Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Architecture and Design
Department of Design

Vegard Wilson Dahl

Efficient use of Signifiers in VR Environments

Master's thesis in Interaction Design
Supervisor: Ole E. Wattne
June 2021

NTNU
Norwegian University of Science and Technology

Vegard Wilson Dahl

Efficient use of Signifiers in VR Environments

Master's thesis in Interaction Design
Supervisor: Ole E. Wattne
June 2021

Norwegian University of Science and Technology
Faculty of Architecture and Design
Department of Design

NTNU
Norwegian University of
Science and Technology

Abstract

As part of a master's thesis at NTNU Gjøvik, spring 2021, a study was carried out examining signifiers in Virtual Reality (VR). The study examined the effect visual and haptic signifiers had on user's mental workload depending on their prior experience in VR. Though based in training simulations, and helping users distinguish certain objects or items, the test application was changed to a puzzle-like task. Users were tasked in guessing the combinations to five keypads, utilizing different kinds of signifiers. They were: None, proximity-based visual positive signifiers, as well as positive/negative, and constant visual positive. Haptic signifiers in the controllers were also added as the technology allowed for it. The users were placed into three different groups, depending on their previous experience with VR, and they were subjected to verbal RTLX surveys after each task to measure their mental workload. These data were analyzed in SPSS to measure their means, significance, and correlation. From the results, it is clear that visual signifiers helped users complete their tasks, and haptics as well. But it is unclear whether the user's prior experience had any effects on their workload without further testing, as the correlations were too weak.

Sammendrag

Som en del av en masteroppgave ved NTNU Gjøvik, ble en studie gjennomført våren 2021 angående signifiers i Virtual Virkelighet (VR). Studien undersøkte hvilke effekter visuelle og haptiske signifiers hadde på brukeres mentale arbeidslast, avhengig av deres tidligere erfaringer med VR. Dog opprinnelig basert på treningsapplikasjoner for VR, i å hjelpe brukere skille mellom viktige objekter og gjenstander, så ble testen endret til en mer utfordringsbasert oppgave. Deltagerne ble bedt om å gjette kombinasjonene til fem kodelåser, ved hjelp av ulike typer signifiers. Disse var: Ingen, nærhetsbasert visuelle positive og positive/negative signifiers, og konstant visuelle signifiers. Haptiske signifiers i selve kontrollerne ble også lagt til da teknologien gjorde dette mulig. Deltagerne ble plassert i tre forskjellige grupper, basert på deres tidligere erfaringer med VR. De ble bedt om å gjennomføre muntlige RTLX skjemaer etter hver oppgave for å måle deres mentale arbeidslast. Disse dataene ble analysert i SPSS for å måle gjennomsnitt, signifikans, og korrelasjon. Fra resultatene så er det klart at visuelle signifiers hjelper brukerne i å fullføre oppgavene, haptiske signifiers også til en mindre grad. Men det er uklart om brukernes tidligere erfaringer har noen effekt på arbeidslasten deres uten videre undersøkelser, siden de statistiske korrelasjonene er for svake.

Preface

This project was written as a master's thesis in Interaction Design at NTNU Gjøvik in the spring of 2021. Parts of the study and research were started in fall of 2020 as a preparatory course. I wanted to do something with Virtual Reality, as I felt it was a field I had contacts, and somewhere I felt there were opportunities. I learned a lot during this time, and I hope to put this new knowledge to good use in my career going forwards.

I would like to extend some thanks to the people who helped me through this project:

Thanks to my supervisor, Ole, for his feedback, advice, and good help during the last year.

Thanks to the boys at Making View for helping me find a field of study. And especially my friend Håkon, for helping me code the prototype.

Thanks to my girlfriend Gabriela, for her support, care, and companionship.

Thanks to Mom and Dad, for letting me turn their home into a makeshift VR lab during easter.

And thanks to the rest of the MIXD class for their camaraderie during the pandemic and thesis.

Contents

List of Figures

List of Tables

List of Abbreviations

NTNU	The Norwegian University of Science and Technology
VR	Virtual Reality
AR	Augmented Reality
XR	Crossed Reality
MR	Mixed Reality
HMD	Head Mounted Display
MW	Making View
PC	Personal Computer
2D	Two-dimensional
3D	Three-dimensional
NASA	National Aeronautics and Space Administration, U.S.A.
TLX	Task Load Index
RTLX	Raw Task Load Index
XP	Experience
K”X”	Keypad “X”
HUD	Heads-up-display
COVID-19	Corona Virus Disease 2019
NSD	Norwegian Centre for Scientific data
IBM	International Business Machines
SPSS	Statistical Product and Service Solutions

1 Introduction

1.1 Topic overview

The concept of virtual reality has been around for a long time, becoming a staple of science fiction and staying in the public imagination as futuristic tech. First developed in the 1960's, mainly for simulator purposes, the technology was steadily developed throughout the 20th century. Though the technology did not become widely available before the 2010's, because of previously low graphical capabilities (Wohlgennant, Simons and Stieglitz, 2020, p. 456).

