument designs that were developed are described. This is followed by a description of the task that the subjects performed.

5.2.1 Simulator Setup

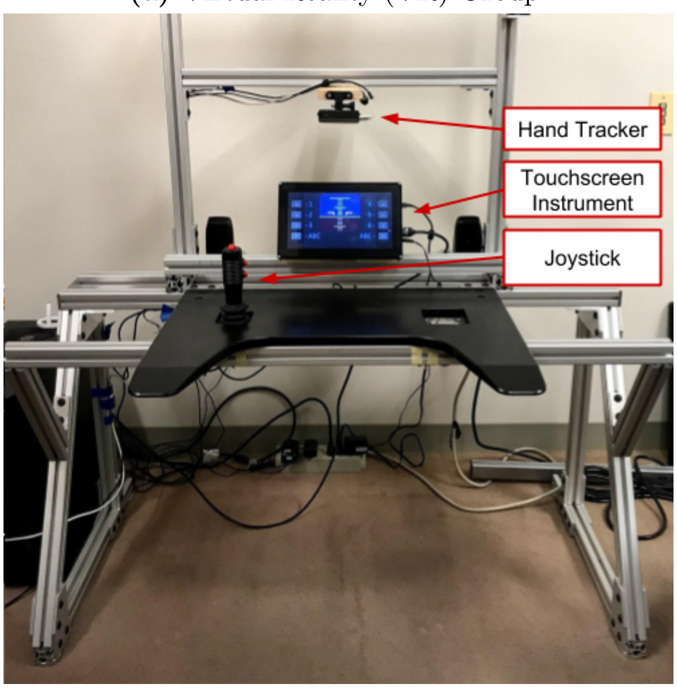
The simulator workstation as configured for each group is shown and annotated in Figure 5.1. Figure 5.2 shows a user in each group operating the simulator. The joystick is mounted on the left side of the desk, and subjects used their left hand to operate it. The instrument is mounted in front of the subject, at a height and position similar to an instrument in a cockpit. The buttons on the instrument were operated with the right hand of the subject. The hand tracker is mounted above the instrument looking down, a similar setup to the previous experiments. Both groups had an aural indication (a click noise of a button being pressed) when a button was pressed on the instrument, using speakers mounted behind the instrument panel. The joystick and instrument were positioned in the same location for each group.

The major physical difference between the two groups was the instrument. For the Virtual Reality (VR) group, 3D printed instruments were used to provide the passive haptics, while the input was read by the hand tracker. The Touchscreen (TS) group used a touchscreen which displayed the instrument buttons, mounted at the same location.

The VR group used the hand tracker activated system previously described in Chapter 2. For this experiment, the buttons were configured to turn a blue color when the hand tracker registered a finger within the

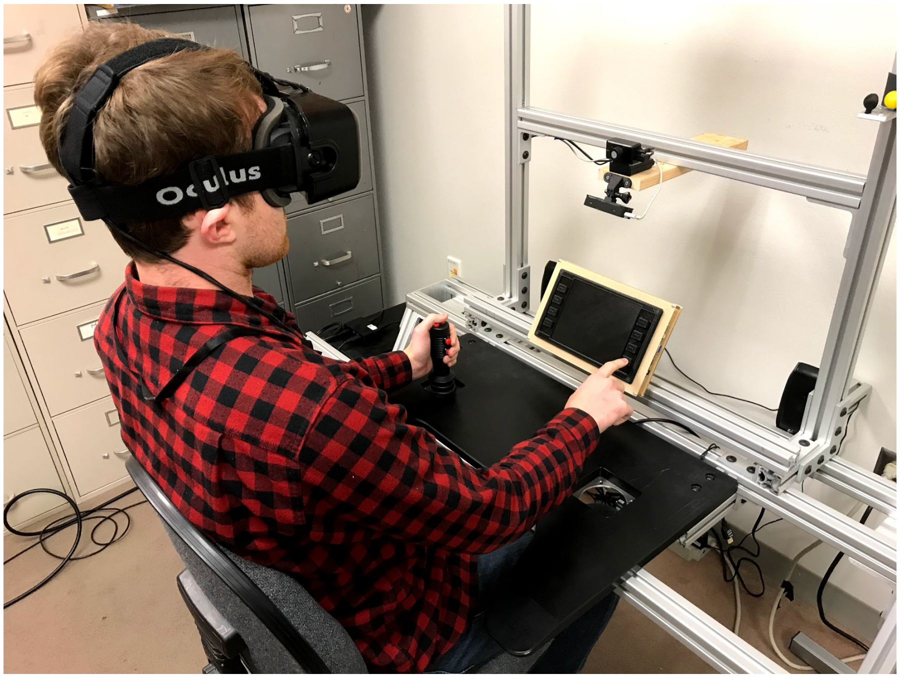


(a) Virtual Reality (VR) Group.

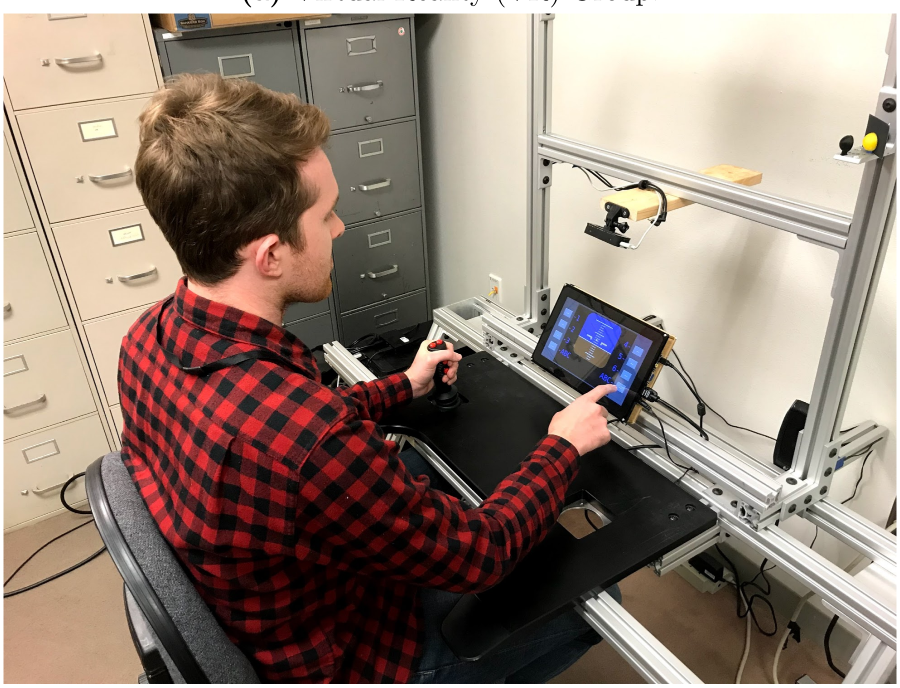


(b) Touchscreen (TS) Group.

Figure 5.1: Simulator configured for each group.



(a) Virtual Reality (VR) Group.



(b) Touchscreen (TS) Group.

Figure 5.2: Simulator being used in each group configuration.

zone. The zones were extended 0.1 in around the border of the button, raised a height of 0.5 in above the surface of the button. When the button was activated after the 150 millisecond delay, the blue highlight would disappear and the button in the virtual world would move inwards as if it were being pushed in[1], the press sound would play, and the response of the instrument associated with pressing that button would occur. A separate release sound would play when the finger left the zone after a successful press, and for the VR group the button would move back to its starting position.

The TS group used a 10.1 in capacitive touch screen with resolution of 1024x600. The active area of the screen was 8.8 in by 5.1 in, with outside dimensions of 10.4 in by 6.7 in. The instrument buttons were rendered in a web browser using standard HTML elements. Javascript press and release events were used to simulate the same behavior as described for the VR group, except for the highlighting before a button press.

5.2.2 Instrument Designs

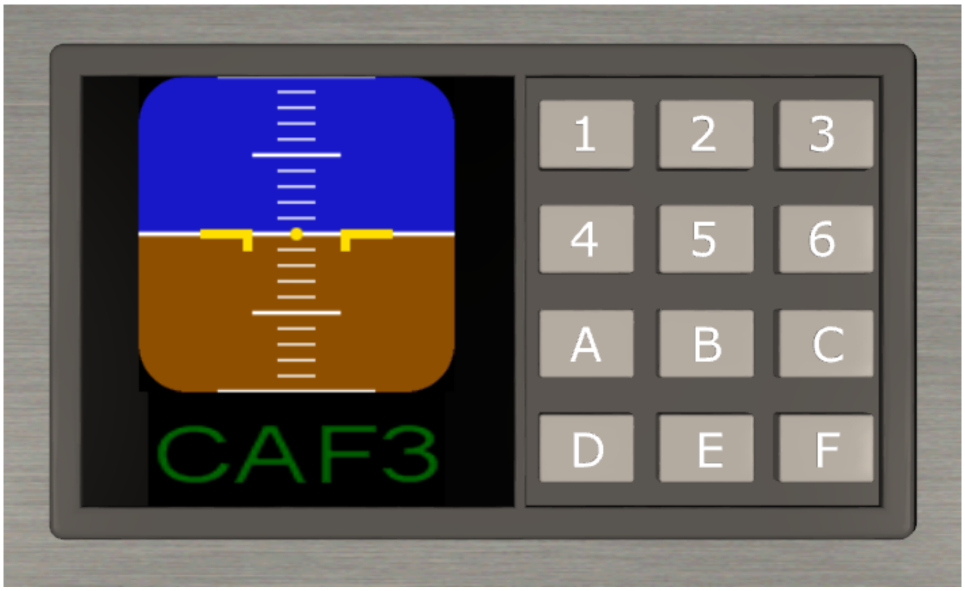
The two different designs used were developed to be both realistic as a cockpit instrument design that would be under consideration, yet still have one design with flaws that should be identified during a design evaluation.

[1] Of course, the physical button could not and did not move.

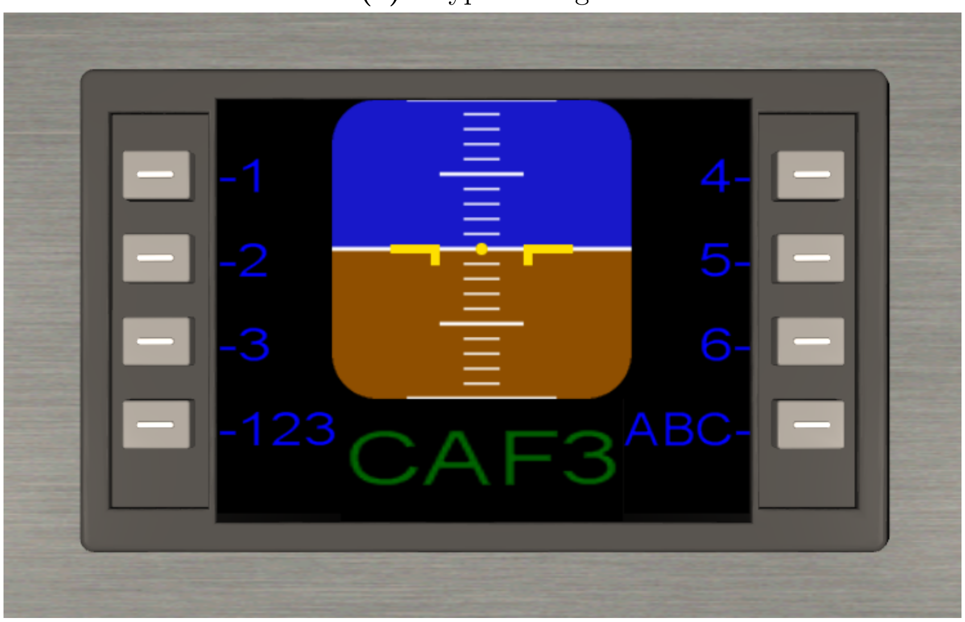
There were three major components of each design: the buttons for text input, a text display area, and an attitude indicator for the flight task. The major difference between our two designs is the location of the buttons and the logic of the text entry task. The 'Keypad' design (Figure 5.3a) has the button keys on the right side and the attitude indicator on the left. The 'Edgekey' design (Figure 5.3b) has the button keys split on either side of the attitude indicator. Both designs placed the text display area underneath the attitude indicator. The logic of the text entry task will be described in more detail in the next section, but to understand the two designs it is only necessary to know that the entries were limited to the letters 'A' through 'F' and the numbers '1' through '6'.

The two designs are pictured in three different versions developed for the experiment. Figure 5.3 shows the instruments as they were rendered in the virtual world for the VR group. The 3D printed versions that provided the passive haptics for the VR group are photographed in Figure 5.4. The TS group used a fullscreen web browser to render the instrument elements on the touchscreen. Screenshots of each design are shown in Figure 5.5.

