Gaze Controlled Interactions for AR/VR

by

Busra Demirci

May 2020

Presented to the

Division of Science, Information Arts, and Technologies

University of Baltimore

In Partial Fulfillment

of the Requirements for the Degree of

Master of Science

Approved by: Kathryn Summers (Jul 17, 2020 09:32 EDT)

[Dr. Kathryn Summers, Thesis Advisor]

[Dr. Greg Walsh, Committee Member]

Abstract

This simulation study assessed the impact of gaze and hand gestures as a primary input method for the aerospace industry by utilizing Augmented Reality (AR) and Virtual Reality (VR). Fifteen NASA employees with a varying range of AR/VR expertise participated in an interactive simulation of driving a rover on Mars using a mechanical robot, 360 Fly Camera, HTC Vive, and Leap Motion. Each participant received two prototypes to interact with, a control and an experiment, where they were instructed to drive a rover simulation by interacting with on-screen buttons for the control and directional hand gestures for the experiment. Due to the limitation in technology, eye movements were not used. Participants were evaluated on rate of success when navigating between 3D printed rockets, perceived ease of use, and qualitative preference. AR/VR slightly impaired the perception of target objects in the physical environment. Results exhibited preference for the hand gestures with a near significant rate of success and perceived ease of use despite the small sample size.

Acknowledgments

It is with sincere appreciation that I thank the following individuals for their contributions:

To Dr. William Little at NASA for his expertise, consultation, assistance, and advocacy in driving this project and providing the technology and tools.

To the engineers and supervisors at NASA for their voluntary participation and cooperation.

To the members of the advisory committee, Dr. Kathryn Summers and Dr. Greg Walsh, for their assistance. With special recognition to Dr. Kathryn Summers for her mentorship and dedication to my success throughout this experience.

Lastly, to my family members, my mother, aunt, cousins, and siblings for their feedback, review, and moral support throughout this project.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Augmented Reality and Virtual Reality (AR/VR) technologies have quickly begun shifting interaction paradigms and user expectations as they relate to digital interfaces and experiences across a wide variety of industries. The immersive experience increases the importance of ergonomics and the potential usefulness of gaze as a primary method of navigation and interaction. Yet gaze-based interactions have traditionally been focused on 2D applications and less attention has been paid to the emerging techniques that could be applied to advanced interfaces.

Background of the Study

The evolution of such immersive experiences dates back to 1957 where Morton Heilig, an inventor and filmmaker, invented a simulator also known as The Sensorama Machine, which produced an illusion of reality by utilizing multi-sensory inputs through three-dimensional motion picture, smells, stereo sounds, seat vibrations, and fans to simulate wind (Augmented Reality – The Past, The Present and The Future 2017).

Utilizing such inventions, Heilig patented the next levels of the Sensorama Machine, which they referred to as “The Experience Theater," intended to serve wider audiences, and “The Thrillerama Theater," which projected three-dimensional motion pictures where viewers could interact with images on the screen (Tate, 1996). Although

Heiling's inventions weren't formally part of an AR/VR initiative, they paved the way for engineers like Ivan Sutherland who sought methods that would immerse users in three-dimensional information. They achieved this by formalizing VR viewing capabilities through their head-mounted display, which utilized images that varied in perspective and adjusted as a user moved (Sutherland & Ivan, 1968). The level and speed of growth of AR/VR technology within the industry serve as key indicators that establish foundational interaction guidelines and principles that support usability for wide, globally diverse audiences (Boland, 2018).

Rationale and Purposes of the Study

Cross-functional industries have successfully begun integrating gaze as a supplement to interaction inputs, yet the implications and use within aerospace environments haven't been thoroughly explored and have the potential to enhance the quality and efficiency of information discovery on surfaces outside of Earth. Spaceflight technology and its applications for use in exploration have been focused on optimizing proficient hand movements during Extravehicular Activity (EVA) or spacewalks (Rainey, 2017). The implication of missions with an extended duration and necessity for precision require, "the ability of robots to work alongside the human crew to perform necessary tasks more efficiently. These tasks include those that are monotonous or risky and impose on the available time astronauts have to focus on science experiments" (Rainey, 2017).

With the advancement of AR/VR technologies and the potential for diverse utilization, the need for foundational design principles in an increasingly sophisticated modality become ever more apparent. It is reasonable to suppose that there are AR/VR interaction techniques that are undiscovered and that will require diligent research to validate. Such interaction models could include having traditional tap and controller inputs eventually move towards the use of gaze, gestures, and brain wave activity as primary inputs, thus have influencing the goals of this study.

Problem and Hypothesis Statement

Gaze-based knowledge has traditionally been focused on the function and implication of anatomical behavior as it relates to human-computer interaction, providing optimal cues to prompt focus and 3D component optimization for an individual's field of vision, perception, and preference. The focus of this paper strives to understand the broader research that has been conducted around cognition, ergonomics, and gaze-based interactions to inform the potential challenges that could occur around relying on gaze as the primary input method for navigating, steering, and controlling rover movement in a virtual environment. Introducing dwell time as a threshold for optimal interaction presentation could reduce the speed, effort, and cognitive load involved with the interaction while demanding precision to select a reduced target which could affect selection efficiency.

Earlier research provides conclusions drawn from a mixture of eye-tracking studies and series of qualitative open-ended questions to participants. As a result, the focus of this study will rely on six open-ended questions (reported in Chapter 3: Methodology). Questions asked will correlate to scenes that a participant is prompted to interact with via simulated rover control in a spatial environment to test the following hypotheses. To meet this goal, three specific hypotheses were to be tested:

Hypothesis One: Augmenting gestures with gaze has the potential to solve interaction efficiency challenges while ensuring users have a natural and seamless experience.

Hypothesis Two: The minimum threshold to gaze-based interaction intention is a fixation of three seconds wherein a participant's expectation is to initiate an action or interaction with an object.

Hypothesis Three: Reducing cognitive load by contextualizing menus and interaction-rich experiences allows participants to focus on the environment at hand and reduces the reliance on ergonomic mobility to discover and explore.

The original design for this experiment depended on eye tracking within the AR/VR equipment, using the Tobii VR headset. Unfortunately, security limitations at NASA dictated a change from Tobii VR technology to the HTC Vive and Leap Motion, which do not support eye tracking. Therefore, the following alternative hypotheses were explored:

Hypothesis One: Replacing button controls with hand gestures to control AR/VR inputs has the potential to solve interaction efficiency challenges while ensuring users have a natural and seamless experience.

Hypothesis Two: Participants will readily adapt to hand gestures to control AR/VR inputs if the hand gestures utilize natural mappings.

Hypothesis Three: Reducing cognitive load by contextualizing menus and interaction-rich experiences allows participants to focus on the environment at hand and reduces the reliance on ergonomic mobility to discover and explore.

Chapter 2: Review of Literature

Augmented/Virtual Reality technologies are regularly investigated and referred to as creative advances that are progressively supporting cutting-edge use cases whether they are embedded into micro-interactions and functions with a digital application or utilized as immersive experiences. Four primary areas of research are relevant to the inquiry of the study: 1) cognitive theory research of maintenance and the ability to learn input behavior; 2) steering virtual environments and optimal display methods to support function; 3) reaction time correlations to gaze-based activity; and 4) accelerating gaze-based browsing by analyzing variable dwell times. The main concentration of prior research has centered around general display modalities and the reaction times of gaze-patterns as they relate to virtual environments, wherein lies an opportunity to extend these modalities into enterprise use for information discovery, particularly for aerospace.