Virtual reality (also abbreviated VR) has seen tremendous use in video games the last years, with many studios having embraced and focused on this technology. However, VR has also been used in more practical settings, like serious games made for training or education. Or as interactive storytelling experiences (Wohlgennant, Simons and Stieglitz, 2020, pp. 457-458). Such experiences can be great tools, as it is possible to place users/trainees in realistic simulated scenarios in an immersive manner. Therefore, it is important to design these experiences in such a way that minimizes distractions and frustrations.

This study aims to see how visual signifiers can help users lower their mental workload and identify key objects in an environment. And discuss whether this can be applied to real immersive training scenarios.

1.2 Justification and contribution

The inspiration for this project came from a meeting with some employees at Making View. Making view (hereby abbreviated "MW") is a company based in Hamar, utilizing game mechanics and technology to make virtual experiences for several clients. These experiences are being used for training, education, and more. After having a meeting with some employees, issues surrounding usability and interaction design were discussed. According to MW, many new users struggle with virtual reality at first.

Since many of MW's projects revolve around education or training, their VR-experiences are used by several users from different fields and with dissimilar skillsets. They often run into users feeling disconnected due to the controllers connected to the headset, or they struggle to interact with the 3D environment, struggling with distinguishing what elements they can interact with, and which they cannot. This could help users find critical and important items in virtual reality scenarios, and potentially help bridge the gap between the virtual and real world.

The goal for this paper is to contribute to the field of educational/instructional VR experiences. It aims to do so by helping establish the effectiveness of different kinds of signifiers in distinguishing important elements, and how they perform towards users of varying expertise.

1.3 Paper structure

This paper will be divided into 6 main chapters, following a modified IMRaD structure. The chapters consist of *introduction, background, methods, practical and ethical concerns, results, discussion, and conclusion*.

1.4 Research questions and hypothesises.

This study will explore the following research questions:

1.4.1 Research questions:

1. How do visual and haptic signifiers impact user's mental workload in VR experiences?
2. What kind of signifiers offer the least mental workload toward the user?
3. How does different kinds of signifiers impact mental workload in users of different experience levels?

1.4.2 Hypothesises:

Visual signifiers in diegetic interfaces will greatly reduce workload, especially concerning Mental demand.

Inexperienced users will respond with a lower mental workload better to highlighted and color-based stimuli/signifiers.

1.5 Keywords

Virtual reality, VR, Affordances, Signifiers, usability, workload, TLX, ….

2 Theory and Background

2.1 Virtual Reality

2.1.1 Virtual Reality background

The term Virtual Reality refers to a technological concept, where a user is subjected to an immersive virtual environment. Most commonly by the help of a mounted headset with screens fitted for each eye. These types of headsets are commonly referred to as VR-headsets, or Head Mounted Displays. Some of these headsets also use paired wireless controllers, to allow the user to interact with objects in the virtual space as well.

In their article, *"Virtual Reality",* Wohlgennant, Simons and Stieglitz (2020), discuss the history and applications of Virtual Reality as a technology. They define Virtual Reality as a concept is usually characterized by three factors, presence, interactivity, and immersion. Eirik Helland Urke is a Norwegian Journalist and editor, working for the technical magazine teknisk ukeblad. Urke has written an introductory book on the field of virtual and augmented reality with help from crossed reality developers throughout Norway (Urke, 2018, pp. 9-10). The book is meant to give insight and inspire new developments within Crossed reality in Norway. Urke also describes the basic technologies and inner workings of most forms of Crossed reality. Crossed reality or Extended reality as Urke calls it, is a collective name for virtual and augmented reality, as well as relevant technologies. Crossed reality is often abbreviated to "XR", as the X presents the unknown variable, be it augmented or virtual.

He describes several subfields and possibilities in his book. He defines Virtual Reality is defined as a series of technologies that create a virtual environment through digitally produced sensory inputs, most often in the form of stereoscopic video and audio. The goal of virtual reality is to achieve the feeling of presence, where the user feels fully immersed in the virtual world. This is not to be confused with immersive effect, which describes a technology's effect of helping the user achieve the feeling of presence (Urke, 2018, pp. 21-23). These technologies can range from relatively simple 360° video to VR glasses with bigger immersive effect. Urke's definitions does not entirely align with the standards described by Wohlgennant et.al., but the same core concepts are applied here with presence, interactivity, and immersion.