The Keypad design included a button for each character that needed to be entered. The buttons are 1 in by 0.75 in, with about 0.26 in between buttons horizontally and 0.38 in vertically. Each button has the label directly on the top of the button. The 3D printed instrument used for the



(a) Keypad Design



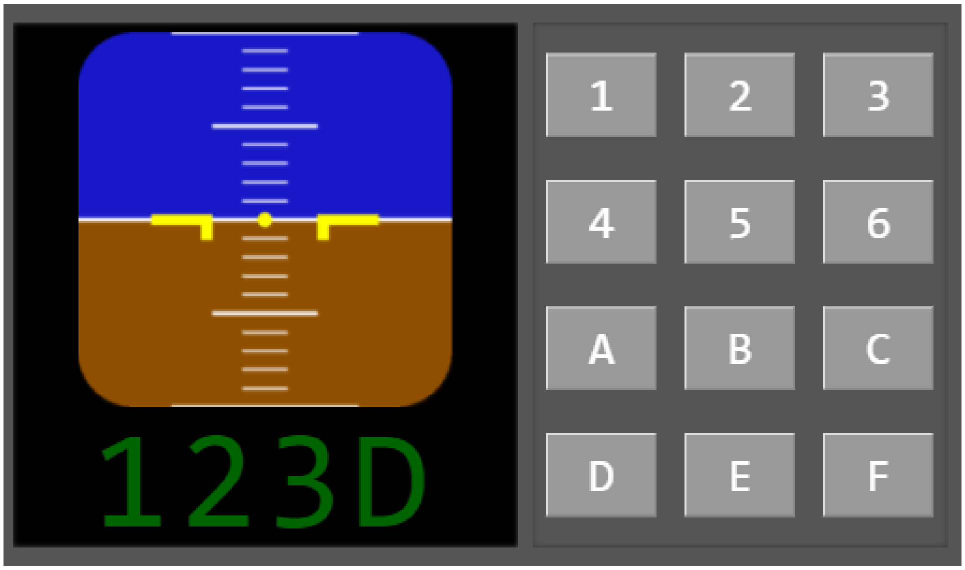
(b) Edgekey Design

Figure 5.3: The instrument designs as shown in the VR group virtual environment.

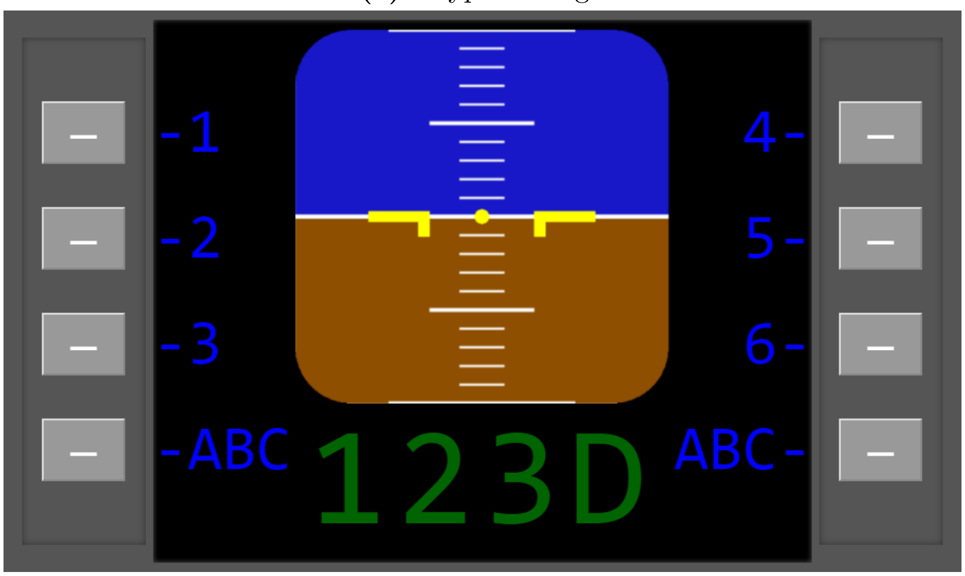
(a) Keypad Design

(b) Edgekey Design

Figure 5.4: The 3D printed versions of the instrument designs.



(a) Keypad Design



(b) Edgekey Design

Figure 5.5: The instrument designs as shown to the TS group on the touchscreen.

VR group had the buttons raised a height of 0.31 in from the surface of the instrument. The button labels were also raised to provide a tactile feedback. The font was approximately 0.36 in tall, and the labels were embossed above the button surface by 0.05 in.

In the Edgekey design, there is not a single button for every number and letter. Instead, the bottom button on either side would switch the behavior (and labels) of the remaining six buttons from being 1 through 6 to A through F. In other words, the bottom "switching" buttons would change the rest of the buttons from the numbers to the letters, and vice-versa. The labels were placed offset from the button on the "screen" portion of the instrument, allowing them to change dynamically. The fonts were approximately 0.32 in tall, and were blue on the screen. The buttons are smaller in this design, at 0.76 in by 0.55 in. A smaller button size was needed to fit the labels and the buttons side by side. The spacing between buttons vertically is the same as the Keypad design at 0.38 in. The center to center distance between the two sides of the button rows is 7.3 in. The 3D printed instrument version had raised nubs on each button covering half the width, 0.08 in tall and raised 0.05 in. As with the Keypad design, the buttons had the same height of 0.31 in from the surface.

5.2.3 Task Design

The subjects were performing two tasks simultaneously: a flight task with their left hand and a text entry task with their right hand. The text entry task required subjects to read and memorize a short string of characters and enter it back using the buttons on the instrument. We call the text they read and enter the ‘prompt’, and thus the text entry task is called the ‘prompting task’. For the flight task, the subjects were asked to fly in pitch only using an attitude indicator. Their task was to keep the pitch at zero (level flight) while disturbances were introduced. This is a standard compensatory tracking task. The flight task is called the ‘tracking task’. These two tasks are described in detail in this section.

Prompting Task

The prompting task was designed to be both a realistic task for a cockpit as well as a demanding task when done concurrently with the tracking task. The task required the subjects to read and memorize a short string of characters and enter it back using the buttons on the instrument. To limit the task physically (by number of buttons) and mentally, the characters used were the numbers 1 through 6 and the letters A through F. The prompts were 4 characters long. Once the subject started entry, the prompt would disappear, forcing them to hold it in working memory. It has been well established that humans are capable of holding about seven digits in working memory (Miller, 1956; Baddeley, 1992).

The sequence of the prompts was separated into 10 second "windows". The prompt would appear randomly between 2 and 3 seconds of the start of the window. From the time of appearance, subjects were given seven seconds until timeout. When the subject pressed the first button of the prompt, the prompt itself was cleared and asterisk symbols ($*$) were shown in place of the prompt for each button entry by the subject. If the subject ran out of time, the text in entry area would return to black. Although subjects were briefed on the timeout and given practice to learn the pace, no warning or indication of time left was shown during the trials. Whether they completed the prompt within the time limit, or they timed-out, this process was repeated every 10 seconds.

The prompts themselves were always composed of three numbers followed by a letter or three letters followed by a number. This structure was decided upon to provide a consistent pattern, yet still utilize both letters and numbers in every prompt. The prompts were randomly chosen but were not allowed to have repeat numbers or letters. The selection of letters or numbers as the first three characters was randomly chosen as well, with an equal weight to each.

The Edgekey design had fewer buttons, but the logic of the buttons generated higher workload demands. While some of the more subtle differences were expected to be noted by the evaluation study (e.g. having smaller buttons, different position of the flight task), the major flaw designed into the Edgekey design was the requirement that the subject uses the switching key to change from letters to numbers and back. This additional action fundamentally changed the demands of the prompting task, as the subjects now had to press this additional button to change labels at least once per prompt. The cognitive workflow of the subject is diagrammed in Figure 5.6. The additional mental effort of the Edgekey design is shown in the dashed box, where the subject has to verify the state of the instrument and possibly press the switch button before they press the buttons of the prompt. Since the prompts were kept as a consistent format of three of the same type and fourth of the other, this extra work was easily skipped for most buttons, and anticipated between the third and fourth button. There was no guarantee that the next prompt would start with the instrument in the correct state for the new prompt, so there was always an additional cognitive load in determining whether a switch was necessary at the beginning of the prompting window, which would be accompanied with the physical effort if the switch was needed.

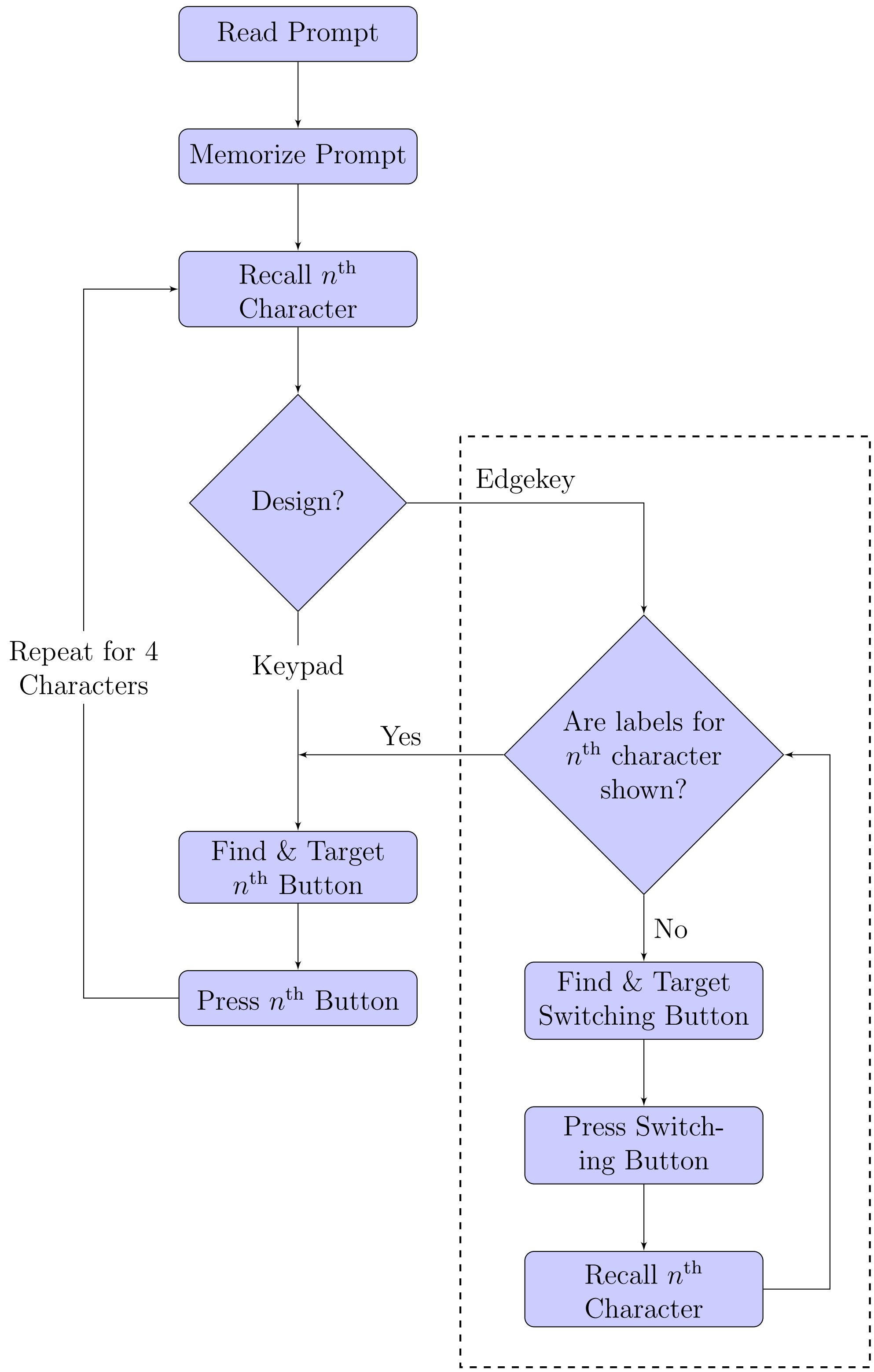


Figure 5.6: Prompting Task Flowchart of Cognitive Work for Each Design. Extra work of Edgekey design enclosed in dashed line box.

Tracking Task

The tracking task display was a standard attitude indicator display, as can be seen in Figure 5.3. Each tick corresponds to 1 degree in the dynamics simulation, with major ticks at intervals of 5 degrees. The attitude indicator was rendered to the size of 3.4 in square on the instrument. Subjects controlled the one-dimensional (pitch only) task using a joystick with their left hand. The joystick is pictured in Figure 5.1.