Cognitive Theory

The effects of innovations in AR/VR interaction are investigated through interdisciplinary cognitive science. The brain makes use of explicit regions dedicated to completing a task that span behavioral and cognitive patterns in order to learn a new ability. This is specifically evident when a user carries out difficult tasks that necessitate extensive concentration and error prevention during complex interactions such as engaging with enterprise software or analytical synthesis. Conversely, undemanding interactions and intuitive skills, such as engaging in physical activity or operating a vehicle, readily prime neurochemical levels in the body, which means that "learning in this system is mediated by the basal ganglia and is critically dependent on the neurotransmitter dopamine" (Maddox, 2017). Discharging dopamine synthetic compounds permit muscle memory to be prepared when "the learner performs a task and receives feedback" (Maddox & Campbell, 2017). In order for new input methods to be learned they need to be stored, guaranteeing that information is stored in a person's long-term memory requires intermittent preparing and practice to support retention. VR augments habit-forming behavior by presenting interactions in a realistic environment, thus making it a fundamental instrument for training user muscle memory. It is especially helpful in supporting retention, since it lends itself to intermittent practice. The process of information retention and learning is best represented by Hermann Ebbinghaus', "overlooking bend...plots maintenance as a component of time. Introductory maintenance (signified by the blue dab) is 100 percent. The red bend is the overlooking bend. Overlooking is quick, with just 40 percent of the data held one day subsequent to preparing, and just 15 percent held multi-weak in the wake of preparing" (Maddox, 2017).

The demonstration of phases in VR where information is repetitively displayed could expand retention and maintenance by creating connections with enhanced cognitive substances that allow retraining moments to occur. Maddox's (2017) research indicates that the initiation of learning triggers 100% maintenance, yet quickly diminishes thereafter. When students in the study were retrained on information that was seemingly overlooked and re-tested a few days later, maintenance was back to 100% (Maddox, 2017). This is predominantly due to the presence of relevant signals that are embedded within a situation where an individual can focus, hence improving the efficiency and quality of maintenance, "Testing and focused retraining invigorates long haul memory stockpiling. Rehashing this procedure a second, third and fourth time decreases overlooking. Testing and focusing on retraining is at the core of preparing for maintenance" (Maddox, 2017).

Understanding the function of memory and cognition is imperative to interaction research especially with new input types and methods that users need to learn with technologies they haven't previously interacted with. Cognitive maintenance behavior illustrates the importance of retraining and the potential to establish retained interaction habits in individuals when placed in an untrained environment. The need for maintenance may also be reducible through coach-marks and contextual onboarding where information is optimally surfaced for the user to understand how to direct interactions like gaze in a manner that elicit desired outputs.

Steering Virtual Environments

Designing for interactions like steering a vehicle or rover in a virtual environment can be eased by leveraging intuitive methods like gaze as an input instead of requiring a user to recognize or recall specialized actions and gestures that can be performed. Sophie Stellmach and Raimund Dachselt performed research on the feasibility of utilizing gaze as a steering method in virtual environments, finding that “gaze-based interaction is still error-prone and requires a deeper investigation for the design of eye gaze interfaces. The Midas Touch issue poses one of the major problems of gaze-based interaction” (2012). With the challenge of differentiating between a fixation for information processing versus interaction prompts, they hypothesized that utilizing dwell times as a measuring parameter could help surmount such issues.

Similar to a majority of overlays in 3D environments, Stellmach and Dachselt superimposed 2D objects within a 3D virtual environment to measure inputs in the form of continuous and discrete gaze, “...as soon as an interface element (e.g., an active region) is looked at, an associated event is (continuously) issued. Once an area is not viewed anymore, the movement stops, whereas with discrete input, an action can be switched on or off by issuing a discrete event (2012). The advantage of such a model allows the user to focus on the environment at hand and use gaze as a way to inform next steps in his/her experience. Although the velocity of such an interaction may not be as efficient as tapping a button or other physical initiation, it may assuage the burden and ergonomic limitations of some control movements. Through two iterations, Stellmach and Dachselt discovered that participants preferred continuous gradient-based inputs, “combining different motion directions and velocities within one UI, which reduced the need for time-consuming dwell activations” (2012). With the gradient-based designs creating a continuous field of view effect, participant task completion was not only quicker, but was qualitatively perceived as intuitive, quick, and typical to organic movement. However, although this research proves the value of certain UI elements and efficiency parameters, it does not address the implications in an environment that houses robust information displays and task capabilities, such as the environment for controlling the NASA rover.

The success and usability of scenes presented by Stellmach and Dachselt (2012) indicate effective task completion and the proposal for use of “Sticky-Gaze Pointers” which function as supplemental visual cues allowing feedback to anchor gaze and reduce the implication of the problems of “overshooting and wandering off.” The implications of such problems may be amplified in high latency environments within subpar visual connections and uncertain situations. Robotic tests were performed at a Mars-like desert in Utah which was controlled via satellite link by Virtual Reality laboratory located in Bremen Germany. This setup presented the advantages of utilizing a virtual control panel to execute missions remotely through satellite video streams. The Virtual Reality laboratory was utilized as the mission control center. 3-dimensional visualizations of the location and the motor positions of the robot were sent to the mission control center which aided operators to move the virtual camera freely as needed. Multiple Robots Coyote III and SherpaTT and also different input devices were simultaneously controlled. The operator had both a better view and obtained sufficient information of the real time current status, even during the sand storms, dusty wind or under bad lighting, to continue the mission safely (Planthaber & Maurus, 2017). Utilization of gradient-based UI could be deployed in the rover context where steering movements are only executed when an individual gazes outside of a designated gaze "resting region" and commands may be issued when gaze is directed at zones that represent directional input, "For example, a slow forward movement will be executed, if a user looks only slightly above the resting region. For the gradient-based design, the active regions need to be bigger than for the discrete input to provide a better interaction with the underlying gradient mask" (Stellmach & Dachselt, 2012). A sense of continuity can intuitively be displayed through visual UI components and patterns that extend a sense of embodied cognition and pre-disposed user expectation.

Reaction Time Correlations

Much like smartphone tap targets and hotspots, 3D interaction in AR/VR settings require aiming the required input method "to communicate with objects in the virtual environment (VE) using gaze offers new opportunities over traditional VR devices" (Cámara et al., 2018). The type of new opportunities refers to how the possibilities for new interaction paradigms could be implemented and assessed for future use.

The advantages of utilizing gaze have been referenced and explored comparing, "the use of gaze pointing and conventional 3-D pointing by hand extension in VR, and [they] found gaze interaction to take significantly less time to complete the same task for distant objects" (Cámara et al., 2018). This would mean that using gaze interaction saves time to aim in virtual environments where even milliseconds can be a determining factor in, for example, a professional e-sports event with the subconscious individual pressure of millions of dollars on the line. Hansen et al. (2003) studied the efficacy of hands-free interaction by comparing the use of a mouse click to gaze dwell mechanisms using a Japanese keyboard. They found that gaze allowed for more vigorous selections, yet mouse clicks were more precise and resulted in 33% more efficient interaction due to the lack of dwell dependency (Hansen et al., 2003).

The main reason that gaze technology is currently outperformed by other methods is primary due to the gaming industry, gamers traditionally have been using controllers/ mice for all their needs which is why they are more familiar with the movements. If gaze technology were further tested and users could fully acclimate themselves to the use of such technology, gaze interaction would probably be able to outperform interaction control using a handheld device. (Dean et al., 2011). According to Dean et al. (2011), "Eye–hand coordination depends on a combination of retinal and extra-retinal signals necessary for accurate movement." Although gaze would be an optimal replacement for gesture-based interactions, limitations in current hardware would need to be optimized to support delayed visual feedback of eye movement in some devices, especially with the 50 to 200 ms delay it can take to update device displays as an output to touch, which can result in a delay in physical perception of direct inputs (Ng et al., 2012). Such delays can become more prominent considering objects displayed in behavior and result.

When observing the reflex arc of an input, the further a signal has to travel from the integrating center in the brain the longer it will take for an effector (neuron) to react. This would mean that an individual could react more efficiently by using gaze as an input method "In most cases, small delays such as the normal delay between a hand moving a computer mouse and the resulting motion of the cursor on the screen are not disturbing. However, longer delays may require adaptation before one can meaningfully use the available feedback to successfully control the device" (Cámara et al., 2018).

Even though the dwell time for the visual user includes a delay time for a handheld device, if dwell times could be reduced during interaction-based intentions, gaze software would allow for a remarkable difference in reaction time pertaining to 3D environments and support the necessity for quick, accurate responses.