2.1.2 Augmented reality and the reality-virtuality continuum

Augmented reality, or "AR", is the other big field of crossed reality. Augmented reality can be explained as the act of placing virtual objects or artefacts, over a real-life scenario. Usually by means of camera lenses or specially made glasses. It does this by recognizing patterns or locations and placing an object according to these coordinates. Urke mentions the social media platform snapchat as a common example of AR, as snapchat allows for face recognition filters that can alter or apply layers to the user's face (Urke, 2018, p.21).

Mixed Reality (MR) can be considered an evolution of AR, as it entails overlaying virtual objects that interact or react with the real environment. The line between AR and MR is quite blurry, and there is according to Urke a lot of disagreement of where the line is, or whether it even exists (Urke, 2018, p.20-21) (Milgram et.al., 1994, p. 291). These findings are corroborated by Wohlgennant et.al, but they also discuss the field of extended reality (XR), which encompasses all "real-and-virtual human-machine interactions, generated by computer technology and wearables. In other words, XR refers to everything on the Reality-virtuality continuum, except for reality itself. Including VR (Wohlgennant, Simons and Stieglitz, 2020, pp. 456-457).

Paul Milgram (et.al) established the concept of the reality-virtuality continuum, a linear scale which ranks the immersivity of crossed reality technologies. The scale goes from full real environment, to a fully virtual one (Milgram et.al., 1994, p 283). With AR and MR being placed somewhere in the middle of the scale, whereas VR would be considered fully virtual. For the purposes of this paper, the focus will be on VR, specifically VR in head mounted displays (HMD), as these offer the most immersive and sophisticated virtual experience on the available consumer market. Relevant studies regarding other forms of crossed reality will also be appropriated for use in VR if applicable.

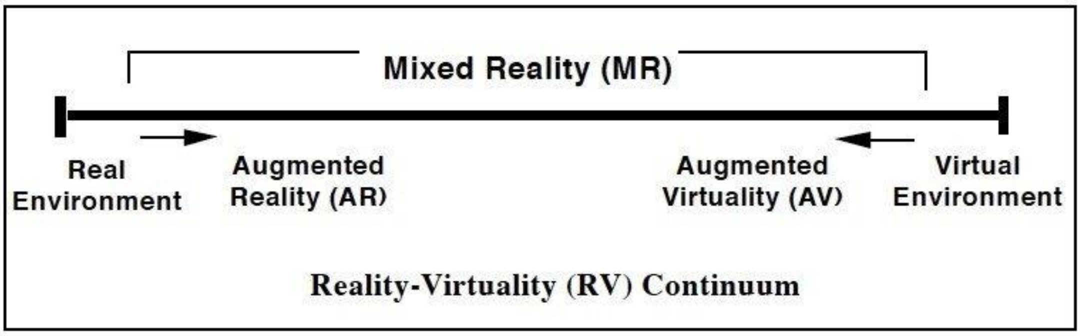


Figure 1: Illustration of Reality-virtuality continuum, illustrating how MR fits between the real and the fully virtual environments.

2.1.3 Head mounted displays and Reality systems.

In order to experience VR, one needs access to what is called a reality system. Reality systems are combined hardware and software that communicates sensory input and output between the system and user.

The most common way to experience VR today is through what is called a head mounted display, or an HMD for short. HMDs are attached to a user's head and uses screens placed over the user's eye to convey visual output. These screens also help block out the outside world, further increasing the sense of immersion. These headpieces also often feature earphones for audio output, and microphones for audible input. The HMD is also tracked in the virtual space, further helping the factor of presence, movements forced upon the headset will be displayed in the output, ensuring synchronized movement in the real and virtual world. Jason Jerald also describes the terms "non-see-through HMDs" and "video-see-through HMDs", meaning HMDs that allows the user to see the real and virtual worlds simultaneously, either through physical transparency or live video feeds. Though, this could be argued to be augmented reality, and not virtual reality, as it does not attain the total feeling of presence, something the author acknowledges (Jerald, 2016, pp. 32-33).

To attain the interactive quality, HMDs often come with hand controllers that allow the user to put their hands in the virtual space as well. These controllers use buttons and pressure- and/or touch-sensors to allow the user to interact with the environment. Using a trigger located near an index finger, or simply gripping the controller tighter can be potential ways to grip an item in VR. Other input tools in VR can be things like motion platforms and treadmills, that allows the user to enable more parts of their body. Although these items are less common, they can help negate factors such as motion sickness (Jerald, 2016, pp. 39-43).