The block diagram of the dynamics is shown in Figure 5.7. The dynamics model was updated and recorded at a rate of 125 Hz. The output of the joystick, r_{js}, varies from -1.0 to 1.0, and the gain of $10°$ was chosen to ensure the pilot had enough control authority to complete the task. The flight dynamics model (Aircraft Dynamics) of the simulator was a stability derivative based model for a Boeing 747 from NASA CR-2144 (Heffley and Jewell, 1972), listed as "Flight Condition 2". The model was linearized from sea-level flight at an airspeed of 335 ft/s. The configuration of the airplane had the gear up, no flaps, total weight of 654 000 lbs, and an angle of attack of 7.3°. This dynamics model was chosen due to availability and the specifics of the dynamical model were not important other than providing a response similar to an aircraft in flight. The transfer function of the aircraft dynamics is given as:

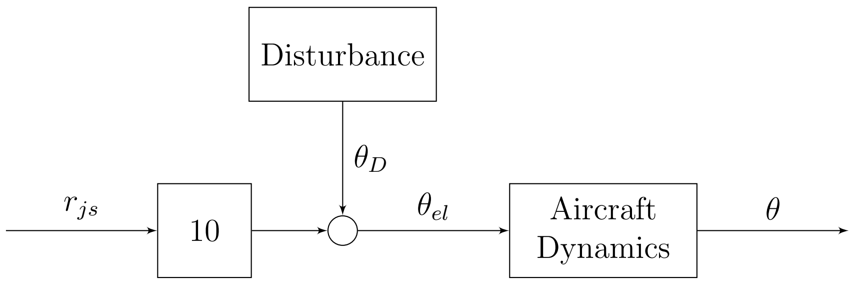


Figure 5.7: Tracking Task Dynamics Block Diagram. r_{js} is the joystick input. θ_{el} is the elevator input. θ_D is the disturbance. θ is the pitch of the aircraft.

$$\frac{\theta}{\theta_{el}} = \frac{-0.572(s+0.553)(s+0.0396)}{(s^2+2\zeta_1\omega_1+{\omega_1}^2)(s^2+2\zeta_2\omega_2+{\omega_2}^2)} \tag{5.1}$$

$$\omega_1 = 0.0578, \quad \zeta_1 = 0.0160$$

$$\omega_2 = 1.12, \quad \zeta_2 = 0.798$$

The disturbance model is from Sweet and Trejo (1999). It is designed to provide a broad spectrum of frequencies for human pilot frequency response identifications (McRuer and Krendel, 1974). The disturbance is a sum of sines described by:

$$\theta_D = K \sum_{i=1}^{12} \left[a_i \left(\frac{2\pi k_i}{240} \right) \sin \left(\frac{2\pi k_i}{240} t + \phi_i \right) \right] \tag{5.2}$$

The k_i terms are given as:

$$k_1 = 7, \quad k_2 = 11, \quad k_3 = 16, \quad k_4 = 25, \quad k_5 = 38, \quad k_6 = 61$$
$$k_7 = 103, \quad k_8 = 131, \quad k_9 = 151, \quad k_{10} = 181, \quad k_{11} = 313, \quad k_{12} = 523$$

The amplitude terms are $a_i = 0.5$ for $i <= 6$ and $a_i = 0.05$ otherwise. The phase terms, ϕ_i, were randomly selected on the $(-\pi, \pi)$ interval ensuring a uniform distribution (Sweet and Trejo, 1999). This random selection was precalculated for each trial, however the order was repeated for each subject so there was no between subject variance in the disturbance signal. Furthermore, each subject received the same sequence of disturbance signals for each instrument design. The disturbance amplitude, K, was chosen such that the root-mean square (RMS) of the signal was 3.5 degrees. The value of the RMS was chosen through pilot studies to ensure the task was challenging but not overwhelming.

5.2.4 Experiment Design

Subjects were divided into the two groups, Touchscreen (TS) and Virtual Reality (VR). The overall sequence of the experiment started with a training session on the simulator and the task, followed by an evaluation session for each of the two designs, finishing with questionnaires asking subjects to evaluate the two designs. The timeline of the experiment was the same for each subject, except for counterbalancing the order that the designs were evaluated. The training portion started with a slide deck explaining the tasks, the simulator that the subject was using (depending on which group they were in), and the functionality of the two designs. Next, they performed practice trials with just the tracking task and then just the prompting task. The practice trials of the tracking task were 60 seconds long and repeated until the subjects' performance had flat-lined. This took between three to six trials for each subject.

For the evaluation sessions with each design, they performed six trials with both tasks. The first three were a minute long, and were considered practice trials, and not included in the data analysis, though this was not communicated to the subjects. The following three trials were two minutes each, and were used for the analysis. Each evaluation session concluded with a two minute trial of just the tracking task without the prompting task. This was included to investigate if the subject had improved or fatigued at the tracking task throughout the experiment.

Subjects spent between one to two hours in the experiment (typically longer for the VR group due to additional setup). The training and setup took approximately fifteen to thirty minutes, and each of the two evaluation sessions took about fifteen to twenty minutes to complete. The remainder of the time was spent on the subjects completing the questionnaires.

The independent variables of the experiment are Group and Design. The Group is the simulator the subject used, a between subjects factor, and either TS or VR. The Design is a within subjects factor, the two instrument designs that every subject evaluated — Edgekey and Keypad.

5.2.5 Dependent Measures

The dependent measures were chosen to evaluate the performance of each task individually as well as the workload of the subject. For the tracking task, the root-mean square error (RMSE) was calculated for each trial (Harris, 2011). The subjects performed a compensatory tracking task, attempting to keep the attitude indicator at zero degrees while disturbances were introduced. This means that the error is the pitch shown on the attitude indicator, since the goal at all times was zero pitch.

The prompting task has two dependent measures, for speed and accuracy. For speed we consider the *response time*, defined as the time between when the prompt is first shown to the subject and when they press the first button of their response entry. The accuracy is measured by how many prompts they complete correctly. Twelve prompts are shown to the subject within each trial. The response time did not include trials which used the switch button of the Edgekey design, or trials where the first button pressed was incorrect. The response time was averaged per trial first and then per design for each subject, and the number of correct prompts is averaged per design for each subject.

A NASA Task Load Index (TLX) survey was administered after they completed each design to measure the workload of the subject. The TLX survey records a rating of workload between 0-100 for the following subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. Our implementation allowed selection of the ratings within increments of 5, and included anchors of "Low" and "High" at the extrema of 0 and 100, respectively (except for Performance, which uses "Good" and "Bad"). The midpoint was also visually indicated with a larger tick. The ranked pairs modification was used and completed for both times the subject took the survey. This modification asks the subject, for each of the pairwise combinations of subscales, which they felt contributed more to their workload. The number of times they select each subscale is used a weight to calculate a weighted mean for the total TLX score.

Finally, the subjects were given a questionnaire asking for their feedback on each instrument design. For each design, the subjects were asked the following questions:

- Please comment on any difficulties you had performing the prompting task with this design especially in contrast to the other design.

- Please comment on anything you liked in this design.
- Please comment on anything you did not like in this design.
- Any other comments?

Additionally, the following questions were asked:

- Which instrument design did you prefer? Why?
- Did you experience any physical fatigue during the experiment? Where?
- Any other comments?

An open form text box was used for the response field for each of these questions.

In a standard design evaluation study, the feedback received from the users in this questionnaire (and other debriefing interviews) would often be the main source for carrying out re-design. The purpose of this feedback in this experiment is to determine and document in which ways does this feedback differ between groups. For example, if most subjects in one group noted issues with the size of a button, while no one in the other group found an issue with that button, this would indicate that using this VR system may not highlight the same issues regarding button sizes. The groups were purposely left ambiguous in the example, as it does not matter which group found the flaw and which group did not comment on it. Although we could postulate as to which group are "correct" in their evaluation of the instrument, it is not a useful exercise, as the only result is to document what potential differences could arise so that users of this system can be aware.

With that goal in mind, the analysis of the feedback questions seeks to identify differences between the groups. The sentences from the open form responses were first separated into single feedback comments, and summarized using common language. If a single subject repeated the same comment in the answers to multiple questions, they were only counted once. Each of these simplified feedback comments were assigned to a category or overall summary of their feedback. This process was completed separately for each group. We aim to look for feedback that is unique to a certain group or feedback that receives a higher frequency of comments in one group. This will provide a summary of where the groups provide the same feedback and where they provide differing feedback.

5.2.6 Hypotheses

The main hypothesis of this experiment is that the use of a VR/R3C simulator will not affect the conclusions of a design evaluation study, compared to a traditional touchscreen simulator. We do expect that some of the dependent measures may have a significant difference in Group (one Group will perform a task better overall, no matter the Design) or a significant difference in Design (subjects will perform better with one Design, no matter the Group). The more important measure for us, however, is the interaction effect. This will test if the change between Designs is similar for the two Groups. If this is the case, then it may indicate that an evaluation study using one of these simulators could draw differing conclusions of an evaluation study using the other. Statistically, we will test the hypothesis that there exists no interaction effect between Group and Design for any of our dependent measures.

Additionally, the two tracking only trials performed at the end of each evaluation session, as well as the final tracking only training trial, will be used to investigate if the subjects were still learning the tracking task. The concern if subjects became more trained in the tracking task is that it could lower their amount of attention to that portion of the task, causing a change in performance on the prompting task that was not due to the design change. These hypotheses are enumerated here:

H1. The tracking task RMSE will have no interaction effect between Group and Design

H2. The prompt response time will have no interaction effect between Group and Design

H3. The percent of correct prompts will have no interaction effect between Group and Design

H4. The NASA-TLX scores will have no interaction effect between Group and Design

H5. The tracking task RMSE for the last training trial and the tracking only trials will not change throughout the experiment

5.2.7 Statistical Tests

The quantitative dependent measures are tested with a two-way ANOVA, with one within subjects factor (Design) and one between subjects factor (Group). The Design factor contains two levels, the two designs each subject tested, Edgekey and Keypad. The Group factor also contains two levels, the VR (Virtual Reality) group and the TS (Touchscreen) group. When the ANOVA showed significance in the interaction test, post-hoc repeated measured t-tests were undertaken to determine the significance of Design within each Group. Independent samples t-tests were used to test the significance of Group within each Design. The last hypothesis testing the effects of learning on the trials with only the tracking task will be tested with a two-way ANOVA, with the Group as a between subjects factor, and the trial number as a within subjects factor. The trial number is chronological in the order the subjects performed them. The first trial was the last tracking only training trial, and the next two were tracking only trials at the end of each design evaluation.

Statistical significance level was corrected using the Bonferroni correction considering the 5 hypotheses being tested. All effects were considered statistically significant at the 0.01 level ($\alpha = 0.05/5 = 0.01$). Effects which have a significance level between $0.05 < p < 0.01$ are considered to be marginally significant.

5.3 Results

5.3.1 Demographics

Twenty-three subjects were recruited from the UC Davis engineering undergraduate and graduate student population. Twelve subjects were placed in the VR group, and the remaining eleven in the TS group. The mean age was 21.0 ($\sigma = 3.14$), with 19 male and 4 female subjects. The genders were balanced between the two groups. Two of the subjects were left-handed, with one in each group. Most subjects had no flight experience (two were student pilots), and all of the VR group subjects indicated that

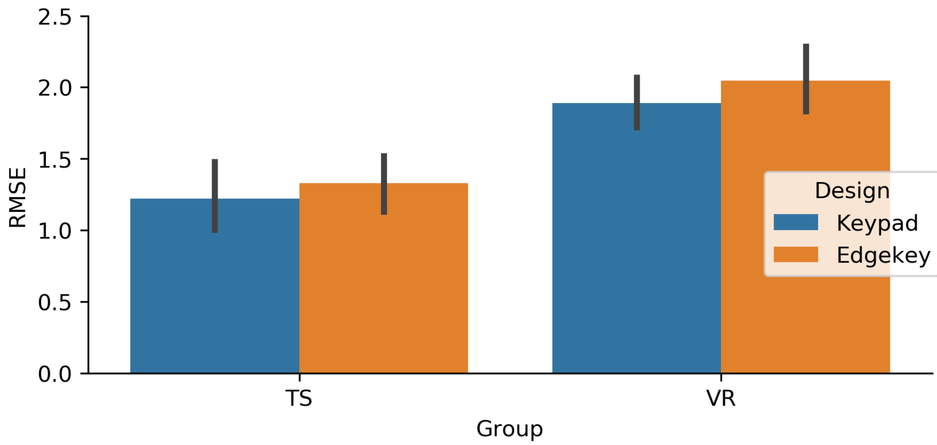


Figure 5.8: RMSE (root mean square error) of the tracking task by Group and Design.

they had less than one hour of experience using virtual reality headsets. It should be noted that the subjects are not the beneficial population of the research. The task and experiment was designed with this in mind and mitigated through training and the simplicity of the task design.

5.3.2 Performance Measures

Tracking Task RMSE

The performance of subjects performing the tracking task was measured using the root-mean square error (RMSE) of the pitch. The RMSE of each Group and Design is shown in Figure 5.8. The effect of Group yielded an F ratio of $F(1, 21) = 21.4, p < 0.001$ indicating a significant difference between VR ($M = 1.28°, \sigma = 0.38°$) and TS ($M = 1.97°, \sigma = 0.38°$).

In both groups, subjects were performing the tracking task using their left hand to control the same joystick. There are two potential factors that could contribute to the decreased performance in the tracking task for the VR group. The most direct factor is the loss of visual acuity in the tracking task display (attitude indicator) due to the resolution of the VR head-mounted display. Indirectly, the additional workload of the prompting task for the VR group could be taking attention away from the tracking task. The effect of Design indicated a marginally significant difference ($F(1, 21) = 5.94, p = 0.024$) for the tracking task RMSE between Keypad ($M = 1.57°, \sigma = 0.51°$) and Edgekey ($M = 1.70°, \sigma = 0.52°$). The only change in the tracking task display between the two instrument designs is a small change in position. It moves from being on the left side for the Keypad to the middle for the Edgekey. Since there was no change otherwise, this suggests that any difference on the tracking task performance between the designs would be related to additional workload from the prompting task.

The main hypothesis (H1. The tracking task will have no interaction effect between Group and Design) was supported by the statistical test. The interaction effect was not significant ($F(1, 21) = 0.17, p = 0.69$). This indicates that even though the VR group had a higher RMSE overall, the difference between designs was consistent between the groups. An evaluation session performed with only the VR system would find the same difference between designs with the RMSE measure as a traditional simulator.