Accelerating Gaze-Based Browsing

Unlike other modes of interface that utilize already learned skills such as pointing, grabbing, and moving objects, the eye is limited in that these learned skills do not yet exist. Only through the observation of the eye's natural movements can we establish intuitive patterns for the use of gaze-based interfaces. Thus our aim should be "to obtain information from a user's natural eye movements while viewing the screen, rather than requiring the user to make specific trained eye movements to actuate the system" (Jacob, 1995). One set-back in the development of gaze controlled interfaces is the issue of the "Midas Touch" where the gaze of the user may trigger unwanted commands (Jacob, 1995). In order to circumvent this issue, dwell times can be utilized in order to interpret the user's intended commands. Chen and Shi (2018) used variable dwell times in their study to improve gaze-based selection. Although this method does improve the selection process, it potentially places the eye in the position of needing to learn trained eye movements. Even in this case there may still be some aberrant movements since gaze can be difficult to fully control through intention. At the same time, due to preset dwell times the time to complete tasks will consequently be longer than a modality which utilizes more natural dwell times.

In a study analyzing a novel operating system interface; the location of the gaze point and the task context was utilized in an algorithm to alleviate the problems faced in interpreting user eye movements in a gaze-based interface. Salvucci and Anderson (2000) presented some of the issues regarding gaze-based interfaces and how they could potentially be addressed. The aim, in this case, was to utilize a gaze-added interface in order to solve issues such as noise due to the eye-tracking equipment and the dissociation between the gaze point and the user's visual attention. The algorithm used in this study was based on a probabilistic model which interprets user gazes and points to the items to which the user is likely attending. However, "a more rigorous design could determine better priors empirically by observing long-term behavior [with] the interface" (Salvucci & Anderson, 2000). As Salvucci et al. state, when they are able to calibrate users accurately on the eye tracker, users commented that the gaze modality was "smooth and seamless," clearly indicating the importance and the necessity for further long-term analysis on gaze. Salvucci and Anderson (2000) does mention, however, that “a detailed study that quantitatively measures the effect of intelligent gaze interpretation on user performance and ease of use” would be beneficial. The authors concluded that a quantitative measure of these parameters would allow for an enhanced image of the improvement in quality of the interface experience, as a result of more specifically studying gaze movements such as dwell time in different environments.

Eye-tracking is a method generally used in understanding and measuring gaze-patterns as they relate to gaze-based behaviors. The success of gaze-based detection user interfaces depends on the equipment used and the sensitivity of the deployed scheme. By adjusting the scheme to address the reality that a perfect environment is not always possible, we can increase the success rate of this user interface. Increasing the success rate of gaze-based schemes that can track eye movement in an environment with noise and interference will help us to improve this technology for average consumers. However, as Liao and Yu’s study (2010) pointed out, a more robust algorithm is needed to make a user interface that can make real-time tracking for interactive applications possible.

Liao and Yu (2010) recognized that corneal reflection can be a significant obstacle in an interface. Unlike the starburst algorithm, which uses adaptive thresholding scheme to remove the corneal reflection, the authors tackle this problem by breaking apart the pixel data sets to extract the points on the contour of the pupil.

Even if one section of the image, split into four parts, is completely thrown off by corneal reflection, the algorithm is able to correct itself by omitting the data set of this section. By doing this we are still able to find the center and reconstruct the ellipse using the data sets from the other corresponding sections of the image (Liao & Yu, 2010).

Using this method to extract data points of the contour allows us to skip the pre-processing step that is used in the starburst algorithm. Thus, using this method, the risk of failure is reduced due to the imaging process and the algorithm can approach the image in its raw form. Unfortunately, the study does not provide a quantitative comparison of the iterative shifting and noise removal in the method proposed by the authors, to the success rate without this scheme, making it hard to evaluate the study results. Liao and Yu's study (2010) also mentions that this method might not be viable for real-time applications in its current state, but with improvements to hardware and to the algorithm used it may be practical for real-time application in the near future. Thus, the existing eye-tracking technology, although improved, may not provide the most accurate quantitative results, especially in the context of the "Midas Touch" where accuracy issues occur when solely relying on gaze as an interaction modality, when "the system is required to distinguish between (1) gaze intended to gather visual information and (2) gaze intended to activate a specific command" (EyeTracking Inc., 2012).

In order to alleviate the problem of the Midas Touch, Chen and Shi (2018) proposes a gaze-based browser utilizing a two-step selection policy with variable dwell time. The authors develop a "new probabilistic model of eye gaze behavior during web surfing." Using information from the user's natural gaze in order to estimate the probability of the user selecting each hyperlink, for each hyperlink is assigned a unique dwell time, taking into account the likelihood of that link being selected. Thus dynamic dwell times are assigned to each hyperlink in the study by Chen and Shi (2018) The parameters of assigning dwell times is determined by the researchers themselves, which they state is "to achieve the best tradeoff between selection speed and false selection." The performance of the method was evaluated quantitatively via simulation and experiment. The variable dwell time policy resulted in a value similar to a 100ms uniform dwell time policy; however, the error rate was reduced by 50%. Their method also achieved an accuracy similar to a 300ms uniform dwell time policy and the error rate decreased by 60%.

Chen and Shi's (2018) study proposes that their probabilistic model takes into account spatial information, thus enabling this model to account for noise in eye tracking and the possibility that the hyperlink may not be fixated upon exactly. In addition to this, temporal information is also incorporated into the model, which takes into account the past history of the user's individual gaze behavior by using hidden Markov models. The model parameters can be established using machine learning, where data is collected from the users while they browse a uniform dwell time policy, gaze-based browser. Also, the model that is proposed here can further on be customized according to the user's gaze data.

Gaze versus Gesture Utilization for Hands-free Interaction

Although the foundational focus of this study was around gaze as a primary interaction pattern and its potential advantages, due to security limitations, the AR/VR headset available for the purposes of this experiment supported hand gestures as an alternative way to research the hypotheses. The premise of gesture recognition is to allow non-verbal communication between a person and a device, "gesture recognition is said to be the interpretation of said movements by motion sensors, accelerometers via various mathematically applied computer algorithms. This is a form of perceptual computing which allows humans to execute commands based on the gestures recognized by the devices they wish to control" (Eisenberg, 2017).

Common human-computer interactions today are generally centered around smartphone touchscreens where users interact with inputs directly on the display via tapping, swiping, and hard pressing. Default interactions currently present in AR/VR involve the use of hand controllers to point at interface elements similar to smarthphone interactions. With the intention of smoothly transitioning users from touchscreen devices to AR/VR headsets, Moberg and Pettersson (2017) researched the interaction possibility by mimicking smartphone behavior and replacing it with a real world hand gestural equivalent using a Leap Motion Controller in a Unity virtual environment. The focused scope of their study measured the efficacy of text typing using hand gestures by quantitatively measuring accuracy, completion time, and usability across five keyboard implementations, "The research concludes that hand gestures are reasonable to use as input technique to accomplish certain tasks that a smartphone performs. These include simple tasks such as scrolling through a website or opening an email. However, tasks that involve typing long sentences, e.g. composing an email, is arduous using pinch gestures. When it comes to typing, the authors advice developers to employ a continuous gesture typing approach such as Swype for Android and iOS" (Moberg & Pettersson, 2017). Such findings imply how the type of gesture can contribute to the usability and long-term translation of input methods for AR/VR; since hands are the most common method of interacting with technology today, validating the use of gestures could prove to be an effective transition to gaze as AR/VR becomes more widely utilized. Similar to gaze research, input type impacts gesture as well, the findings relating to accuracy in Moberg and Pettersson's (2017) study parallel gaze findings where calibration and dedicated hotspots play a key role in accurately capturing inputs that may slightly vary per user by allowing for a wide target area, thus reducing the potential for margin of error.

When considering hands-free input methods ergonomic implications can result in fatigue for the user leading to strenuous task completion and a smaller window of accurate input mapping. Satriadi et al. (2019) found that navigational gesture interactions in AR/VR were prone to arm fatigue when participants utilized a "mid-air" arm trajectory held for a prolonged period of time, also known as the "gorilla-arm effect." They proposed a potential solution of "micro-gestures" where fatigue could be reduced by encompassing a shortened range and duration of movement combined with the ergonomic improvement of, "a lowered arm posture with a bent elbow," which would require more advanced sensing technology to fully validate (Satriadi et al., 2019). Although such considerations can improve the magnitude of fatigue in a user's interactions, strain can still exist from continuous interaction. Since gaze is a natural action performed to observe an individual’s surroundings, the implication of fatigue could potentially be furthered reduced with sufficient fixation and interaction prompt detection once AR/VR technology advances to support such calibration and input mapping.