In the HMD market for PC, there are two leading brands, HTC's Vive, and Facebook's Oculus. Both are similar in the sense that they use paired wireless controllers as input. But there are some differences with regards to both software and hardware. The only truly relevant difference that had a say in the choice of HMD was sensors. Both HMD use USB-linked sensors attached via the computer, to measure the HMD and controllers' location. The Vive's sensors come with roof/wall mountings and need to be attached higher up- While the Rift comes with very basic standing sensors, that can be put down on desk-height surfaces. The tradeoff being that with the Vive, set-up is more difficult and dependent on the area, but allows for top-down scans. Helping negate signal-blocks if the subject/user were to turn around. The sensors are also less vulnerable to accidental movement. The Rift does not have this bonus, being more prone to signal-block, but its ease of set up and transportation was decided to be more useful for this study. Therefore, a Rift was chosen for this study.

All though HMDs might be the most common for their ease of use, cost, and efficiency, they are not the only options. There are also world-fixed displays and hand-held displays (Jerald, 2016, pp. 33-34). World fixed displays aim to immerse the user by displaying graphics on surfaces within a physical space, either using screens or light projection. To ensure high quality immersion through this method requires specialized rooms and can therefore be quite costly.

Hand-held displays are output devices small enough to be held in one's hands, as the name implies. These can often be quite cheap, and a common example is cardboard goggles with plastic lenses that can serve a smartphone.

Both methods can be argued to not be true VR. The handheld displays often do not allow for enough interactivity or presence, as there is often a lack of input and motion tracking. The author also argues that world-fixed displays are greyer on the Reality-virtuality continuum, as they can be perceived as augmented reality instead (Jerald, 2016, p. 34).

2.1.4 VR-factors and training uses.

As previously mentioned, VR is usually defined by three factors, these being presence, interactivity and immersion, these factors are generally agreed upon among researchers, bar Immersion, which is still a subject to discussion (Wohlgennant, Simons and Stieglitz, 2020, p. 457). Presence refers to the users subjective feeling of being somewhere other than their physical space.

While the factor of interactivity dictates that the user has some sort of real-time control or influence over their surroundings, which is the focus of this study. Examples of interactivity in VR could be moving oneself by using a locomotive method. Or picking up objects or interacting with a menu.

Immersion is as previously mentioned a more divisive concept, whereas some treat it as a more objective or quantifiable concept, others treat it like a subjective or emotional factor. Immersion is defined as an absorbing involvement or learning from extensive exposure to surroundings and native conditions (Immersion, 2021). This term is commonly used term in video games, as games viewed to have settings and rules true to the game setting are deemed immersive.

As stated in the introduction, VR is being used for many different uses (Wohlgennant, Simons and Stieglitz, 2020, pp. 457-458), the background for this study is rooted in training. Joseph Psotka from the U.S. Army Research institute published an article in 1995 about the possibilities with VR for training purposes (Psotka, 1995), though somewhat dated, the paper describes the factor of immersion and its effects on training/education in VR, beyond that of novelty.

Psotka seems to use the terms Immersion and Presence interchangeably, using both to refer to the users feeling of being there. He goes on to argue that VR provides more intuitive and human ways for human-computer interactions, as using the senses is the way evolution prepared humans (Psotka, 1995, p. 410). VR essentially simulates learning from experience, using human senses like vision and hearing to gain insight into the virtual environment, like the real world.

2.2 Affordances and signifiers

2.2.1 Affordances

The term "affordance" was coined by American psychologist James Gibson, expert in visual perception. He introduced the term in his book, "*The Ecological Approach to Visual Perception*". In the book Gibson describes vision and perception through the lens of ecology. He describes an affordance as an attribute an object or environment offers an individual, unrelated to the individual's perception (Gibson, 1986, pp. 126-129). Using the ground as an example, the ground offers support to animals and humans alike and therefore has the affordance of support. However, as he explains, an affordance relies on the relationship between a subject and the object, and not its perception or perceptive qualities. For example, a surface may be strong enough to hold a human or small animal, but not a larger animal, even if it may look like it does. In that case, the affordance of support would only apply to the humans and smaller animals. As the affordance of support does not exist for the larger ones.

Although Gibson coined the term, there has been much discussion in the field of design and psychology surrounding the term affordances. (Torenvliet, 2003, pp. 13-15). Don Norman appropriated the term for his book "*The psychology of everyday thin*gs", (later "*The design of everyday things*") released in 1988. Norman originally described affordances as an object's perceptible qualities, informing a user of its potential uses. As explained by Torenvliet, despite Norman introducing the public to affordances, affordances were described and widely misunderstood as an objects visual attributes that enable interaction. Not the relationship between actor and object itself (Torenvliet, 2003, pp. 15-16). This study will use the definition as set by Gibson, with affordances being the relative relationship between an actor and an object.

Norman seems to have rectified his definition some in the revised edition of "*The design of everyday things"*. Here he defines an affordance as a relationship between an object and an interacting agent. He also explains how affordances differ for users of different abilities, using the example of a heavy object's "liftability" only being an affordance for agents strong enough to lift it (Norman, 2013, pp. 10-11). Norman did not however abandon his previous stance, but he appropriated them into the concept of "signifiers".