If the visual acuity was the only factor affecting performance between groups, we would expect to see a similar effect for the trials where the subjects were only performing the tracking task. The subjects ran a single trial that was just the tracking task without the prompting task at the end of each evaluation session. The RMSE results for these tracking only trials are shown in Figure 5.9 The effect of group on RMSE for the tracking-only trials yielded a marginally significant difference ($F(1, 21) = 4.81, p = 0.039$) between the VR Group ($M = 1.32°, \sigma = 0.50°$) and the TS Group ($M = 0.91°, \sigma = 0.43°$). There was no significant difference for the effect of Design ($F(1, 21) = 0.068, p = 0.80$). The interaction effect between Group and Design was also not significant ($F(1, 21) = 3.21, p = 0.087$). This indicates that when the subjects were focused on the single task, they were able to mitigate most of the visual resolution differences between using a touchscreen and the virtual reality screen.

Although the tracking only trials found a marginally significant difference for the group, the difference was much more distinct for the trials with both tasks. Additionally, the marginally significant difference between the designs for the trials with both tasks was reduced to no significance when

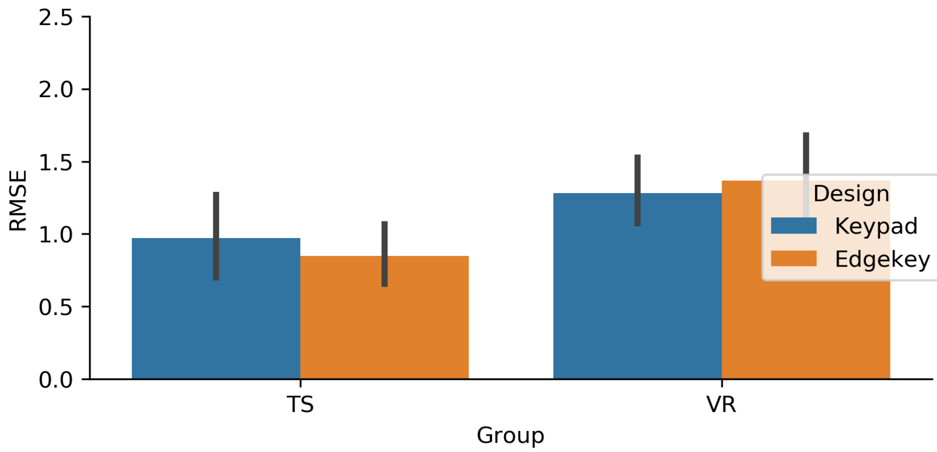


Figure 5.9: RMSE of tracking task for tracking only trials by Group and Design.

the additional prompting task was removed. This also points to the additional workload of the prompting task causing a performance drop on the tracking task. The factors leading to the added workload of the prompting task are investigated in the next performance measures discussed.

Prompt Response Time

The first measure of the prompting task is the response time of the subject. The response time is defined as the time from the prompt is shown to each subject until they press the first button of the prompt. The response time of a prompt was not included if the subject did not press the correct letter or number for the first button (6.5% of prompts). However, the correctness of their response after the first button was not considered. For the Edgekey design, it would be possible that the subject had to start

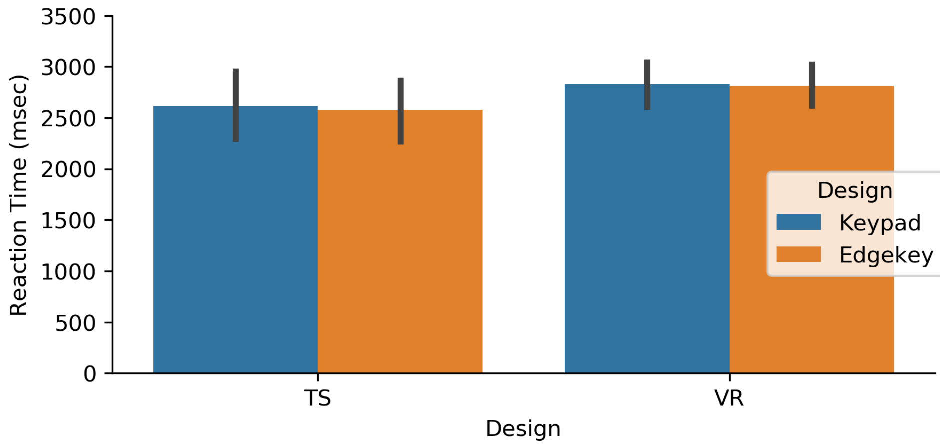


Figure 5.10: Prompt Response Time (milliseconds) by Group and Design.

with the switching button if the new prompt did not start with the same mode (letters or numbers) as the previous prompt (see Figure 5.6). Since this button would not clear the prompt when it was pressed, it is not considered the first button of their entry. However, this would still require an additional movement of the subject, adding additional time. For this reason, the prompts which required the subject to start with the switch key are filtered out of this analysis. After filtering, 885 of the total 1700 prompts recorded for the Edgekey design were kept.

The response time by Group and Design is shown in Figure 5.10. The response time was unique among the dependent measures, as all tests were insignificant. The effect of group yielded an F ratio of $F(1, 21) = 1.19, p = 0.29$ indicating no significant difference between VR ($M = 2812\,\text{ms}, \sigma = 383\,\text{ms}$) and TS ($M = 2594\,\text{ms}, \sigma = 567\,\text{ms}$). One factor that could influence the response time between groups is the additional time to activate a button in the VR environment versus the touchscreen. The touchscreen subjects were using a familiar interface for activating the buttons, while the VR subjects needed to activate the button with the virtual hand. We previously saw the difference in performance between targeting buttons in the real world and virtual world in the first experiment (Chapter 3). However, a large portion of the response time for the subject is their cognitive processing of the prompt – recognizing the new prompt has appeared, reading it, then memorizing it. Beyond potential differences in the visual environment, the cognitive portion should not take more time for one group or the other. A potential reason that there could be a lower than expected difference between the group means is that it was observed during the experiment that some VR subjects learned to keep their hand extended closer to the instrument so that the hand tracker could keep it in view. When the hand tracker lost view of the hand, the re-acquisition time could be significant, so holding it close to the instrument would prevent this from happening. This issue comes up again when looking at the subjects' response to questions about fatigue.

The effect of design was also insignificant ($F(1, 21) = 0.68, p = 0.42$) between Keypad ($M = 2728\,\text{ms}, \sigma = 512\,\text{ms}$) and Edgekey ($M = 2687\,\text{ms}, \sigma = 471\,\text{ms}$). The biggest difference between the two designs is the switching key on the Edgekey design. As described above, the need for an additional switch press before the first prompt button was filtered out, so we are only comparing prompts where the first button was available right away to the subject. Even though the physical requirements were filtered out, subjects still need to verify that the labels are in the correct state for starting entry. Since the Edgekey design had more time pressure due to the need for the switch key, subjects could have learned to respond quicker to adapt for this. However, these differences in the design did not appear to have a significant effect on the response time.

Finally, the interaction effect was not significant ($F(1,21) = 0.001, p = 0.96$), supporting the hypothesis (H2. The prompt response time will have no interaction effect between Group and Design). Unlike with the RMSE, however, we did not see a difference between groups or designs. This indicates that for a task that includes a large cognitive workload, the performance measures may provide the same result with a VR system than would be expected with a traditional simulator.

Prompt Completion Accuracy

The second measure of the prompting task is the accuracy of the subjects in correctly completing the prompt. The prompt task consists of two sequential components: memory and execution. First, they must re-

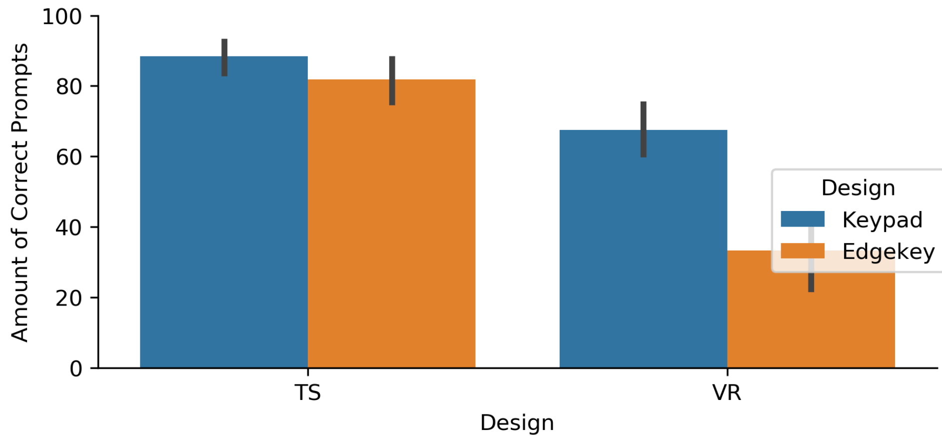


Figure 5.11: Percent of Correct Prompts per Trial by Group and Design.

member the prompt as they enter it, and second, they must be able to physically press the buttons within the seven second response window. For the statistical test, we are using the percent of how many prompts each subject completed successfully per trial. Among the incorrect prompts, we can differentiate between whether the subject entered the prompt incorrectly (failure to remember the prompt) or whether the subject ran out of time (failure to physically press the buttons). These are reported to help analyze the results, but are not used in the statistical tests. There were 12 prompts per trial, and every subject completed three trials for each design.

The percent of correct prompts are reported on here, shown for each Group and Design in Figure 5.11. The percent of correct prompts had a significant interaction effect between group and design ($F(1, 21) = 27.8, p < 0.001$), meaning the main effects must be interpreted with the post-hoc tests as well. This is the first dependent measure that rejects the null hypothesis (H3. The percent of correct prompts will have no interaction effect between Group and Design). Both main effects were significant, the effect of group yielded an F ratio of $F(1, 21) = 43.9, p < 0.001$ while the effect of design yielded an F ratio of $F(1, 21) = 64.1, p < 0.001$.

For the effect of design on the VR group, the repeated measured t-test indicated a significant difference ($t(11) = 8.0, p < 0.001$) between the Keypad ($M = 67.6\%, \sigma = 13.47\%$) and the Edgekey ($M = 33.3\%, \sigma = 19.8\%$) The TS group had a marginally significant difference ($t(10) = 2.28, p = 0.045$) between Keypad ($M = 88.4\%, \sigma = 8.0\%$) and the Edgekey ($M = 81.6\%, \sigma = 11.7\%$). These results indicate that both groups had trouble with the additional time pressure caused by the Edgekey design requiring the use of the switch key. The TS group performed closer to their performance in the Keypad design, however, with only about 1 fewer prompts correct on average (6.8%). The VR group had much more difficulty in the Edgekey design, correctly completing about half as many as they completed in the Keypad design. They also had more difficulty in both designs compared to the TS group.

The post-hoc test for differences between groups within each design also found that for both designs the VR group had more difficulty. These tests had significant effects for both the Keypad design ($t(21) = 4.44, p <$

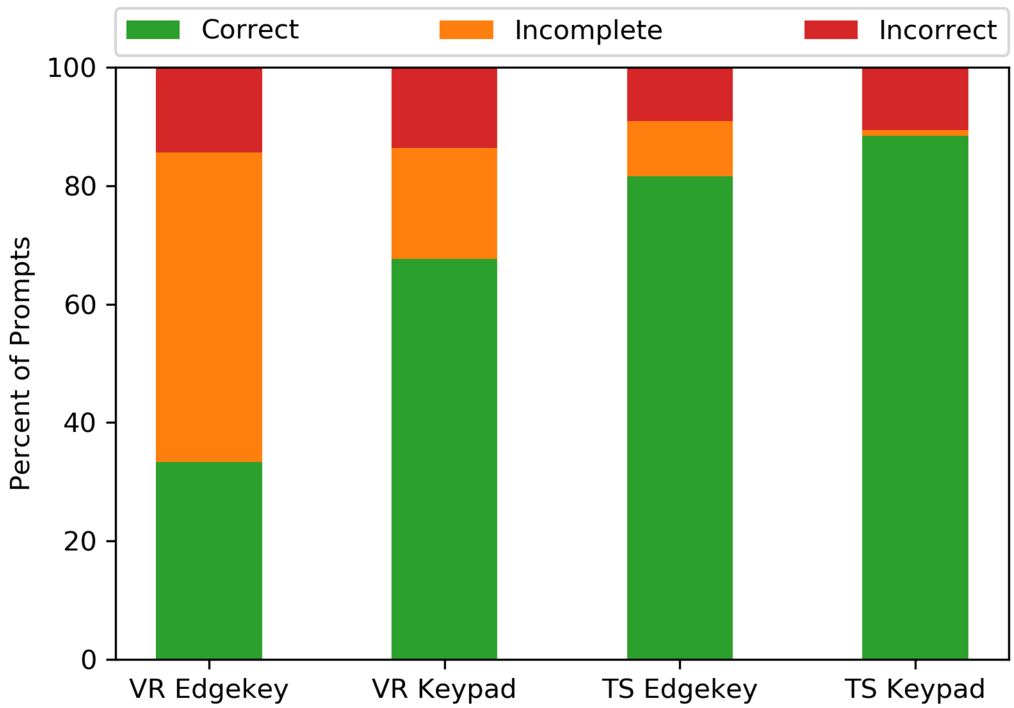


Figure 5.12: Average percent of correct, incorrect, and incomplete prompts by Group and Design.