The advantages and disadvantages of gaze versus gesture as input methods has been tested through awareness cues presented in virtual environments, Piusomboon et al. (2019) used a baseline condition and compared it to environments with collaborative variables where users were prompted to find objects and interact with placement tasks using gaze, head, and gesture movements separately. The study conducted consisted of awareness cue variables like Field-of-view (FoV) where users were able to establish an understanding of orientation in the virtual space and in comparison to a partner sharing the same space. The next two variable conditions combined FoV with either head-gaze or eye-gaze which would serve as the origin for identifying the center of the FoV to indicate where the user was looking to prompt an interaction (Pisusomboon et al., 2019). “We particularly noticed that in the Head-gaze condition the number of hand gestures used was the lowest among all the condition. The number of pointing gestures used was highest in the Baseline condition and lowest in the Eye-gaze condition. The Head-gaze condition has a similar number of pointing gestures as the Eye-gaze condition. This finding makes sense as participants tried to use the ray available in Head-gaze and Eye-gaze conditions in place of the hand pointing in the Baseline and FoV conditions, which is an indication of later two conditions being physically demanding" (Pisusomboon et al., 2019). When comparing the use of head and gaze conditions in the study, Pisusomboon et al. (2019) allowed the simultaneous use of three hand gestures where participants had the ability to "point, grasp, and neutral" and found that the lower use of gestures during head-gaze conditions indicated a more natural and intuitive approach than utilizing gaze or gesture alone. Since head movements signify prompts similar to the hands, the findings of this study imply that a combination of interaction methods may be more ideal than relying solely on gaze or gesture.

AR/VR environments are systems generated by technology to create scenes that allow for rich experiences, interactions are limited to the capabilities of head-mounted displays which, in many cases, are dependent on pointing devices like controllers, "However, for virtual reality, commanding devices which can be manipulated unseen are much preferred for example voice commands, lip reading, interpretation of facial expression and recognition of hand gestures" (Eisenberg, 2017). Gestures allow the dependency on controller devices to be removed, hence resulting in direct interaction between the user and environment similar to the implication of gaze, therefore, allowing this study to understand the feasibility of hands-free interaction for AR/VR.

Chapter 3: Methodology

Subjects

The subjects for this study were 15 engineers at the Kennedy Space Center location of the National Aeronautics and Space Administration (NASA) located in central Florida, USA. Participants consisted of an equitable representation of gender association, distributed evenly between the ages of 20 - 70.

A total of 20 employees were asked to voluntarily participate in the study through a verbal explanation of the purpose, duration, and overall procedure of the study. Fifteen (75%) employees had the availability to stop by the lab where the test was conducted during times of operation. All participants met criteria of low to high volume past usage of AR/VR technology, of experience with roles where the technology would be of use, and of having subject matter expertise to evaluate the prototypes; all consented to the conditions of the study.

Instruments

The instruments used to evaluate interactions were two prototypes and an open-ended question format, each individually administered to participants. Open-ended questions were used to understand foundational experience with AR/VR technology and qualitative perception of interacting with the control and experiment.

The initial open-ended questions instrument was administered first to reduce bias that could be produced in relation to participants' AR/VR knowledge. The following questions were verbally asked:

1. AR/VR technology exists in a wide variety of apps, tools, and software used for various purposes today. Have you interacted with AR/VR in the past? If so, how frequently?
2. Tell me about the AR/VR applications you have heard of or have experience with.
3. How did you interact with such technology? In what settings?
4. Have you used navigation-based controls in the past? How do you typically interact with this technology?
5. What are some of the main reasons you have taken these approaches?
6. When you were new to this technology, or if you have never used this technology, how would you prefer to learn how to interact?

Standard probes like "can you think of any other examples" and "tell me more about that" were asked to collect further information about the participants' experience. Participants' responses were noted and utilized for qualitative analysis.

Participants were then asked to evaluate two instruments in the form of prototypes that behaved as virtual control panels allowing a participant to control the movements of a robot as if it were a rover. One prototype served as a control, using traditionally exposed interaction controls, and the second prototype served as the experiment. The prototypes were created using an HTC Vive headset to create a Mixed Reality environment where participants could view interactions overlaid onto the real world (see Figure 1). A Leap Motion controller was attached to the front of the headset as a sensor device to capture gestural inputs and commands. To simulate the functionality of a rover, a mechanical robot was used with a sensor receiver to communicate with the inputs and 360fly camera (see Figure 2) for the participant to view the environment through the perspective of the robot while navigating.

Figure 1. AR/VR headset.

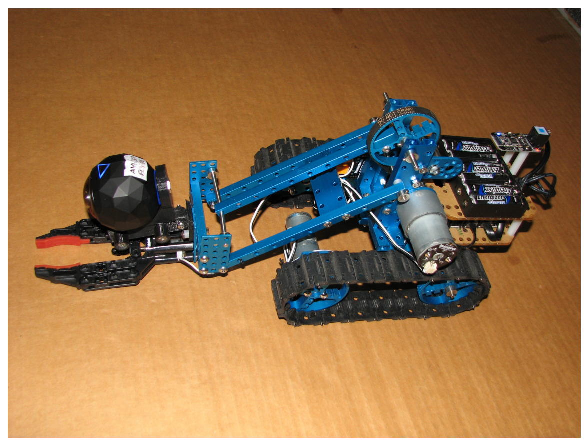

Figure 2. Mechanical robot with 360 fly camera.

The control prototype consisted of interactions all visually exposed to the participant by default. When a participant wore the headset, s/he would view what the robot was seeing along with buttons statically displayed that would allow the participant to move the robot forward, backward, left, and right (see Figure 3). Game objects are used as buttons in the virtual environment where box colliders create triggers prompted by the rigid body of a participant’s hands coming into contact with the floating surface area of each displayed interaction. Pointing a finger at a 45-degree angle and pushing a button overlay in the environment creates an event of the two bodies colliding to prompt the action/ commands available: left, right, forward, and reverse.

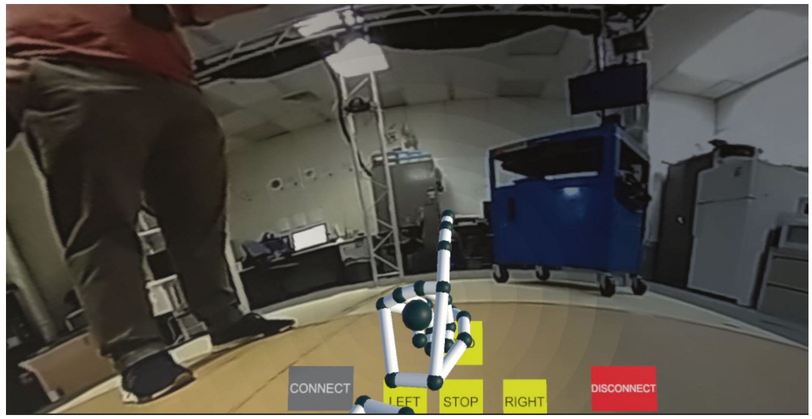

Figure 3. Control prototype.

The experimental prototype was intended to understand the gaze patterns a participant would utilize as a replacement for the static button controls. Due to the limitation in technology where the HTC Vive does not enable eye tracking, gestures were used to augment gaze direction. Gestures were calibrated utilizing pitch and yaw axis rotations where pitch represents the vertical axis and yaw represents the horizontal axis of one's gaze. The Leap Motion documentation (see Figure 1.) describes pitch as, "...the angle between the negative z-axis and the projection of the vector onto the y-z plane. In other words, pitch represents rotation around the x-axis" and yaw as, "...the angle between the negative z-axis and the projection of the vector onto the x-z

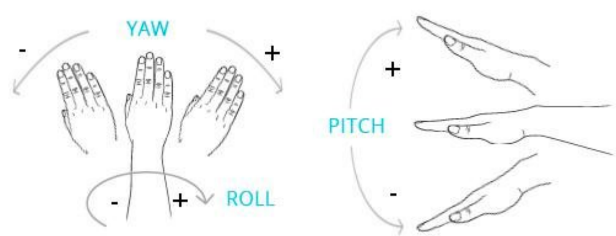

Figure 4. Leap Motion gesture documentation (ResearchGate).

plane. In other words, yaw represents rotation around the y-axis" (Vector). The pitch and yaw ranges were piloted and calibrated by testing wrist movements. Placement on the axis would correspond to an interaction prompt to control the robot. With the ability to detect between the right hand and the left hand, hand movements were used to replicate gaze patterns where right-hand movements would be used to control pitch, and left-hand movements would be used to control yaw. The ranges were defined as gestural inputs, "f" represents the float point data type to capture decimal values in the development environment (see Table 1).