2.2.2 Signifiers

In the realm of design, the concept of signifiers arose as designers would misuse the term affordances to indicate a possibility of action, Norman defined the term signifiers to signify a point to the user can interact with the affordance. Signifiers are not to be confused with perceived affordances, as they only convey what an agent may believe the affordance may be. Signifiers are the visual, audible, or contextual (or other forms of sensory input) cues that imply interactivity, reusing the same example of a heavy object, handles or bars on this object would signify lifting to a person (Norman, 2013, pp. 13-19). These definitions of affordances and signifiers are also retold and reiterated by Jerald in his book (Jerald, 2016, pp. 278-280)

As signifiers represent affordances, signifiers can occur gradually, or accidentally in the world. Norman uses a pathway as an example, as the path works as a guide for walkers that others have used this shortcut earlier (Norman, 2014, p. 14) (Jerald, 2016, p. 279-280). Other such elements can be general wear-and-tear on buttons or other everyday objects, revealing that they have a function. In general, through consistent human use, natural signifiers will arise in the environment.

For this study, the focus will be on signifiers as defined by Norman, and the role they play in signifying affordances in VR experiences.

2.2.3 Affordances and signifiers in VR and VR environments

Applying these concepts to VR can be done similarly to how they are incorporated in conventional 2D interfaces. In traditional interaction design, signifiers are there to help the users reveal affordances. Highlighted buttons, underlined hyperlinks, recognizable elements like hamburger- and drop down-menus. But VR can also offer the same bonuses through immersive VR spaces that real world environments can offer as well. Furthermore, VR offers the ability to signify objects in a way that is not possible in "the real world". Like highlighting objects in the environment itself, rather than in conventional menus. Some applications will often highlight selected and/or selectable objects to signify interaction.

This study aims to investigate signifiers in this way, using three kinds of visual signifiers in the form of highlights. The three different ways are proximity-based positive signification, highlighting interactive object when the users hand gets near. Proximity-based positive and negative signification, providing colored highlights on all objects, with the color indicating the interactivity of the object. And finally constant positive signification, which will always signify the interactive object, independent of the user's proximity. These three ways were chosen as they aim to solve the problem presented by the MW employees, helping inexperienced users in VR-training scenarios distinguish important elements in the environment. Especially in Virtual Environments that may include many static and/or interactive objects. Haptic instances in the form of vibrations in the controllers was added as another proximity-based signifier as well.

Diegesis is another term that could be deemed appropriate for VR. Diegesis is defined by Merriam-Webster as *"the relaying of information in a fictional work (such as film or novel) through a narrative"* (Diegesis, 2021). A common example for film is diegetic music, where the music is relayed to the viewer as narrative piece, while also being part of the films setting. Say for example if the setting is a concert, both the films viewers and characters in the audience experience the music.

Salomoni et.al. (2017) extends that definition within media and explains it within the context of VR interfaces. In games and VR-applications, diegesis is often used in interfaces. Diegetic interfaces are more common in VR as overlaid menus can be considered disruptive to the experience, diminishing the feeling of presence and immersion. Examples of diegetic interfaces can include the user physically holding a map to ascertain their position or having an inventory/small menu attached to a hand/controller. In their study, Salomoni and accomplices find that users appreciate diegetic interfaces in VR, as they scored higher than non-diegetic ones, as it enhanced their feelings of presence and immersion (Salomoni et.al., 2017, pp. 180-183).

2.2.4 Signifiers and Feedforward

Some would argue that/mislabel these signifiers as feedback, it is important to distinguish. The same way that feedback is information given to the user to communicate action. Giving negative or positive feedback helps a user perceive whether their task was successful or not. Feedforward describes the information that helps guide the user to do that task (Norman, 2014, pp 71-73). Norman explains that feedforward is accomplished by a series of factors, including affordances and signifiers.

One could argue that the signifiers to be tested in this study are more akin to feedforward, however, feedforward relies on several factors, including signifiers. As the signifiers themselves are the main variable, the only real changing part from task to task within the test, the signifiers remain the focus of the study.

2.3 Workload

Sandra Hart and Lowell Staveland worked on NASA's Task Load Index during the 1980's, a tool for measuring mental workload. In a published paper, outlining the findings of their research, they give a definition for mental workload. For their study, they defined workload as a human-centered construct, representing cognitive cost in an agent for achieving a set task (Hart and Staveland, 1988, p. 140). This cost is derived from a series of factors in both the individual, environment, and task. Thus, they describe workload as subjective, as different agents with different qualities will perceive tasks and situations differently. Imposed workload for example, an adult performing a puzzle will probably experience lower workload than a child carrying out the same task.