0.001) between the VR group and the TS group, and the Edgekey design ($t(21) = 7.05, p < 0.001$) between the VR group and the TS group. The main effect of group clearly has a meaningful effect, which found the VR group ($M = 50.5\%, \sigma = 24.1\%$) had significantly fewer correct prompts than the TS group ($M = 85.0\%, \sigma = 10.4\%$). This difference is largely due to subjects not being able to complete the prompt. Figure 5.12 shows the breakdown of the mean result of each trial for each group and design.

Across all groups and designs, between one and two prompts on average were incorrectly completed in each trial (between 9.4% and 14.4%). Most of the difference in number completed correctly is due to the incomplete prompts. A major contributing factor for this could be the method of button activation used for the VR group combined with the time pressure. Another contribution could be the limitations of the hand tracker. When the hand tracker lost tracking or gave bad information, it became hard or impossible for the subject to activate a button until the hand tracker returned to normal. When this happened in the middle of a prompt, the amount of time it took to recover from the bad tracking would lead to a timeout on the prompt entry, causing an incomplete prompt. The variance of number correct was also much larger in in the VR group, which could be caused by some subjects adapting to the unfamiliar VR environment more rapidly.

For this dependent measure, the null interaction effect hypothesis was rejected. We expect that this was caused by the time pressure and the additional work required to activate the buttons in the VR system. This was not completely unexpected as we have seen in previous experiments time penalty of using the hand tracker for pressing buttons in the VR system. An evaluation study which requires time pressured button entry may not get the same results using a VR system due to this limitation. However, in our case the same effect was found between designs within

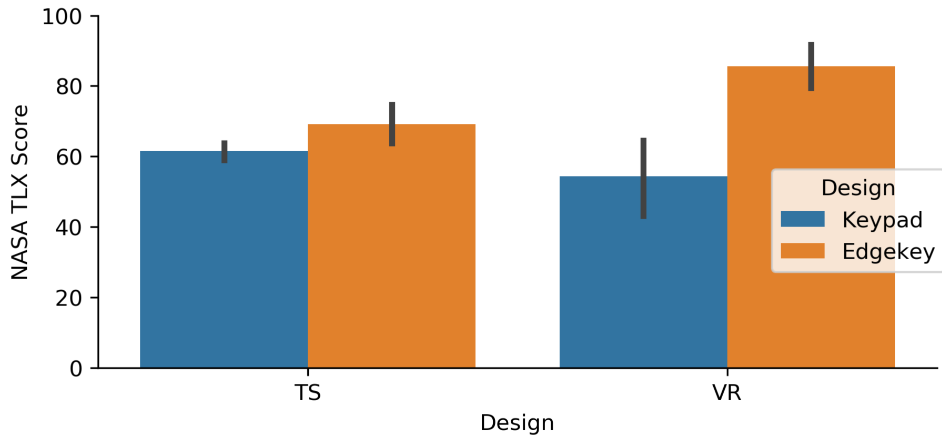


Figure 5.13: NASA-TLX workload scores by Group and Design.

each group. The interaction effect existed because the VR group found the Edgekey design more challenging compared to the Keypad than the TS group did.

NASA-TLX

After the subject completed their trials for each design, they filled out a NASA-TLX workload survey. Their scores, weighted means by the pairwise comparisons, are used here as a measure of their self-reported workload. The factor plot is shown in Figure 5.13. The interaction effect between group and design was found to be significant ($F(1, 21) = 8.25, p < 0.001$). The main effects showed a significant difference in design ($F(1, 21) = 23.6, p < 0.001$), but not in group ($F(1, 21) = 1.69, p = 0.21$). This could mean that the group did not affect the TLX score, but in the presence of an interaction effect, the post-hoc tests guide the interpretation.

The repeated measures t-tests indicated significance between designs for the VR group ($t(11) = -4.20, p = 0.001$) between the Keypad design ($M = 54.4, \sigma = 20.4$) and the Edgekey ($M = 85.6, \sigma = 11.2$) There was a marginally significant difference between designs for the TS group ($t(10) = -2.72, p = 0.02$) between the Keypad design ($M = 61.5, \sigma = 4.46$) and the Edgekey ($M = 69.2, \sigma = 10.1$). The effect of design was much stronger in the VR group, but both groups indicated respectively higher workload on the TLX scores for the Edgekey design. This follows from the experimental design which predicted that the Edgekey design would be more difficult. One factor that could have contributed to a larger difference in scores for the VR group could be the increased difficulty subjects had in completing the prompt, as seen in the results of the number of incorrect and incomplete prompts for the VR group using the Edgekey design (Figure 5.12). The effect of group was not shown to be significant in the ANOVA analysis, but the independent samples t-test showed a significance for the Edgekey design ($t(21) = 3.69, p < 0.01$) between the VR Group ($M = 85.6, \sigma = 11.2$) and the TS Group ($M = 69.2, \sigma = 10.1$). With the Keypad design, The effect of group was not significant ($t(21) = -1.13, p = 0.27$) between VR ($M = 54.4, \sigma = 20.4$) and TS ($M = 61.5, \sigma = 4.46$). These tests further illustrate that the VR group found a higher workload for the Edgekey design specifically, as both groups rated the workload in the Keypad design similarly.

The subject perceived workload scores rejected the null interaction hypothesis (H4. The NASA-TLX scores will have no interaction effect between Group and Design). In this case, the effects changed between the groups. The VR group reported significantly more workload for the Edgekey design, while the TS group reported only a marginal increase. We expect that this change was likely due to the increased work required to complete the prompt, the same reason for the change in the prompt accuracy measure which also rejected the null hypothesis.

Tracking Task Learning

Throughout the experiment the subjects performed trials with only the tracking task, instead of both the tracking task and the prompting task. Initially, they performed a number of training trials at the beginning with only the tracking task, and then after each evaluation session there was a single trial of just the tracking task. In this section we will test the RMSE of their final training trial and the two after-evaluation trials for any significant learning effects. The trial number is chronological throughout the timeline of the experiment for each subject. Since the visual environment of the tracking task was quite different for each group, the Group factor

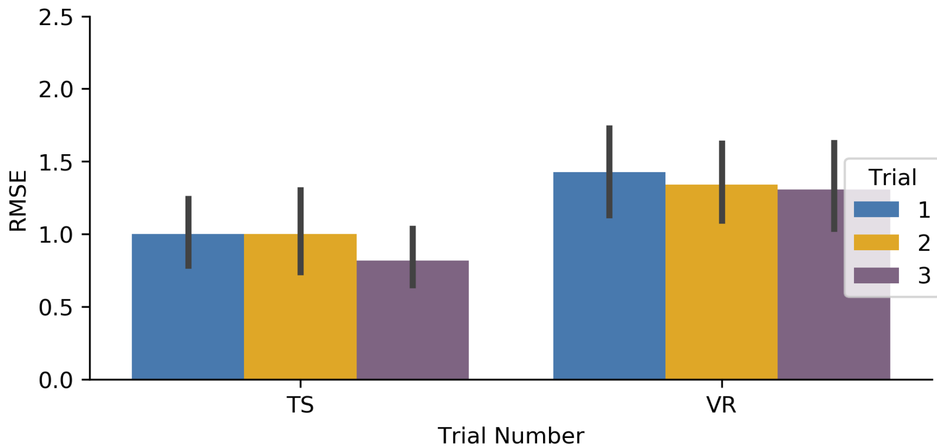


Figure 5.14: RMSE for Tracking Only Trials by Group and Trial.

is included as a between subjects factor. The RMSE of the tracking only trials, by Group and trial number, is shown in Figure 5.14.

The two-way ANOVA found Group to be a marginally significant factor ($F(1, 21) = 4.94, p = 0.037$). The TS Group ($M = 0.94°, \sigma = 0.43°$) had an RMSE 0.42° lower than the VR Group ($M = 1.36°, \sigma = 0.51°$). The effect of group on the tracking task was already established, so the marginal significance found here is not unexpected. Trial number was found to have no significant effect ($F(1, 21) = 3.65, p = 0.069$) between the three trials. The means of the three trials, in order, are 1.23° ($\sigma = 0.54°$), 1.18° ($\sigma = 0.51°$), and 1.07° ($\sigma = 0.51°$). Even though the statistical test indicates no significance, the means do decrease as trial number increases. This combined with the large variance suggest that some subjects were potentially experiencing some training effects, but overall the effect of training is not statistically significant. The interaction effect of Group and trial number had no significance ($F(1, 21) = 0.16, p = 0.69$).

The final hypothesis (H5. The tracking task RMSE for the last training trial and the tracking only trials will not change throughout the experiment) was supported by these results. We found that the subjects did not experience significant training in the tracking task throughout the experiment.

Summary

A summary of the significance results from the ANOVA and post-hoc t-tests for all the performance measures are shown in Table 5.1. The significance is indicated by '$*$' for $p < 0.01$, '+' for $0.01 < p < 0.05$, and '-' for no significance. For the measures with significant interaction effect, the post-hoc t-tests are shown per group and per design. We found no interaction effect with the tracking task RMSE and prompting task Response Time, supporting the first two hypotheses from Section 5.2.6 (H1 and H2). The prompting task accuracy (Correct Prompts) and the workload (NASA-TLX) both rejected the no interaction effect hypotheses (H3 and H4). We attribute the interaction effects found in those two dependent measures to be related to the increase in difficulty to activate buttons using the hand tracker. The results will be discussed further (Section 5.4) after

	ANOVA			
	Group	Design	Group:Design (Interaction)	Hypothesis
Tracking RMSE	*	+	-	H1 Supported
Response Time	-	-	-	H2 Supported
Correct Prompts	*	*	*	H3 Rejected
NASA-TLX	-	*	*	H4 Rejected
	t-tests			
	Design		Group	
	VR Group	TS Group	Keypad Design	Edgekey Design
Prompts Correct	*	+	*	*
NASA-TLX	*	+	-	*

Table 5.1: Statistical Significance Test Results. '*' indicates significance at the $p < 0.01$ level, '+' indicates marginally significant ($0.01 < p < 0.05$), and '-' indicates no significance. Hypotheses are from Section 5.2.6, which stated that the dependent measure has no interaction effect.

reporting on the results of the feedback questions.

5.3.3 Design Feedback

As discussed in Section subsection 5.2.5, the long-form feedback questions were synthesized and summarized into categories. The categories and the counts of comment occurrence for each group is summarized in Table 5.2. Categories which only received one comment are not included in this table in interest of brevity, but the full table is shown in Appendix A.26.

Topic	Feedback Summary Category	VR Group	TS Group
Edgekey	Switch Difficult	14	12
Keypad	Familiar	6	11
Edgekey	Centered Flight Task Better	3	13
Keypad	Buttons Proximal	6	7
Keypad	Buttons Always Visible	5	5
Other	Hand Tracking Issues	9	0
Edgekey	Hand Blocks View	3	4
Fatigue	Prompting Arm	4	1
Edgekey	Clean Design	3	2
Fatigue	Fatigue from Joystick	0	4
Edgekey	Easier	0	4
Keypad	Buttons Confusable	0	4
Other	Colors Disliked	2	2
Fatigue	Eye Fatigue	3	0
Keypad	Easy Focus Switch	2	1
Keypad	More Mistakes	1	2
Edgekey	Accuracy Worse	1	2
Keypad	Buttons Bad Layout	2	0

Table 5.2: Counts of Design Feedback Comments per Group. Sorted by sum of comments.

By far the issue that received the most feedback was the difficulty of using the switch key (Edgekey, Switch Difficult). Most of the complaints stated the extra difficultly of having to press another button. Some of the other complaints from this category were: it took extra time (with no extra time given), it added to the mental demands of the task, and it was difficult to see which mode the instrument was in. Both groups disliked the switch key, and mentioned it just as frequently.

> "Switching from numbers to letters was hard, especially if I was trying to compensate for turbulence and was struggling at the time." (TS Subject)

> "I did not like how much extra work it was. It took so much extra focus that I forgot I was flying with the joystick" (VR Subject)

Many subjects noted the familiarity of the Keypad design (Keypad, Familiar) and that having the buttons close together (Keypad, Buttons Proximal) as things they like about that design. The familiarity was noted more often for the TS Group, but both were some of the more frequent comments within each group.

One comment about the Edgekey design that got more frequent mentions from the TS Group was that they found having the flight task in the middle of the display, centered between the buttons, was preferred (Edgekey, Centered Flight Task Better). The subjects who chose the Edgekey as their preferred design nearly unanimously cited this as their reason for their preference[2]. The comments that fed into this category also included subjects who noted the difficulty of splitting their focus back and forth with the Keypad design. Interestingly, two of the TS Group subjects noted that they would have found the Keypad easier if they had tactile feedback to guide their input. This could suggest that the reason the VR

[2] The one holdout did not explain why they preferred the Edgekey design.

Group subjects did not find the centered flight task advantageous is because with the tactile feedback of the 3D printed instruments they were able to keep visual focus on the left half of the screen in the Keypad design, thus not seeing benefit from the centering of the flight task display.