Table 1.

Gestural Input Ranges

Pitch ranges (right hand)	Yaw ranges (left hand)
Stop: > = 0.8f, < = 1.2f	Stop: > -0.3f, < 0.3f
Forward: > = -0.2f, < = 0.2f	Right > = -0.7f, < = -0.3f
Backward: > = 1.8f, < = 2.2f	Left: > = 0.3f, < = 0.7f

Utilizing these definitions, this instrument presented users with a clear field of view where no buttons were present on the screen and the participant would utilize the gestural controls to direct the robot in the direction of his/her gaze.

In addition to the prototype instruments, participants were verbally asked a series of questions pertaining to their experience. The open-ended questions were used to consistently understand what participants' reactions and preferences were when interacting with the prototypes. During each session, participants were encouraged to think aloud, and notes were taken relating to their initial comments and answers to the following:

1. Based on your experience, what were your impressions about navigating using this prototype?
2. What did you learn about the rover/robot and its functions based on your interactions?
3. Was there anything else you would have expected to see? If so, what would you want to see and why?
4. On a scale from 1 - 5, 1 being very difficult and 5 being very easy. How would you rate this prototype? Why did you give it that rating?
5. How did this experience compare with other control panels that you have previously interacted with?
6. Which prototype version did you find the most useful? Why?
7. If you could change anything about the prototype what would it be? Why?
8. Do you have any remaining questions or concerns relating to the prototype? If so, what are they?
9. Based on your experience using the virtual control panel, how likely would you be to utilize it as a tool on a scale from 0 – 10, 0 being not at all likely and 10 being extremely likely?

Participants' responses to the follow-up questions were noted and utilized as a way to support observational findings.

Procedures and Tasks

All participants were tested at the AR/VR lab at the Kennedy Space Center where they are employed. Each evaluated the series of test instruments during single, 30-minute sessions. Tests were conducted in a lab room separate from work areas and desks where sound levels were controlled and surroundings were free of distractions.

Each participant was asked to participate in the testing area and provided with the following scenario:

"Today, you will be evaluating a Virtual Control Panel prototype to simulate navigating a rover on Mars. I want you to imagine that you are an astronaut on Mars and are controlling the actions of a rover from the comfort of the ship. This will assist you in exploring the terrain without exposure to potentially harmful situations. I want to mention a few things before we get started:

1. **Everything you share with me today will be confidential.** All information gathered during this session will be combined with other data to give us a big picture of what everyone had to say.

2. **I want you to feel free to speak your mind.** This isn't a test where there are right or wrong answers. We are here to listen to what you have to say and talk about your opinions.

3. **You will be evaluating a prototype.** This means that there will be only a few functions available to utilize and you may have to adjust your tap angle to ensure the input is accepted.

4. **This prototype is a draft.** I'm just here to collect your honest thoughts and opinions and you won't hurt my feelings if there are things that you don't like. Do you have any questions for me before we begin?"

No participants had concerns and were willing to take part in the study. When the session began, the study was prefaced with: "Before we begin, I'd like to learn more about you," followed by the open-ended questions to understand the level of experience each employee had with AR/VR technology.

Upon completion of the questions, the participant was assisted with comfortably wearing the HTC Vive headset and the following scenario was explained: "Let's say that you were on Mars and wanted to interact with objects found on the surface. In a moment, you will interact with a prototype where you will be instructed to find your way around the terrain. You will see a series of actions present within your view. Use these actions to explore with the rover. Navigate and let me know what you see. It would be very helpful if you could provide comments about the prototype as you are interacting with it." Participant comments and impressions were noted. Shortly after, the participant was instructed to navigate the robot from one object placed in the environment to another: "Now that you have learned how to control the rover, navigate from the 3D printed Delta IV rocket (see Figure 5) to the 3D printed V2 (see Figure 6). It would be very helpful if you could provide comments about the prototype as you are interacting with it." After completion, participants' ability to successfully carry out the task were noted followed by open-ended questions about their impressions of navigating using the prototype.

Figure 5. Delta IV.

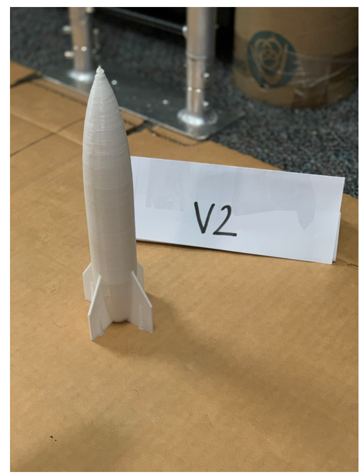

Figure 6. V2.

Next, participants were instructed to interact with the experiment prototype instrument where they were presented the same scenario as the control, but with different interaction instructions: "You will now evaluate an alternative version of the prototype. In this version, you will only use hand gestures to control the rover's behavior. Hand gestures will be used to simulate your eye movement. When using your hands, make sure only one hand is within view at a time. With your palm facing down, use your right hand and move your wrist up to move forward, and down to move backward. With your palm facing down, use your left hand and move your wrist right to move right, and left to move left. It would be very helpful if you could provide comments about the prototype as you are interacting with it." Participant perceptions were noted in the same way as the first prototype followed by the same navigational task utilizing the new interaction pattern. To conclude the study, participants were asked open-ended questions to determine preference among the prototype instruments, qualitative perception of interaction inclination, and Likert rating of ease of use and likelihood of utilization.

Analysis Plan

The intention of the research was to fulfill four analyses focused on qualitative measures.

1. Precursory analyses were conducted to evaluate pre-existing knowledge and level of exposure to AR/VR technology and interaction patterns. Notwithstanding that this type of evaluation is prone to error, the foundational examination was focused on contextual exposure an employee had both within his/her role at NASA and outside of their field of work. Understanding the correlation between a participant's experience and task success was imperative in factoring out pre-existing bias.

2. The prototype analyses concentrated on the perceived ease of use and the ability to successfully navigate using the interactions present in the control instrument. Assessments were based on the efficiency of task completion, success or failure of navigation, and participants' qualitative responses to the prototypes.

3. Analyses of the open-ended question responses were read and grouped against all qualitative findings to identify correlations between experience, success, and perceived interaction preference among participants.

Verbally expressed phrases and thoughts were listed as findings throughout the evaluation. Identifying a list of responses to each question and task served as units of measure where, at times, multiple responses were elicited resulting in a greater number of responses than participants in the study. Each response was categorized by similarity in an iterative process. Utilizing a grouping mechanism allowed responses to create trends in findings based on discovery rather than predetermined importance. In addition to qualitative response analyses, qualitative Likert scores were averaged across participant response regarding ease of use for each overall prototype experience. The outcome was a weighted rating of interaction effort, trend in preference, and categorized themes of responses.

Chapter 4: Results

The research contained in this study was intended to investigate the feasibility of utilizing gaze as a primary interaction pattern when utilizing augmented/virtual reality systems, specifically within the industry of aerospace. Instead, the usefulness of gestural support for gaze interaction was explored. The following hypotheses were tested:

Hypothesis One: Replacing button controls with hand gestures to control AR/VR motion has the potential to solve interaction efficiency challenges while ensuring users have a natural and seamless experience.

Hypothesis Two: Participants will readily adapt to hand gestures to control AR/VR motion if the hand gestures utilize natural mappings.

Hypothesis Three: Reducing cognitive load by contextualizing menus and interaction-rich experiences allows participants to focus on the environment at hand and reduces the reliance on ergonomic mobility to discover and explore.