Despite workload having several varied and differing definitions through the times, as retold by the user manual, and accompanying paper themselves. The given definition in these papers will be used for the purposes of this study, as the study relies on NASA's index as a central tool (Hart and Staveland, 1988, p. 140) (Human performance research group, n.d., p. 2).

3 Methods

3.1 Test description

The objective of this study is to find to what degree visual assistance in diegetic VR interfaces helps reduce mental workload, in the form of signifiers. To examine this, a test was designed. The test was designed to take place in a virtual space, which the user interacted with through a head mounted display and controllers. The model is an Oculus Rift, borrowed from the VR-Lab at NTNU Gjøvik, consisting of an HMD as well as controllers for each hand.

The test takes place in a simple VR experience, based on earlier 3D assets. The test was to consist of the user being asked to perform a series of tasks, such as finding and picking up a certain object (called interactive object), from a series of static ones. This test would be repeated, with varying degrees of signifiers for each repetition. The interactive object should also be changed from test to test, so that the test subject is unable to memorize the objects. This was later changed, as simply interacting with an object was deemed too simple to offer any real challenge. Instead, a puzzle consisting of a user entering a passcode into a diegetic keypad interface was selected. This would make the test harder to guess, leaving the subjects to rely on the given signifiers in the keypads interface.

Following each task, the subject will be asked to answer a RTLX form as the facilitator notes answers. This is done so that the users can remain immersed in the VR space, as removing the HMD several times during the test would be disruptive.

3.1.1 Pre-task

Before each test takes place, each participant will be informed through the consent form (see appendix A), subjects will be asked to consent with a signature before the test starts. Each subject will then be assigned a number for the sake of data collecting, starting at #1 and going up.

The participant will be guided through a short pre task, being asked to perform basic movements with the controllers, and having a run through of the controller scheme if the subjects are not familiar. Afterwards the subject is shown the RTLX sheet and explained the six subdivisions explained to them, and how the RTLX will be used in the test. This acts as a way of making sure the subject has some familiarity with both VR and the RTLX form before the test starts.

3.1.2 Main tasks

The prototype is to consist of a flat wall with four keypads spaced out evenly and numbered in order. Each keypad looks identical, consisting of a screen with room for three digits and 12 buttons, one for each number, one to zero, as well as a backspace and an enter. The backspace key will remove the last entered digit, and the enter button will submit the current code for validation. Each keypad has a unique three-digit code, selected from a random number generator to avoid patterns. Subjects are tasked with guessing and entering the correct code for each numpad, with help from the signifiers given in the keypads interface.

The type of signifier changes for each keypad. The subjects will start in front of keypad number one and will be teleported to the next when the right code has been entered and submitted, continuing until all keypads are completed. Keypad 1 is the only keypad without a signifier, to serve as a control.

The remaining pads are displayed in figure 2, where the yellow points indicate the position of the subject's index finger. Keypad 2 uses a highlight effect when the subjects hand gets close to the correct button, allowing for a more immersive and visually subtle type of signification. This is displayed in the illustration, as the correct number is 2. Keypad 3 uses both positive and negative highlights when the hand is in proximity to the buttons. The right button will be highlighted green, while the others red. In the illustration, the correct button is shown to be number 5, but the buttons surrounding the subject's finger also provide negative visual stimuli. Keypad number 4 goes back to simple highlights, however now the highlight is constant, foregoing the need for proximity. This is also shown in the illustration, as the correct button is on the opposite side of where the subject is pointing. This will allow the user to see the correct option immediately, although at the cost of immersion.