> "[The Edgekey design] forced me to pay more attention to what I was typing, this wouldn't have been a problem if the keypad was a physical device that allowed me to locate the numbers and letters without looking, much like the dots on a computer keyboard." (TS Subject)

> "I like that the flight control was cent[e]red, so you could see it even when you were looking at the buttons." (VR Subject)

The most notable exceptions to providing similar feedback between groups are the categories that relate to fatigue issues. Four subjects in the TS group noted fatigue caused from using the joystick, yet none in the VR group did, despite using the same joystick setup, and seated in the same location. The VR group did note more fatigue in their other arm that was used for the prompting task. This fatigue seemed to be caused by the additional effort needed to have the hand tracker recognize the hand. For example, one subject wrote:

> "My right wrist was somewhat fatigued. Though I think this is mostly from positioning my hand for the simulator to recognize my input." (VR Subject)

Some of this additional effort was due to subjects learning to hold their prompting task hand "hovering" while waiting for the next prompt. This was done to keep the hand in view of the hand tracker as when the hand leaves the field of view, the re-acquisition will slow down the entry of first button. Many subjects organically learned this, and kept their arm in front of the instrument between prompts.

Similar to the fatigue issues being different, there were some comments that were due to the technology being used more-so than the designs themselves. Obviously, the subjects who noted difficulty using the hand tracker, or the one subject who mentioned touchscreen issues, are specific to the simulator technology the used. However, some of the other categories had comments that may have been indirectly caused by the different technologies and their limitations. For example, some subjects noted the keypad design caused them to make more mistakes. For the TS Group, this was due to the touchscreen being too responsive to the button presses:

> "[S]ince I was able to go more quickly with this layout, I had more mistakes in the entry." (TS Subject)

One subject in VR who complained of more mistakes in the Keypad design identified a common problem caused by the hand tracker. When the hand tracker was having registration issues it would sometimes mistakenly place the other fingers in the activation zone of the buttons underneath the one being targeted, causing multiple buttons to be pressed in a short period of time.

> "There's more unintended register since other fingers might trigger the buttons." (VR Subject)

Although only one subject noted this, it was observed happening to many subjects. In fact, for the VR group, eight of the twelve subjects had the wrong button register within 200 milliseconds of the last button in the Keypad design. In the other designs and groups this happened to only one or two subjects.

5.4 Discussion

Our results suggest that tasks or performance measures which are dominated by a cognitive portion, such as the prompt response time, provide similar results. Tasks which rely on visual resolution or time pressured responses may not produce the same results between designs using the R3C system. None of the effects reversed slope between designs, however, and the only change is in magnitude of the effect. In fact, for both the number of prompts correct and the workload ratings, which had significant interaction effects, the use of the VR system amplified the effect of design within the groups from a marginally significant effect to a significant effect.

Many design evaluation studies would be concluded with both paper questionnaires as well as open interviews to receive the feedback from the subject. Our experimental design avoided the use of the interview for two reasons. First, since our subjects were not subject domain experts or experienced evaluators, we wanted to ensure that the prompting of the questions were consistent. Second, the primary goal of the design feedback for this experiment was not to evaluate the designs, but rather to compare evaluations. The use of a proctor interviewing the subjects could introduce accidental bias into the responses of the subjects. This can often be useful when evaluating a new interface, for example, an interviewer could ask subjects about a flaw they had not mentioned yet to determine if they did not notice it or did not care about it. However, in our case, we omitted this additional information to ensure no bias was introduced in the collection of their opinions.

Our initial motivation is to provide a virtual reality simulator to be used with an existing mockup. In this experiment, we compared the performance of a touchscreen simulation to the use of our R3C system. A touchscreen simulation is a higher fidelity simulation than what may exist during the mockup phase. Our goal in this experiment was to show that the R3C system could provide the same feedback from a higher fidelity simulation. This provides motivation for using the R3C system earlier in the design process (during the mockup phase). Our results indicate that the VR/R3C system does provide the same feedback from evaluations, and dependent measures will transfer favorably except for time-pressured dependent measures.

This was a limited study of the utility of VR/R3C for design evaluation purposes. The task and instrument design was kept simple in nature for this study in order to limit the amount of confounding variables as well as keep it easy to learn for the subject population. Future studies could investigate this system in a more involved design study, with multiple instruments or designs, or more complex behavior in the cockpit. At this point, it would become more essential to use subject domain experts (i.e. experienced pilots) in order to validate these results.

5.5 Conclusion

The motivation of this experiment was to determine the differences between using an R3C simulator system and a traditional simulator system to perform a design evaluation experiment. We had two groups of subjects perform the same evaluation task on two different designs of a cockpit instrument, one group using the R3C system and the other a touchscreen system. The evaluation task included two simultaneous tasks: a pitch disturbance tracking task with their left hand and a call and response prompting task with their right hand. In addition to the quantitative performance measures of the task, subjects were asked for their feedback on the two designs at the conclusion of the experiment.

The results are summarized using their two independent variables: Group and Design. Group, a between subjects factor, refers to the technology the subject used: either Virtual Reality/R3C (VR) or Touchscreen (TS). Design is a within subjects factor, and is the instrument design the subject was evaluation: Edgekey or Keypad.

The VR Group had worse performance than the TS Group with the RMSE of the tracking task. Subjects from both groups had a marginally significant difference in tracking task performance due to Design, with subjects performing better with the Keypad design. It was also shown that, on control trials that had only the tracking task (no prompting task), the effect of Group was reduced to marginally significant. The response time of the prompting task had no significant effect based on Group nor Design. Neither the RMSE nor the response time had interaction effects between Group and Design. This indicates that the differences in these two dependent measures would be found whether the evaluation was performed using the VR/R3C system or the touchscreen system.

The percent of correct prompts and NASA-TLX workload score did have a significant interaction effect between Group and Design. The TS Group was able to complete significantly more prompts correctly overall than the VR Group, but the VR Group had a significant effect with Design and the TS Group only had a marginally significant effect. The TLX scores for the VR Group showed a significant effect in Design, with subjects rating the Edgekey design over 30 points higher than the Keypad design (averages of 54.4 to 85.6, respectively). However, like the percent of correct prompts, the TLX score was found to be only marginally significant for the TS group, rating the Keypad at 61.5 compared to the Edgekey at 69.2. We attribute the rejection of the null interaction hypotheses to the increased workload of activating buttons in the R3C system coupled with the time pressures of the prompting task.

The results of the subjective feedback analysis found that there was no omission of major feedback items on the design of the two instruments from either group. The only feedback comments that did not transfer were the fatigue issues, and technology-specific issues. We did discover that some issues were mentioned at differing frequencies, which is to say, one group would have more subjects mention it than the other. These results suggest that the use of the R3C system for receiving feedback from a design would be appropriate.

Chapter 6

Conclusion

6.1 Prototype Development

The creation of the Rapidly Reconfigurable Research Cockpit prototype represented a large amount of integration work. The use of the hand tracker in the virtual environment was a novel use at the beginning of the project, but now LeapMotion (the company behind the hand tracker) has changed focus to develop for the specific use case of hand tracking in virtual reality. The capacitive touch sensors provided a useful countermeasure to the initially inaccurate registration between the virtual world and real world. As the technology matured, and with the development of the calibration algorithm, the original goal of non-functioning geometric mockups was realized. The development of the calibration mechanism proved to be an important step in the development process, as an accurate registration between the physical components and the virtual-visual world provided a much more convincing simulation to new and experienced users alike.

6.2 Experimental Findings

The first experiment (Chapter 3: Pointing Experiment) provided an initial study into targeting performance in various fidelities of the prototype technical approach. Eight subjects were tasked with targeting buttons on a four-button keypad, repeating the same task for four conditions. The conditions varied three independent variables: the use of virtual reality (VR and real world), haptics (passive haptics and no haptics), and button detection method (hand tracker and capacitive touch). The effect of each independent variable was determined on the time, accuracy, and success rate of targeting a button. It was found that the subjects could perform the targeting task almost twice as fast in the real world (1.57 seconds to 3.15 seconds) than in the virtual reality environment. The center of pressure of the touch was within error between conditions, but the virtual reality condition had a slightly wider distribution. There was a significant effect on the button detection method, with the capacitive touch taking an average of 3.15 seconds to the hand tracker method taking an average of 3.78 seconds. There was no effect due to the passive haptics. In all conditions, the selection task was performed nearly without error (98.4% correct) and no difference was found between conditions.

The second experiment (Chapter 4: Passive Haptics Experiment) performed a Fitts' Law analysis of the technology, and characterized the differences between using the passive haptics and not using any haptic feedback. Twenty subjects performed a Fitts' circle ISO-9241-9 task using a virtual reality HMD and activating the button targets with a hand tracker. The subjects repeated the task for a variety of indices of difficulty and also for two haptic conditions: passive haptics and no haptics. It was found that the Fitts' throughput was significantly higher when using the passive haptics (4.25 bps) compared to no haptics (3.76 bps). Subjects spent significantly more time in the corrective phase of their trajectory without passive haptics, indicating that the passive haptics helps with the final phases of targeting movements. Self-reported arm fatigue was also found to be lower for subjects who completed the first condition with passive haptics, though scores converged for the second conditions. The presence survey found a marginally significant increase in presence using passive haptics. Subjects did strongly indicate their preference for using the passive haptics, despite the small difference found in the presence survey results.

The third experiment (Chapter 5: Design Evaluation Experiment) had subjects perform a mock design evaluation using one of two simulators: an R3C based virtual reality simulator or a touchscreen based simulator. The 23 subjects were assigned one simulator technology and then evaluated two designs of an instrument, performing a flight task and working memory prompting task using the instrument. The two instrument designs were the same for each group, and subjects gave their subjective feedback after they went through a series of tasks on each design. The instrument designs placed buttons for the prompting task response in two different layouts, an "Edgekey" design and a "Keypad" design. The results were analyzed for interaction effects between group and design for the dependent measures. This would highlight the possible differences for an evaluator if the virtual reality (VR) group evaluation led to a different outcome than the touchscreen (TS) group. The subjective feedback was also categorized and compared to see how the two groups differed in response.

There was no significant interaction for the root mean square error (RMSE) of the tracking task. The VR group had worse tracking performance (1.97° vs. 1.28° for the TS group), but the differences between instrument design were the same. The prompting task response time had no effect for design or group, but the amount of correct prompts between designs was significantly different for each group. The VR group had significantly fewer correct prompts with the Edgekey design (44.0%) than with the Keypad design (67.5%). The TS group had only a marginally significant difference between designs, with 88.3% and 81.8% in the Keypad and Edgekey designs, respectively. Similarly, the workload scores (NASA-TLX) had a significant difference in both designs, but the effect was stronger in the VR group. The Edgekey design had a higher workload score than Keypad in both groups.

Finally, the subjects were asked for the opinions on each design. The responses from the subjective feedback questions found that both groups were able to identify the same major categories, for both positive and negative comments on the designs. Some issues were not noted as much in one group or the other, but major issues were noted by both groups. For example, 13 comments were made in the TS group noting that the flight task being centered in the Edgekey design was helpful, but only 3 subjects noted this in the VR group. However, both groups noted at about the same frequency that the switch key in the Edgekey design was troublesome (14 comments in VR, 12 in TS). Overall the feedback received is encouraging that the R3C system would highlight the same usability issues in a design being considered.

6.3 Research Questions

In Chapter 1.3, we listed seven research questions that we aimed to answer by this research. Here we summarize our answers to these questions.

Can the technical approach of a mockup providing passive haptics with a virtual-visual overlay using a head-mounted display be used successfully? Yes, it can be used successfully, and this answer is supported by all three experiments, but most notably the first and last experiment. The first experiment (Chapter 3) found that subjects could select buttons in the R3C system just as successfully as in the real world. The third experiment (Chapter 5) had subjects successfully use the R3C system in a more realistic flight task, using 3D printed instruments.

Can a user select a button in the R3C prototype with the passive haptics and hand tracker as quickly and accurately as one in the real world? Yes, they can select a button as accurately, but not as quickly. In the first experiment (Chapter 3) it was shown that the subjects could target a button as accurately as in the real world, but they took more time to accomplish this.

Does the use of passive haptics change the performance of subjects targeting a button in the virtual environment? Yes, subjects were able to target buttons with a higher throughput with the passive haptics than without (Chapter 4).

Does the use of passive haptics increase the presence of subjects using the virtual environment? In Chapter 4 we found that the presence was marginally higher with the use of passive haptics.

Does the use of passive haptics change the trajectory formation of subjects targeting a button in the virtual environment? The trajectory phase analysis in Chapter 4 found that the use of passive haptics had the largest change to the corrective phase of the trajectory. Subjects spent more time in the corrective phase without the passive haptics.

Can the R3C prototype be used effectively as a design evaluation tool? The third experiment (Chapter 5) performed a design evaluation of two mock instrument designs. Subjects were able to learn and use the R3C system to test the instrument designs and provide feedback.

How does the R3C prototype compare to other design evaluation tools? In the third experiment (Chapter 5) a design evaluation study was performed with both a R3C simulator and a touchscreen simulator. We found that some performance measures (tracking task performance and response time to a memorization task) provided similar results in both simulators. Tasks which were time pressured provided some difficulties to the R3C group. The most promising result was that the subjective feedback from both simulator groups provided the same conclusions about the designs being evaluated.