To address these inquiries, NASA employees were recruited from the Kennedy Space Center office located in Brevard County, Florida. The previous methodology chapter outlined the process of recruitment resulting in 15 participants in the study where data was collected in the AR/VR laboratory. Participant ages ranged from 20 to 70 years old where 3 were female (20%) and 12 were male (80%) partaking in the study.

The results of the study are outlined in four sections below, consisting of: demographics of the sample, performance of the sample, interaction analysis of preferences based on comparison between control and experiment prototypes, and qualitative analysis based on participant responses to open-ended questions.

Demographics of the Sample

The following table outlines the demographic attributes of participants who took part in the study. Values are presented as percentages based on a value out of the total sample "N" where N=15 (see Table 2). This data was collected as both information based on previous knowledge about the employees and open-ended questions asked prior to participation.

Table 2.

Demographics of the Sample

Attribute	N	%
Gender of participant		
Male	12	80%
Female	3	20%
Role/occupation of participant		
Software development engineer	4	26.7%
Engineer	4	26.7%
Strength analyst	3	20%
Launch services engineer	2	13.3%
Aerospace engineer	1	6.7%
Mechanical engineer	1	6.7%
Past experience with AR/VR		
Yes	13	86.7%
No	2	13.3%
Frequency of experience with AR/VR		
Daily	6	40%
Weekly	3	20%
Monthly	2	13.3%
Annually	2	13.3%
N/A	2	13.3%
AR/VR devices previously used		
HTC Vive	5	33.3%
Oculus	3	20%
Magic Leap	2	13.3%

Google Daydream	3	20%
Smartphone	8	53.3%
Purpose of use		
Profession	5	33.3%
Gaming/leisure	8	53.3%
Education	3	20%
Demonstration/simulation	6	40%

As reported, a majority of participants were software development (26.7%) or general engineers (26.7%) and had interacted with AR/VR technologies in the past (86.7%). Responses relating to frequency and type of device used were skewed towards daily interaction (40%) and Smartphone (53.3%) followed by the HTC Vive headset use (33.3%). Approximately half of the participants utilized AR/VR for gaming and leisure purposes (53.3%) which ultimately contributed to the frequency of use.

Performance of the Sample

Only success rate was captured for the four tasks assigned to participants. Values like time to complete and margin of error were not measured. Among the 15 participants, all (100%) were able to successfully fulfill the basic navigational interactions of tasks one and three where they were instructed to move the rover utilizing buttons or gestural inputs respective to each prototype.

Tasks two and four proved a bit more challenging, and the success and failure rates varied among users. Success rate was higher for the control prototype compared to the experimental prototype; calculated by denoting a binary 1 (success) or 0 (failure) for completion where the control prototype was (86.7% on average) with a confidence interval of 0.8667 ±0.172 and standard deviation of 0.34; completion rate for the experimental prototype was (73.3% on average) with a confidence interval of 0.7333 ±0.224 and standard deviation of 0.44. However, perceived ease of use, based on a Likert scale average of 1 (very difficult) to 5 (very easy), was lower for the control (2.93) with a confidence interval of 3.13 ± 0.445 and standard deviation of 0.88 compared to the experimental prototype (3.73) with a confidence interval of 3.8 ±0.46 and standard deviation of 0.91.

Interaction Analysis

Interaction was measured by observing the participants' comments and interaction efficiency when testing each prototype. When interacting with the control, participants were required to point a finger at a 45 degree angle and to utilize a pushing motion to interact with the buttons displayed in the environment. When used in practice, participants were generally successful in initiating desired actions. Margins of error when attempting to "push" the buttons fell into three thematic categories: success in tapping the targets, accuracy of angle, and confusion caused by lack of feedback. Tapping targets proved challenging as participants parsed various actions on the screen. Often miss-taps occurred when another finger would interact with an undesired button while an index finger was reaching towards a particular target button, resulting in delay where a user had to engage in an unintended interaction in order to undo the mistake. Mistakes with accuracy of angle would cause a participant to attempt pushing buttons multiple times in order to elicit a response from the robot. It was difficult for users to achieve a 45 degree finger trajectory upon tapping the virtual buttons. Although explicit calibration training may have helped with angle accuracy, the interaction was clearly not intuitive or easy to master.

Lastly, the lack of feedback when pushing a button resulted in participants tapping multiple times, since the virtual button did not display a “pushed in” visual treatment to allow the user to understand that an interaction was in progress. This is an interaction design flaw that could presumably be rectified easily; however, the visual feedback for the experimental prototype was natural and intuitive and thus required no special planning or design complexity. Due to this lack of button-based feedback, participants made multiple attempts to “press” buttons even when the robot had already started to move and the VR simulation showed movement.

Interacting with the experimental prototype was perceived as much more intuitive than the control due to a more natural mapping relationship and a more immediate recognition of robot response to movement directions. Although hand gestures were used as a way to augment gaze patterns, participants reported that the interactions felt smoother and easier due to a match in cognitive expectations. Participant challenges were categorized as ergonomic strain and lack of visual cues. When utilizing hands to initiate interactions, participants reported strain in the range of motion of their wrists and muscle fatigue when attempting to interact in a particular way for an extended period of time. It is possible that eye strain may also occur when utilizing gaze patterns to initiate similar actions. Lack of visual cues were also reported as challenging since participants were viewing navigation of the robot through the lens of the camera, so that understanding orientation and overall location sometimes required additional cognitive load for the participant to synthesize location and position in order to plan their next steps.

Although both the control and experimental prototypes proved to have a unique set of challenges for the participants, preference of interaction type skewed towards the experiment (93.3%) where participants reported a high level of agency and ease due to the more intuitive nature of the interaction type.

Qualitative Analysis

Qualitative perception of the participants were collected by open-ended questions asked after interacting with the prototypes. Overall impressions were that the intentions of each prototype were straightforward and were grasped manageably. A majority of participants (80%) expressed issues with the latency in interaction delivery between the system and the robot that was a result of communication interruptions within the technology.

In addition to preferring the experimental prototype, participants expressed a desire for more robust controls, such as a way to throttle the speed of the robot, with the expectation that a higher level of navigational accuracy could be achieved by increasing or reducing the amount of power exerted by the movement of the robot. Due to the high level of familiarity with gaming among participants, there was an expectation of joystick-like movements and predilection for reduced rigidity, where angled directions would be possible to mimic the general steering functionality often present in real-world vehicles.

Chapter 5: Conclusion and Recommendations

The intention of this research was to understand the feasibility of gaze as a primary input when interacting with AR/VR technology, specifically with an emphasis on utilization in the context of aerospace. Discussion around the research discovery will focus on the synthesis of findings and their implications on testing the hypotheses established prior to the experimentation.

Test of Hypotheses

Hypothesis One. Hypothesis One defined the ability for hand gestures to increase interaction efficiency by allowing a real-world application of interaction inputs to exist where a user's gestural direction would signify a parallel in navigating through AR/VR. The foundational premise for this hypothesis was proven when observing participants interacting with both prototypes; participants were able to initiate their desired inputs more accurately with the gestural motions rather than static buttons in the virtual environment. This was also apparent in the ease of use reported by participants where the gestural prototype received a Likert scale average of 0.8 more ease compared to the control. Although the time on task of participants was not measured during the study, the qualitative interpretation of efficiency may be due to the absence of ergonomic restrictions found in physical initiations as reported by Stellmach and Dachselt (2012).

Hypothesis Two. Hypothesis Two stipulated the adaptability of participants to utilize hand gestures when natural mappings reflected consistency in desired inputs. This hypothesis was partially successful as participants easily learned directional cues due to the similarity in orientation. The hypothesis was partially unsuccessful due to the response time of the system, glitches in the communication between the robot and the inputs received from the user, and the fish-eye effect of the 360 fly camera which made gauging perception and distance less accurate. As a result, participants questioned the accuracy of their gestures and increase the number of attempts. Referencing Satriadi et al.'s (2019) study findings show that the introduction of hybrid methods where participants use gestures and hand controllers, "provide a good compromise in performance, comfort, and ease of learning between the use of either method alone."