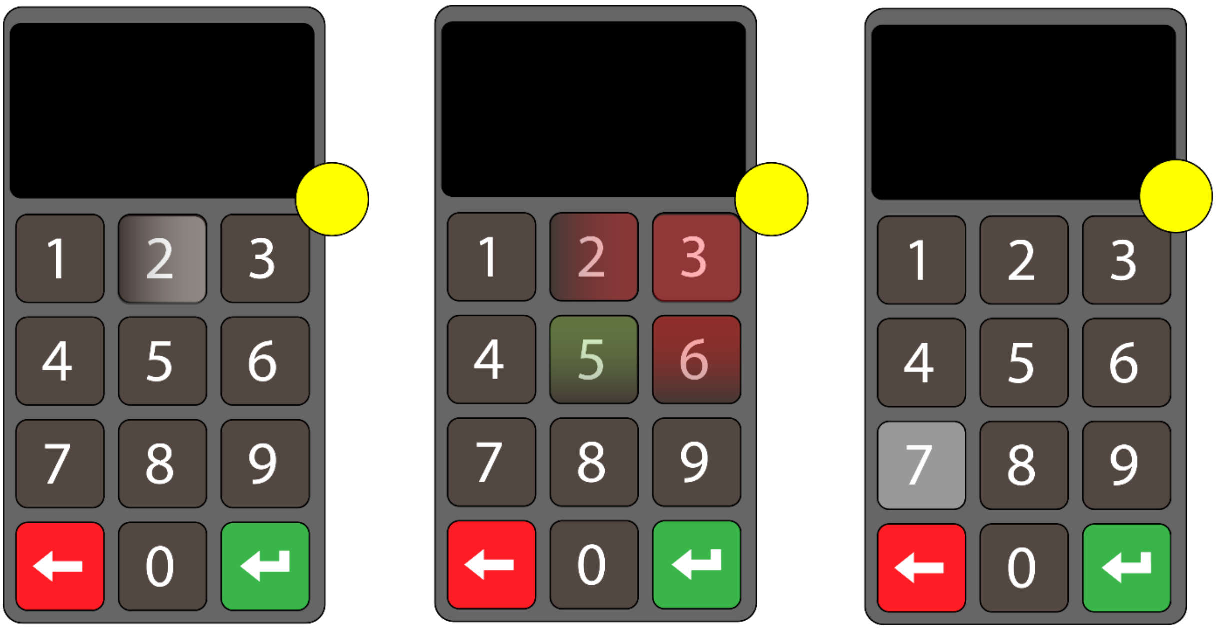


Figure 2: Illustration showing how the buttons react to the finger's position (yellow dot). The correct highlighted buttons are 2, 5 and 7 respectively.

In addition to the signifiers, feedback was also integrated into the buttons, giving a soft haptic vibration on successful presses, and negative audible feedback on presses of a wrong button. If one were to notice this, one could correctly guess the combinations by trial and error, listening for these.

3.1.3 3.2.3 Wrap-up

After the test has finished, the subject will be asked some questions. Whether they have experienced VR before, or if they are used to navigating 3D spaces, such as 3D modelling software or video games. Every subject will be given a score from zero to four, depending on their experience. A score of zero indicates no or very little 3D-spacec experience. A score of one to two indicates that the user has some familiarity, three for a lot of experience on 2D screens, such as video games. A score of four points would indicate that the user is experienced in VR environments.

These scores will be used to compare their workloads for each task and investigate whether users of different expertise levels have different opinions on Signifier use.

3.2 Heuristic evaluation

Before conducting the test, a heuristic evaluation was carried out on the prototype to gain insights on the validity of the prototype itself. A heuristic evaluation is a usability engineering method, used to locate problematic areas in interfaces. It is performed by inspecting a user interface from a series of heuristics, as explained by Jakob Nielsen and Rolf Molich (1990). By having several people look at an interface, different problematic areas and systems can be identified. This method is to be used to inspect the VR prototype program, and check for variables and factors that may affect the test. As the keypad the user interacts with can be considered their interface.

This could not be called a true heuristic evaluation however, as Nielsen and Molich themselves explain that such a task would be difficult for a single person to do, as one could never find every problem by themselves, and that different individuals excel at locating different issues. Therefore, an optimal heuristic evaluation should include three to five evaluators (Nielsen, 2007) (Nielsen and Molich, 1990, p. 255). Also, the fact that this is being used to test a VR experience developed by the facilitator themselves could bring some biases and thus discredit the validity of its findings.

As previously mentioned, the evaluation will be derived from a series of heuristics. In their conference paper, Nielsen and Molich (1990, p. 249) use earlier version of what would later be called Nielsen Heuristics. For this evaluation, an updated list derived from the more widely established 1994 version of Nielsen's heuristics will be used (Nielsen, 2020). These are in short:

Visibility of system status: keeping the user in the know of the systems status. Providing information and feedback. **Match between system and world:** Using real world conventions and language towards the user. **User control and freedom:** giving the users the ability to undo and redo, exit unwanted situations.

Consistency and standards: having consistent language and labelling. **Error prevention:** designing interfaces to prevent misuse. **Recognition, not recall:** keeping actions and options visible to mitigate memory load/workload. Instructions should be easily available.

Flexibility and ease of use: system should cater to both novice and experienced users. Allow user to customize shortcuts if applicable. **Aesthetics and minimalist designs:** No unnecessary text or information that can distract from the core information and/or content. **Help user recognize and recover from error:** Errors should be clearly communicated and communicate the issue clearly. Suggestions for solutions are also recommended. And finally **Help and documentation:** Documentation can be provided if necessary, to help user figure out tasks. Should be readable, precise, and available. (Nielsen, 2020)

3.3 Nasa Task Load Index

To test how the lack of clear signifiers affect a subject's workload, an optimal tool would be the NASA Task load Index, or NASA TLX for short. The NASA TLX is described as a multi-dimensional rating procedure, serving as a survey to rank the importance of six factors, or subscales. These subscales are mental demand, physical demand, temporal demand, performance, effort and lastly, frustration. The survey is handed to the test subject(s) after finishing a set task.