6.4 Future Work

Early in the research, it was decided to focus to only allow for push button interaction based on the technology level of the hand tracker and development time required for more interactions. An extension of this work could address other types of interactions, such as dials or toggle switches. Similarly, the button detection for the push buttons use a naive yet effective algorithm for determining when the user is pushing on a button. The collision detection model could be enhanced to predict when a user is actually moving in to push a button or if they are accidentally moving through the collision volume.

The capacitive touch sensors were removed as a need for our prototype as the hand tracker data and calibration improved. Recently research has been performed that allows for capacitive sensors to be printed directly into a 3D printed object (Shemelya et al., 2013; Kwok et al., 2017). As this technology matures, it could be used to rapidly prototype instrument designs with a highly-reliable method of button detection.

Future virtual reality and augmented reality technologies will also allow for further innovations on the R3C method. Currently, the most promising augmented reality headset is the Microsoft HoloLens, though there are a number of limitations. Most notably, the field of view is drastically smaller than our current setup (about 40 degrees) and the hand tracker we used can not work with the HoloLens. The use of augmented reality instead of a fully virtual setup would provide a drastically different experience as the visuals would not be as immersive. Newer virtual reality headsets are already available that provide small upgrades to the display quality and optics. The most important technical upgrade for the current R3C system would be a more robust hand tracker, as many of the problems experienced were related to the hand tracking reliability.

Bibliography

A. F. Abate, M. Guida, P. Leoncini, M. Nappi, and S. Ricciardi. A haptic-based approach to virtual training for aerospace industry. *Multimodal Interaction through Haptic Feedback*, 20(5):318–325, Oct. 2009. ISSN 1045-926X. doi: 10.1016/j.jvlc.2009.07.003.

J. Accot and S. Zhai. Refining Fitts' law models for bivariate pointing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 193–200. ACM, 2003.

S. Aromaa and K. Väänänen. Suitability of virtual prototypes to support human factors/ergonomics evaluation during the design. *Applied Ergonomics*, 56:11–18, Sept. 2016. ISSN 00036870. doi: 10.1016/j.apergo.2016.02.015.

T. Aslandere, D. Dreyer, and F. Pankratz. Virtual hand-button interaction in a generic virtual reality flight simulator. In *2015 IEEE Aerospace Conference*, pages 1–8, Mar. 2015. ISBN 1095-323X. doi: 10.1109/AERO.2015.7118876.

A. Baddeley. Working memory. *Science*, 255(5044):556–559, 1992. ISSN 0036-8075.

X. Bi, Y. Li, and S. Zhai. FFitts law: modeling finger touch with fitts' law. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1363–1372. ACM, 2013.

M. Bordegoni and G. Caruso. Mixed reality distributed platform for collaborative design review of automotive interiors: This paper presents how mixed reality technologies allow a closer collaboration among designers, final users and engineers and hence reduce the time for reviewing and validating car interior designs. *Virtual and Physical Pro-*

totyping, 7(4):243–259, Dec. 2012. ISSN 1745-2759, 1745-2767. doi: 10.1080/17452759.2012.721605.

G. Borg. *Borg's perceived exertion and pain scales.* Human Kinetics, Champaign, IL, US, 1998. ISBN 0-88011-623-4 (Paperback).

C. W. Borst and R. A. Volz. Evaluation of a haptic mixed reality system for interactions with a virtual control panel. *Presence: Teleoperators and Virtual Environments*, 14(6):677–696, 2005.

L. Bouguila, M. Ishii, and M. Sato. Effect of coupling haptics and stereopsis on depth perception in virtual environment. pages 54–62, 2000.

F. P. Brooks Jr. What's real about virtual reality? *Computer Graphics and Applications, IEEE*, 19(6):16–27, 1999.

G. Bruder, F. Steinicke, and W. Sturzlinger. To touch or not to touch?: comparing 2d touch and 3d mid-air interaction on stereoscopic tabletop surfaces. In *Proceedings of the 1st symposium on Spatial user interaction*, pages 9–16. ACM, 2013.

G. C. Burdea and P. Coiffet. *Virtual reality technology*, volume 1. John Wiley & Sons, 2003.

S. Card, W. K. English, and B. J. Burr. Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT. *Ergonomics*, 21(8):601–613, Aug. 1978. ISSN 0014-0139. doi: 10.1080/00140137808931762.

Y. Cha and R. Myung. Extended Fitts' law for 3d pointing tasks using 3d target arrangements. *International Journal of Industrial Ergonomics*, 43 (4):350–355, July 2013. ISSN 01698141. doi: 10.1016/j.ergon.2013.05.005.

S. Choi, K. Jung, and S. D. Noh. Virtual reality applications in manufacturing industries: Past research, present findings, and future directions. *Concurrent Engineering*, 23(1):40–63, Mar. 2015. ISSN 1063-293X, 1531-2003. doi: 10.1177/1063293X14568814.

K. Chun, B. Verplank, F. Barbagli, and K. Salisbury. Evaluating haptics and 3d stereo displays using Fitts' law. In *Haptic, Audio and Visual Environments and Their Applications, 2004. HAVE 2004. Proceedings. The 3rd IEEE International Workshop on*, pages 53–58. IEEE, 2004.

J. C. Coelho and F. J. Verbeek. Pointing Task Evaluation of Leap Motion Controller in 3d Virtual Environment. In *Creating the Difference, Proceedings of the ChiSparks 2014 Conference, The Hague, The Netherlands*, pages 78–85, 2014.

E. R. F. W. Crossman. The speed and accuracy of simple hand movements. *The nature and acquisition of industrial skills*, 1957.

E. R. F. W. Crossman and P. J. Goodeve. Feedback control of hand-movement and Fitts' law. *The Quarterly Journal of Experimental Psychology Section A*, 35(2):251–278, May 1983. ISSN 0272-4987. doi: 10.1080/14640748308402133.

H. Q. Dinh, N. Walker, L. F. Hodges, C. Song, and A. Kobayashi. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, pages 222–228, Mar. 1999. ISBN 1087-8270. doi: 10.1109/VR.1999.756955.

H. Drewes. Only one Fitts' law formula please! In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 2813–2822. ACM, 2010.

D. Elliott, G. Binsted, and M. Heath. The control of goal-directed limb movements: Correcting errors in the trajectory. *Human Movement Science*, 18(2–3):121–136, June 1999. ISSN 0167-9457. doi: 10.1016/S0167-9457(99)00004-4.

P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology: General*, 47(3):262, 1954.

S. Grant, T. Aitchison, E. Henderson, J. Christie, S. Zare, J. Mc Murray, and H. Dargie. A Comparison of the Reproducibility and the Sensitivity to Change of Visual Analogue Scales, Borg Scales, and Likert Scales in Normal Subjects During Submaximal Exercise. *Chest*, 116(5):1208–1217, Nov. 1999. ISSN 0012-3692. doi: 10.1378/chest.116.5.1208.

T. Grossman and R. Balakrishnan. Pointing at trivariate targets in 3d environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 447–454. ACM, 2004.

P. A. Hancock, D. A. Vincenzi, J. A. Wise, and M. Mouloua. *Human Factors in Simulation and Training*. CRC Press, 2008. ISBN 978-1-4200-7284-6.

D. Harris. *Human performance on the flight deck*. Ashgate Publishing, Ltd., 2011. ISBN 1-4094-2338-7.

S. G. Hart. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, Oct. 2006. ISSN 1541-9312. doi: 10.1177/154193120605000909.

S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988. ISBN 0166-4115.

R. K. Heffley and W. F. Jewell. Aircraft handling qualities data. Technical Report CR-2144, NASA, 1972.

J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. pages 1063–1072. ACM Press, 2014. ISBN 978-1-4503-2473-1. doi: 10.1145/2556288.2557130.

E. R. Hoffmann. Which Version/Variation of Fitts' Law? A Critique of Information-Theory Models. *Journal of Motor Behavior*, 45(3):205–215, May 2013. ISSN 0022-2895, 1940-1027. doi: 10.1080/00222895.2013.778815.

B. E. Insko. *Passive haptics significantly enhances virtual environments*. PhD thesis, University of North Carolina at Chapel Hill, 2001.

International Organization for Standardization. ISO 9241-9:2000, Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices. Technical report, 2000.

R. D. Joyce and S. Robinson. Passive Haptics to Enhance Virtual Reality Simulations. In *AIAA Modeling and Simulation Technologies Conference*, AIAA SciTech Forum. American Institute of Aeronautics and Astronautics, Jan. 2017. DOI: 10.2514/6.2017-1313.

R. D. Joyce and S. K. Robinson. The Rapidly Reconfigurable Research Cockpit. American Institute of Aeronautics and Astronautics, June 2015. ISBN 978-1-62410-355-1. doi: 10.2514/6.2015-2336.

L. Kohli. Exploiting perceptual illusions to enhance passive haptics. In *IEEE VR Workshop on Perceptual Illusions in Virtual Environments*, pages 22–24. Citeseer, 2009.

L. Kohli, M. C. Whitton, and F. P. Brooks. Redirected touching: The effect of warping space on task performance. In *3D User Interfaces (3DUI), 2012 IEEE Symposium on*, pages 105–112. IEEE, 2012.

S. W. Kwok, K. H. H. Goh, Z. D. Tan, S. T. M. Tan, W. W. Tjiu, J. Y. Soh, Z. J. G. Ng, Y. Z. Chan, H. K. Hui, and K. E. J. Goh. Electrically conductive filament for 3d-printed circuits and sensors. *Applied Materials Today*, 9:167–175, Dec. 2017. ISSN 2352-9407. doi: 10.1016/j.apmt.2017.07.001.

G. Lawson, D. Salanitri, and B. Waterfield. Future directions for the development of virtual reality within an automotive manufacturer. *Applied Ergonomics*, 53:323–330, Mar. 2016. ISSN 00036870. doi: 10.1016/j.apergo.2015.06.024.

L. Liu, R. Van Liere, C. Nieuwenhuizen, and J.-B. Martens. Comparing aimed movements in the real world and in virtual reality. In *Virtual Reality Conference, 2009. VR 2009. IEEE*, pages 219–222. IEEE, 2009.

A. Lécuyer. Simulating haptic feedback using vision: A survey of research and applications of pseudo-haptic feedback. *Presence: Teleoperators and Virtual Environments*, 18(1):39–53, 2009.

C. L. MacKenzie, R. G. Marteniuk, C. Dugas, D. Liske, and B. Eickmeier. Three-dimensional movement trajectories in Fitts' task: Implications for control. *The Quarterly Journal of Experimental Psychology Section A*, 39(4):629–647, Nov. 1987. ISSN 0272-4987. doi: 10.1080/14640748708401806.

I. S. MacKenzie. A note on the information-theoretic basis of Fitts' law. *Journal of Motor Behavior*, 21(3):323–330, 1989. doi: 10.1080/00222895.1989.10735486.

I. S. MacKenzie. Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human-Computer Interaction*, 7(1):91–139, 1992.

I. S. MacKenzie. Fitts' Throughput and the Remarkable Case of Touch-Based Target Selection. In *Proceedings of the 16th Interenational Conference on Human-Computer Interaction*, volume 9170, pages 238–249, Cham, 2015. Springer International Publishing. ISBN 978-3-319-20915-9 978-3-319-20916-6.

I. S. MacKenzie and W. Buxton. Extending Fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 219–226. ACM, 1992.

I. S. MacKenzie and P. Isokoski. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1633–1636. ACM, 2008.

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

W. McNeely. Robotic graphics: a new approach to force feedback for virtual reality. In *Virtual Reality Annual International Symposium, 1993., 1993 IEEE*, pages 336–341. IEEE, 1993.

D. McRuer and E. Krendel. Mathematical Models of Human Pilot Behavior. Technical Report AGARD-AG-188, 1974.

D. E. Meyer, R. A. Abrams, S. Kornblum, C. E. Wright, and J. E. Keith Smith. Optimality in human motor performance: ideal control of rapid aimed movements. *Psychological review*, 95(3):340, 1988.

G. A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63 (2):81–97, 1956. ISSN 1939-1471(Electronic),0033-295X(Print). doi: 10.1037/h0043158.

A. Murata and H. Iwase. Extending Fitts' law to a three-dimensional pointing task. *Human movement science*, 20(6):791–805, Dec. 2001. ISSN 0167-9457. doi: 10.1016/s0167-9457(01)00058-6.

M. Nagendran, K. S. Gurusamy, R. Aggarwal, M. Loizidou, and B. R. Davidson. Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database of Systematic Reviews*, (8), 2013. ISSN 1465-1858. doi: 10.1002/14651858.CD006575.pub3.

NASA. NASA Space Flight Human-System Standard Volume 2: Human Factors, Habitability and Environmental Health. Technical Report NASA-STD-3001 Volume 2 Rev A, 2015.

E. Nystad and A. Sebok. A comparison of two presence measures based on experimental results. *EXPLORING THE SENSE OF PRESENCE*, page 266, 2004.

C. Pontonnier, G. Dumont, A. Samani, P. Madeleine, and M. Badawi. Designing and evaluating a workstation in real and virtual environment: toward virtual reality based ergonomic design sessions. *Journal on Multimodal User Interfaces*, 8(2):199–208, June 2014. ISSN 1783-7677, 1783-8738. doi: 10.1007/s12193-013-0138-8.