Hypothesis Three. Hypothesis Three emphasized the reduction of cognitive load through a focused display and ergonomically eased input initiation. The intention was that a more clear FoV with no static buttons or menus would decrease choice paralysis and increase the ability of a participant to interact with the virtual environment in a more focused manner. This hypothesis was proven true through opinions expressed aloud while participants navigated their tasks and responses to open-ended questions. Similar to hypothesis one, the reported ease of use and participant feedback, "Navigating is pretty good, and environment understanding is good." Although a quantitative margin of interaction error was not recorded, participants expressed preference for the experiment and experienced fewer occurrences of inaccurate inputs to navigate the robot.

Recommendations

This study exhibited limitations in technologies used to deploy the control and experiment. Such technologies include the HTC Vive, Leap Motion, 360 Fly camera, and motor robot to explore navigational ease based on gaze patterns. Gestural inputs were used as a foundation to understand how gaze patterns may similarly be perceived since the HTC Vive did not support eye tracking. Due to the deprecation of specific gesture commands for the Leap Motion, the calibration and range of hand motions that were utilized as a replacement for eye tracking were also limited in accuracy and range. The 360 Fly camera distorted user perception of the environment resulting in some confusion around object distance and robot location relative to a given target. The motor robot functioned as expected, yet exhibited some latency challenges that resulted in communication lags between the headset and command system. It is imperative to address the gaps in measurement by utilizing input devices that provide both eye tracking ability and control as a replacement for static inputs, and an upgraded robot camera that would allow users to more accurately understand where the robot is within the physical environment.

As indicated by interaction design research for AR/VR environments, specifically for navigation-based instances, certain visual cues may guide a user's attention toward important information or ease the burden of detecting visual targets (Rusch et.al., 2013). This was especially apparent with the distortion of location perception for participants in the study, a thematic takeaway during observation and feedback was the desire for arrows and distance indicators to facilitate precision when carrying out tasks in the virtual environment. Jonides (1980) reported cues that could contribute to easier navigation like “dynamic blinking” to successfully draw one’s attention, while such cues may help guide participants, the implication of a cluttered FoV would need to be further explored. Yeh and Wickens (2001) found that accuracy increased when noticing a target during aviation exploration tasks when cues were used to make objects more visible when scanning an environment. Fisher and Tan (1989) discovered that participants could locate targets more quickly when highlights were used in displays compared to those without highlighting. Knowing this, the absence of visual cues in both the control and experiment of this study may have contributed to a lower perception of ease of use and decrease in performance of the participants. Creating an experimental environment where navigational tasks are clearly defined by arrows and directional cues in VR may allow users to focus more easily on the input controls rather than detecting a navigational direction and interaction to continue, thus allowing the study to be focused on interaction preference instead of understanding navigational direction and orientation.

Understanding the correlation between the threshold of gaze fixation and user intention is important for both accurately applying interactions that were intended by the user and for preventing potential ergonomic fatigue. Relying on one primary input method could cause muscle strain when exercised for extended periods of time. Utilizing gaze as the sole input could perhaps result in such strain unless fixations reflect and support patterns that are considered natural to the participant. Due to the lack of access to such technology and ability to collect time as a quantitative measure of interactivity, it is recommended that such dimensions are explored in further research.

Foundational findings have contributed to a further understanding of further implications that should be considered and observed when researched. The objective of this study was to investigate the possibility of utilizing gaze as a primary interaction for AR/VR technology. This remains as an outstanding question that can only be accurately acknowledged with proper instruments and targeted methodologies, yet the foundational findings relating to a preference in hand gestures over static buttons establishes a premise for the future of hands-free interaction in AR/VR.

References

Augmented Reality – The Past, The Present and The Future. (2017, March). [web log]. https://www.interaction-design.org/literature/article/augmented-reality-the-past-the-present-and-the-future.

Boland, M. (2018, February 21). Enterprise Xr's Bottom Line Impact. *ARtillry*. https://arinsider.co/2018/02/21/artillry-briefs-episode-13-enterprise-xrs-bottom-line-impact/.

Cámara, C., de la Malla, C., López-Moliner, J., & Brenner, E. (2018). Eye movements in interception with delayed visual feedback. *Experimental brain research*, 236(7), 1837–1847. https://doi.org/10.1007/s00221-018-5257-8

Chen, Z., & Shi, B. (2018). Using Variable Dwell Time to Accelerate Gaze-Based Web Browsing with Two-Step Selection. *International Journal of Human-Computer Interaction*, 240–255.

Dean, H. L., Martí, D., Tsui, E., Rinzel, J., & Pesaran, B. (2011). Reaction time correlations during eye-hand coordination: behavior and modeling. The Journal of neuroscience: the official journal of the Society for Neuroscience, 31(7), 2399–2412. https://doi.org/10.1523/JNEUROSCI.4591-10.2011

Eisenberg, A. (2017, November 14). Gesture Recognition Technology and Its Importance in VR. https://appreal-vr.com/blog/gesture-recognition-in-virtual-reality/

EyeTracking, I. T. H. of E. T. (n.d.). The EyeTracking Blog. http://www.eyetracking.com/News/EyeTracking-Blog/EntryId/57/King-Midas-and-the-Golden-Gaze.

Hansen, J. P., Johansen, A. S., Hansen, D. W., Itoh, K., & Mashino, S. (2003). Command Without a Click: Dwell Time Typing by Mouse and Gaze Selections. *Human-Computer Interaction - INTERACT'03,* 121-128.

Jacob, R. J. (1995). Eye tracking in advanced interface design. Virtual Environments and Advanced Interface Design, 258-288.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye's movement. *Atten. Perform.*, 187–203.

Liao, W. and Yu, L. (2010). Robust pupil detection for gaze-based user interface. In Proceedings of the 2010 workshop on Eye gaze in intelligent human machine interaction. *EGIHMI '10,* 45-50. https://doi.org.proxy-ub.researchport.umd.edu/10.1145/2002333.2002341

Maddox, T. (2017, November 14). Training for Retention in Virtual Reality and Computer-Based Platforms. https://trainingindustry.com/articles/learning-technologies/training-for-retention-in-virtual-reality-and-computer-based-platforms/

Maddox, T., & Campbell, P. (2017, December 20). Muscle Memory and the Brain Science Behind VR. https://trainingindustry.com/articles/learning-technologies/muscle-memory-and-the-brain-science-behind-vr/

Moberg, W., & Pettersson, J. (2017). (thesis). *Interacting with Hand Gestures in Augmented Reality: A Typing Study*. Blekinge Institute of Technology. http://www.diva-portal.org/smash/get/diva2:1138555/FULLTEXT02

Natural user interfaces for human-drone multi-modal interaction - Scientific Figure on ResearchGate. https://www.researchgate.net/figure/Hand-gesture-commands-used-for-drone-flight-In-this-setup-positive-pitch-roll-and-yaw_fig11_304816895

Ng, A., Lepinski, J., Wigdor, D., Sanders, S., & Dietz, P. (2012). *Designing for Low-Latency Direct-Touch Input*. Tactual Labs. https://www.tactuallabs.com/papers/designingLowLatencyDirectTouchInputUIST12.pdf.

Planthaber, S., & Maurus, M. (2017). Controlling a Semi-Autonomous Robot Team From a Virtual Environment.

Rainey, K. (2017, August 3). Robonaut's potential shines in multiple space, medical and industrial. https://www.nasa.gov/mission_pages/station/research/benefits/robonaut/.

Rusch, M. L., Schall, M. C., Jr, Gavin, P., Lee, J. D., Dawson, J. D., Vecera, S., & Rizzo, M. (2013). Directing driver attention with augmented reality cues. *Transportation* *research. Part F, Traffic psychology and behaviour*, *16*, 127–137. https://doi.org/10.1016/j.trf.2012.08.007

Salvucci, D., & Anderson, J. (2000). Intelligent gaze-added interfaces. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems. *CHI '00*), 273-280. https://doi.org.proxy-ub.researchport.umd.edu/10.1145/332040.332444

Satriadi, K. A., Ens, B., Cordeil, M., Jenny, B., Czauderna, T., & Willett, W. (2019, August 15). *Augmented Reality Map Navigation with Freehand Gestures*. https://ieeexplore.ieee.org/document/8798340/authors#authors

Spencer, S. N. (2005). Learning through computer-generated visualization. *ACM SIGGRAPH Computer Graphics (ACM Digital Library)*, *39*(1), 2. https://doi.org.proxyub.researchport.umd.edu/10.1145/2168556.2168577

Stellmach, S., & Dachselt, R. (2012, March). *Designing gaze-based user interfaces for steering in virtual environments*. ACM Digital Library. https://doi.org/10.1145/2168556.2168577.