C. N. Riviere, W. T. Ang, and P. K. Khosla. Toward active tremor canceling in handheld microsurgical instruments. *IEEE Transactions on Robotics and Automation*, 19(5):793–800, Oct. 2003. ISSN 1042-296X. doi: 10.1109/TRA.2003.817506.

J. Schiefele, O. Albert, V. van Lier, and C. Huschka. Simple force feedback for small virtual environments. In *Aerospace/Defense Sensing and Controls*, pages 100–110. International Society for Optics and Photonics, 1998.

A. Sears and B. Shneiderman. High precision touchscreens: design strategies and comparisons with a mouse. *International Journal of Man-Machine Studies*, 34(4):593–613, 1991.

M. Seixas, J. Cardoso, and M. T. G. Dias. One Hand or Two Hands? 2d Selection Tasks with the Leap Motion Device. 2015.

G. A. Sexton. Cockpit–Crew Systems Design and Integration. In D. C. Nagel, editor, *Human Factors in Aviation*, pages 495–526. Academic Press, San Diego, 1988. ISBN 978-0-08-057090-7. DOI: 10.1016/B978-0-08-057090-7.50021-7.

C. E. Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, Jan. 1949. ISSN 0096-8390. doi: 10.1109/JRPROC.1949.232969.

C. Shemelya, F. Cedillos, E. Aguilera, E. Maestas, J. Ramos, D. Espalin, D. Muse, R. Wicker, and E. MacDonald. 3d printed capacitive sensors. In *2013 IEEE SENSORS*, pages 1–4, Nov. 2013. ISBN 1930-0395. doi: 10.1109/ICSENS.2013.6688247.

T. B. Sheridan. Musings on telepresence and virtual presence. *Presence*, 1 (1):120–126, 1992.

M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence*, 3(2):130–144, 1994.

R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, 61(6):751–789, Dec. 2004. ISSN 10715819. doi: 10.1016/j.ijhcs.2004.09.001.

R. Stone. Haptic feedback: A brief history from telepresence to virtual reality. *Haptic Human-Computer Interaction*, pages 1–16, 2001.

D. Swapp, V. Pawar, and C. Loscos. Interaction with co-located haptic feedback in virtual reality. *Virtual Reality*, 10(1):24–30, May 2006. ISSN 1359-4338, 1434-9957. doi: 10.1007/s10055-006-0027-5.

B. T. Sweet and L. J. Trejo. The identification and modeling of visual cue usage in manual control task experiments. 1999.

S. Tachi, T. Maeda, R. Hirata, and H. Hoshino. A Construction Method of Virtual Haptic Space. In *Proceedings of the ICAT'94*, Tokyo, Japan, 1994.

R. J. Teather and W. Stuerzlinger. Pointing at 3d targets in a stereo head-tracked virtual environment. In *3D User Interfaces (3DUI), 2011 IEEE Symposium on*, pages 87–94. IEEE, 2011.

R. J. Teather, D. Natapov, and M. Jenkin. Evaluating haptic feedback in virtual environments using ISO 9241–9. In *Virtual Reality Conference (VR), 2010 IEEE*, pages 307–308. IEEE, 2010.

W. H. Teichner. Recent studies of simple reaction time. *Psychological Bulletin*, 51(2):128–149, 1954. ISSN 1939-1455(Electronic),0033-2909(Print). doi: 10.1037/h0060900.

US Department of Defense. Department Of Defense Design Criteria Standard: Human Engineering. Technical report, 1999.

H. Wan, S. Zou, Z. Dong, H. Lin, and H. Bao. MRStudio: A mixed reality display system for aircraft cockpit. In *2011 IEEE International Symposium on VR Innovation*, pages 129–135, Mar. 2011. doi: 10.1109/ISVRI.2011.5759615.

A. T. Welford. *Fundamentals of skill.* 1968.

B. G. Witmer and M. J. Singer. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, 7(3):225–240, June 1998. ISSN 1054-7460. doi: 10.1162/105474698565686.

J. O. Wobbrock, E. Cutrell, S. Harada, and I. S. MacKenzie. An error model for pointing based on Fitts' law. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1613–1622. ACM, 2008.

R. S. Woodworth. Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements*, 3(3):i, 1899.

I. Yavrucuk, E. Kubali, and O. Tarimci. A low cost flight simulator using virtual reality tools. *IEEE Aerospace and Electronic Systems Magazine*, 26(4):10–14, Apr. 2011. ISSN 0885-8985. doi: 10.1109/MAES.2011.5763338.

C. Youngblut and O. Huie. The relationship between presence and performance in virtual environments: Results of a VERTS study. In *Virtual Reality, 2003. Proceedings. IEEE*, pages 277–278. IEEE, 2003.

L. Zea, S. Over, D. Klaus, J. Tanner, and K. Stroud. Development of a Cockpit Architecture for the Dream Chaser Orbital Vehicle. page 3421, 2012.

Appendices

Appendix A

Result Tables

A.1 Passive Haptics Experiment

Sequence	Haptics	Mean	Std. Dev.
NH First	—	3.877	0.5265
PH First	—	3.99	0.3684
—	NH	3.71	0.3805
—	PH	4.157	0.412
PH First	NH	3.879	0.3892
PH First	PH	4.1	0.3284
NH First	NH	3.541	0.3007
NH First	PH	4.213	0.4934

Table A.1: Throughput Means

Factor	F ratio	p value
Sequence	0.5297	0.4761
Haptics	35.53	1.221×10^{-5}
Sequence:Haptics	9.062	0.007516

Table A.2: Throughput ANOVA

Sequence	Haptics	df	t	p value
PH First	—	9	2.547	0.03135
NH First	—	9	5.503	0.0003787
—	PH	18	-0.602	0.5547
—	NH	18	2.176	0.04309

Table A.3: Throughput t-tests

Condition	Sequence	Haptics	Mean	Std. Dev.
—	NH First	—	3.468	0.6042
—	PH First	—	3.642	0.4628
—	—	NH	3.246	0.3738
—	—	PH	3.864	0.5044
First	—	—	3.462	0.5202
Second	—	—	3.648	0.5533
Second	PH First	NH	3.426	0.4171
First	PH First	PH	3.857	0.4178
First	NH First	NH	3.067	0.2222
Second	NH First	PH	3.87	0.6021

Table A.4: Throughput Unfiltered Means

Factor	F ratio	p value
Condition	0.9756	0.3364
Haptics	4.732	0.04318
Condition:Haptics	52.17	1.017×10^{-6}

Table A.5: Throughput Unfiltered ANOVA

Condition	Sequence	Haptics	df	t	p value
—	PH First	—	9	4.525	0.001436
—	NH First	—	9	5.662	0.000309
—	—	PH	18	-0.05582	0.9561
—	—	NH	18	2.403	0.02727
First	—	—	18	-5.284	5.04×10^{-5}
Second	—	—	18	-1.92	0.07091

Table A.6: Throughput Unfiltered t-tests

Sequence	Haptics	Mean	Std. Dev.
NH First	—	77.5	10.56
PH First	—	71.2	8.764
—	NH	71	9.701
—	PH	77.7	9.565

Table A.7: Presence Score Means

Factor	F ratio	p value
Sequence	4.089	0.05828
Haptics	6.079	0.02396
Sequence:Haptics	0.7164	0.4084

Table A.8: Presence Score ANOVA

Haptics	Cronbach's alpha
No Haptics	0.9139
Passive Haptics	0.9191

Table A.9: Presence Score Cronbachs alpha

Sequence	Haptics	Mean	Std. Dev.
NH First	—	15.07	2.948
PH First	—	13.1	3.523
—	NH	14.72	2.854
—	PH	13.45	3.762
PH First	NH	14.5	3.064
PH First	PH	11.7	3.529
NH First	NH	14.95	2.773
NH First	PH	15.2	3.259

Table A.10: Arm Fatigue Ratings Means

Factor	F ratio	p value
Sequence	2.119	0.1627
Haptics	9.717	0.00595
Sequence:Haptics	13.9	0.001538

Table A.11: Arm Fatigue Ratings ANOVA

Sequence	Haptics	df	t	p value
PH First	—	9	−4.583	0.001323
NH First	—	9	0.4596	0.6567
—	PH	18	−2.304	0.03336
—	NH	18	−0.3443	0.7346

Table A.12: Arm Fatigue Ratings t-tests

Please circle the number that best describes the amount of fatigue you feel in your arm. Try to consider the overall sensation of fatigue, soreness, and exertion level. Consider the arm to include shoulder muscles that you use to move the arm.	
6	No fatigue
7	Very, very lightly fatigued
8	
9	Very lightly fatigued
10	
11	Lightly fatigued
12	
13	Somewhat fatigued
14	
15	Heavily fatigued
16	
17	Very heavily fatigued
18	
19	Very, very heavily fatigued
20	Maximal fatigue

Table A.13: Borg RPE Scale as used

A.2 Design Evaluation Experiment

Group	Design	Mean	Std. Dev.
TS	—	1.277	0.3789
VR	—	1.967	0.378
—	Edgekey	1.704	0.5188
—	Keypad	1.57	0.5068

Table A.14: RMSE Means

Factor	F ratio	p value
Group	21.42	0.000145
Design	5.944	0.02374
Group:Design	0.1669	0.687

Table A.15: RMSE ANOVA

Group	Design	Mean	Std. Dev.
TS	—	2595	567.3
VR	—	2813	383.2
—	Edgekey	2688	471.1
—	Keypad	2729	512.5

Table A.16: Response Time Means

Factor	F ratio	p value
Group	1.199	0.2859
Design	0.6814	0.4184
Group:Design	0.00184	0.9662

Table A.17: Response Time ANOVA

Group	Design	Mean	Std. Dev.
TS	—	84.97	10.35
VR	—	50.46	24.07
—	Edgekey	56.4	29.38
—	Keypad	77.54	15.26
VR	Edgekey	33.33	19.75
VR	Keypad	67.59	13.47
TS	Edgekey	81.57	11.61
TS	Keypad	88.38	8.033

Table A.18: Correct Prompts Means

Factor	F ratio	p value
Group	43.56	1.552×10^{-6}
Design	63.93	8.309×10^{-8}
Group:Design	26.89	3.872×10^{-5}

Table A.19: Correct Prompts ANOVA

Group	Design	df	t	p value
VR	—	11	8.039	6.234×10^{-6}
TS	—	10	2.287	0.04526
—	Keypad	21	4.441	0.0002262
—	Edgekey	21	7.053	5.839×10^{-7}

Table A.20: Correct Prompts t-tests

Group	Design	Mean	Std. Dev.
TS	—	65.35	8.535
VR	—	70.01	22.65
—	Edgekey	77.74	13.4
—	Keypad	57.83	15.18
VR	Edgekey	85.61	11.21
VR	Keypad	54.42	20.4
TS	Edgekey	69.15	10.06
TS	Keypad	61.55	4.468

Table A.21: NASA TLX Means

Factor	F ratio	p value
Group	1.688	0.208
Design	23.57	8.455×10^{-5}
Group:Design	8.252	0.009113

Table A.22: NASA TLX ANOVA

Group	Design	df	t	p value
VR	—	11	−4.205	0.001474
TS	—	10	−2.718	0.02164
—	Keypad	21	1.132	0.2703
—	Edgekey	21	−3.693	0.001351

Table A.23: NASA TLX t-tests

Group	Trial	Mean	Std. Dev.
TS	—	0.9402	0.425
VR	—	1.359	0.5176
—	1	1.225	0.5421
—	2	1.177	0.5144
—	3	1.074	0.5053
VR	1	1.428	0.5754
VR	2	1.34	0.4877
VR	3	1.308	0.5244
TS	1	1.002	0.4223
TS	2	0.9994	0.5038
TS	3	0.8188	0.3488

Table A.24: Tracking Only Trials RMSE Means

Factor	F ratio	p value
Group	4.937	0.0374
Trial	3.649	0.06987
Group:Trial	0.1621	0.6913

Table A.25: Tracking Only Trials RMSE ANOVA

Topic	Feedback Summary Category	VR Group	TS Group
Edgekey	Accuracy Worse	1	2
Edgekey	Busy Design	0	1
Edgekey	Centered Flight Task Better	3	13
Edgekey	Clean Design	3	2
Edgekey	Easier	0	4
Edgekey	Familiar	1	0
Edgekey	Hand Blocks View	3	4
Edgekey	Labels Easy to Read	1	0
Edgekey	Labels Hard to Read	1	0
Edgekey	Switch Difficult	14	12
Edgekey	Switch Neutral	0	1
Fatigue	Eye Fatigue	3	0
Fatigue	Fatigue from Joystick	0	4
Fatigue	Prompting Arm	4	1
Keypad	Buttons Always Visible	5	5
Keypad	Buttons Bad Layout	2	0
Keypad	Buttons Confusable	0	4
Keypad	Buttons Proximal	6	7
Keypad	Easy Focus Switch	2	1
Keypad	Familiar	6	11
Keypad	Labels Good	1	0
Keypad	Labels Hard to Read	1	0
Keypad	More Mistakes	1	2
Keypad	More Successful	0	1
Keypad	Tracking Easier	1	0
Other	Colors Disliked	2	2
Other	Hand Tracking Issues	9	0
Other	Prompt Location Bad (Both)	1	1
Other	Single Finger Hard	1	1
Other	Touchscreen Issues	0	1

Table A.26: Full Feedback Comments by Category.