Sutherland, I. E. (1968). *A Head-Mounted, Three-Dimensional Display*. AFIPS Proceedings of the Fall Joint Computer Conference, Part I, 757-764.

Tate, S. (1996). Virtual Reality: A Historical Perspective. http://ei.cs.vt.edu/~history/Tate.VR.html

Vector. (n.d.). https://developer-archive.leapmotion.com/documentation/java/api Leap.Vector.html#javaclasscom_1_1leapmotion_1_1leap_1_1_vector_1a781d0ab346c40cb3b3c6ee550b933acd

Yeh, M., & Wickens, C. D. (2001). Display signaling in augmented reality: effects of cue reliability and image realism on attention allocation and trust calibration. *Human factors*, *43*(3), 355–365. https://doi.org/10.1518/001872001775898269

Appendix A: Consent Form

This research study is for a thesis at the University of Baltimore. If you choose to participate in this study, you will interact with objects in a virtual environment using three different methods: by looking at things with all the controls visible, by clicking to see the controls, by controls dynamically appearing based on how long you look at things. By clicking on this link, you agree to provide feedback about these interaction methods, including measured dwell times of your gaze within the virtual environment and you confirm that you are over the age of 18. Your participation is voluntary and you may withdraw from the study at any time without penalty.

The study will take about 30 minutes to complete where you will be asked to follow step-by-step instructions on tasks to perform while interacting with a prototype simulating the navigation of a rover on Mars. Your participation will benefit research for new interaction methods for emerging AR/VR technologies. Note, that interacting with AR/VR technology may result in eye strain, motion sickness, anxiety, or related reactions. If you have any sensitivities to these, you should decline to participate.

Your feedback will be confidential and all results will be reported anonymously. No identifying information will be collected or recorded. The only data collected will be observational notes for reference purposes and the thesis report will be an aggregate of all participant feedback. Notes will be password protected and destroyed once the project is complete.

The aggregated, anonymous information collected in this study may be used for future research studies or distributed to another investigator for future research studies.

If you have questions about this study, please contact the researcher Busra Demirci at 561-376-7786, or busra@demirci.com or Dr, Kathryn Summers at ksummers@ubalt.edu or University of Baltimore IRB coordinator at (410) 837-4057, or irb@ubalt.edu.

Are you willing to participate in this study? **Yes** **No**

Appendix B: Study Instrument

Test Overview & Methodology

Overview

The study will employ one-on-one in-person testing sessions. Participants will be asked to review the live AR/VR prototype for the Virtual Control Panel using an HTC Vive and Leap Motion and asked to answer a series of questions pertaining to their experience. The moderator will use this discussion guide as an overall guide but will not be followed verbatim so as to allow for other relevant discussions that may arise.

Research Approach

Session Agenda

Introduction (<5 mins)

Background Questions (5 mins)

- What is their role and occupation?
- How do they evaluate controlling equipment?
- What information would they find useful?

Prototype Evaluation (15-20 mins)

- How do they interact with the prototype?
- What are their thoughts and opinions?

Wrap up (5 mins)

- What were their overall thoughts about the prototype?
- What did they think about interacting with the virtual control panel based on their experience?

Introduction and Preliminary Discussion

I'd like to thank you for spending part of your day talking to me. Today, you will be evaluating a Virtual Control Panel prototype to simulate navigating a rover on Mars.

I want you to imagine that you are an astronaut on Mars and are controlling the actions of a rover from the comfort of the ship. This will assist you in exploring the terrain without exposure to potentially harmful situations. I want to mention a few things before we get started:

1. **Everything you share with me today will be confidential**. All information gathered during this session will be combined with other data to give us a big picture of what everyone had to say.

2. **I want you to feel free to speak your mind.** This isn't a test where there are right or wrong answers. We are here to listen to what you have to say and talk about *your* opinions.

3. **You will be evaluating a prototype**. This means that there will be only a few functions available to utilize and you may have to adjust your tap angle to ensure the input is accepted.

4. **This prototype is a draft**. I'm just here to collect your honest thoughts and opinions and you won't hurt my feelings if there are things that you don't like.

Do you have any questions for me before we begin?

Background Questions

I'd like to learn more about you.

1. Please describe your role and occupation.
2. Have you interacted with AR/VR in the past? If so, how frequently have you had exposure to the technology?

AR/VR technology exists in many apps and software today. I'd like to hear more about your experiences with this technology.

3. Have you used any AR/VR applications in the past? If so, which ones? How did you interact with them?
4. In what settings do you typically interact with navigation-based controls? How do you typically interact with the inputs?

5. What are some of the main reasons that you take this approach?

6. If you were interacting with a Virtual Control Panel, what information would you first want to see?

7. How would you prefer to learn more about how to interact?

Control Prototype Evaluation

Task 1

Instructions:

Let's say that you were on Mars and wanted to interact with objects found on the surface. In a moment, you will interact with a prototype where you will be instructed to find your way around the terrain. You will see a series of actions present within your view. Use these buttons to explore with the rover. Navigate and let us know what you see. It would be very helpful if you could provide comments about the prototype as you are interacting with it.

Observations:

- What were some of their initial comments?
- What information do they spend the most time reviewing?
- How far did they navigate?

Post-task questions:

Based on your experience:

- What were your initial impressions about navigating?
- What did you learn about the rover and its functions?
- Was there anything else you would have expected to see? If so, what would you want to see and why?

Task 2

Instructions:

Now that you have learned how to control the rover, navigate from the 3D printed Delta IV rocket to the 3D printed V2. It would be very helpful if you could provide comments about the prototype as you are interacting with it.

Observations:

- What were some of their initial comments?
- What information do they spend the most time reviewing?
- How far did they navigate?

Post-task questions:

Based on you experience:

- What were your initial impressions about navigating?
- What did you learn about the rover and its functions?
- Was there anything else you would have expected to see? If so, what would you want to see and why?
- On a scale from 1 - 5, 1 being very difficult and 5 being very easy. How would you rate this prototype and why?

Experiment Prototype Evaluation

Task 1

Instructions:

You will now evaluate an alternative version of the prototype. In this version, you will only use hand gestures to control the rover's behavior.

When using your hands, make sure only one hand is within view at a time. With your palm facing down, use your right hand and move your wrist up to move forward, and down to move backward. With your palm facing down, use your left hand and move your wrist right to move right, and left to move left. It would be very helpful if you could provide comments about the prototype as you are interacting with it.

Observations:

- What were some of their initial comments?

- What information do they spend the most time reviewing?
- How far did they navigate?

Post task questions:

Based on you experience:

- What were your initial impressions about navigating?
- What did you learn about the rover and its functions?
- Was there anything else you would have expected to see? If so, what would you want to see and why?

Task 2

Instructions:

Now that you have learned how to control the rover with this version, navigate from the 3D printed Delta IV rocket to the 3D printed V2. It would be very helpful if you could provide comments about the prototype as you are interacting with it.

Observations:

- What were some of their initial comments?
- What information do they spend the most time reviewing?
- How far did they navigate?

Post task questions:

Based on you experience:

- What were your initial impressions about navigating?
- What did you learn about the rover and its functions?
- Was there anything else you would have expected to see? If so, what would you want to see and why?
- On a scale from 1 - 5, 1 being very difficult and 5 being very easy. How would you rate this prototype and why?

Debriefing

Final Questions

- Overall, what were your impressions of the prototypes?
- What information about this virtual control panel did you find most useful and/or interesting?
- How did this experience compare with other control panels that you have previously interacted with?
- Which prototype version did you find the most useful? Why?
- If you could change anything about the prototype what would it be? Why?
- Do you have any remaining questions or concerns relating to the prototype? If so, what are they?
- Based on your experience using the virtual control panel, how likely would you be to utilize it as a tool on a scale from 0 – 10, 0 being not at all likely and 10 being extremely likely?