These subscales are measured in two parts, weighing and rating. After the task has been performed, the subject is asked to weigh the six subscales against each other, with each subscale being paired against each other, for a total of 15 pairings. The subscales are then tallied for the number of times they took precedence, which represents the subscale's "weight" (Human Performance Research Group, n.d., pp. 2-5).

Afterwards the subject fills out the rating sheet, where they rate the effect of each subscale on a scale from 0 to 100, with increments of 5. This is called the "raw rating". The "adjusted rating" for each subscale is calculated by multiplying that subscales weight and raw rating. The adjusted rating for every subscale is summed, and then divided by 15 to get the final weighted rating.

The NASA TLX is a proven and reliable tool for measuring a subject's workload, but some have sought to improve it.

3.3.1 RAW Task Load Index

An alternative to the NASA TLX is the RAW Task Load Index, called RTLX for short. The RTLX foregoes the comparing in the weighing, and instead focuses solely on the ratings. The final rating is achieved by simply calculating the mean ratings of the subscales.

A study published in 1989, comparing the RTLX to the standard NASA TLX, was carried out. They concluded that the RTLX is a good alternative to the TLX, as their studies found that it had a lower std. deviation by omitting the weighing phase (Byers, Bittner, and Hill, 1989, p. 484). This, in combination with its simpler design and process makes it an optimal tool for testing on larger amounts of users, or for faster testing overall.

The RTLX was chosen for this study, as it made for faster evaluation in the subjects, as they would have to do a TLX several times during the study. The RTLX was chosen as the quicker alternative as it foregoes the weighing, which is the most time-consuming part of the survey. It was feared that conducting the whole survey would take too long and break the subject's immersion. By simply asking the subject's how they felt about the six factors on a scale of 0-100 would be faster. This way the subjects can stay within the virtual space and simply answer verbally. This would be more difficult to do with a standard TLX.

3.4 Test development

To develop the test, a series of prototyping and animation tools were used. Some efforts were made to develop the software, as it would be preferable to not have to rely on outside factors. The two most notable applications tested for these purposes were the VR-sketching applications Tvori and Sketchbox. These however turned out to be insufficient for the test goals, and outside help was hired to help finish the test software, hereby referred to as "the prototype".

3.4.1 Figma

Figma is an online prototyping tool, allowing users to create and script prototypes in real time. Figma is well suited for screen based 2D applications, but there have been efforts to translate Figma to Virtual reality and 3D spaces. Using a 360° as a background, a simulated VR experience on a 2D screen can be created in the form of a panorama. This can be done by placing 2D overlays over the panorama, emulating a HUD.

Figma could serve as a plan B if testing in Virtual reality would not be an option, and plans were made accordingly. However, simply testing the signifiers in a 2D-based Figma prototype would be unsuitable, as there would be a significant difference from a 3D space. If was to be made a backup-platform, a panorama-view application would be the best bet, as it would arguably be more representative of a VR experience than a flat screen.

This never materialized, as an HMD was acquired, and a fully interactive VR-based prototype was developed.

3.4.2 Tvori and Sketchbox

Tvori is an animation software, it allows the user to pose models in environments, and animate them with effects along a timeline. This allows users to create and share animations where the users are the camera. Some experimentation was done with Tvori, as it was thought that the animation could be used to animate the different kinds of signifiers on models.

This was possible within the software but getting the animation to correspond with the users' motions was found to be difficult. Although there are many opportunities in Tvori, making it immersive was a challenge. In the sense that the animation itself would play out in a predetermined manner, and not according to the user's input. For this reason, it was decided to try out other solutions as well and come back to Tvori if no alternatives were found.

Another free software, Sketchbox was tried for development. Sketchbox describes themselves as the number one Design and Collaboration tool for AR and VR. Sketchbox allows users to collaborate on shaping and designing environments and scenarios, much like a typical wireframing toll such as Figma, just in VR instead of a flat screen. Shaping environments in Sketchbox was easier than in Tvori, as the software allows one to choose from several shapes and online-models. However, Sketchbox was more limited in terms of animation and effects, and thus could not be a reasonable alternative for a high-fidelity prototype. Even more so, as there was built in animations on objects that had been selected by the user, making it even more difficult for testing animated signifiers.

Although both Tvori and Sketchbox were designed around VR-prototyping and allowed for quick starts and ease of use, they were deemed Unsuitable. Their animation capabilities were either too simple or too specific for the needs of the prototype, and considerable time and resources would be used on making something adequate. Sketchbox would see use the development of the prototype, in its intended area, as it was used to map out the test area and visualize the types of signifiers. This would turn out to be helpful for the final step of development. Screenshots from Sketchbox can be seen in figure 3 below.

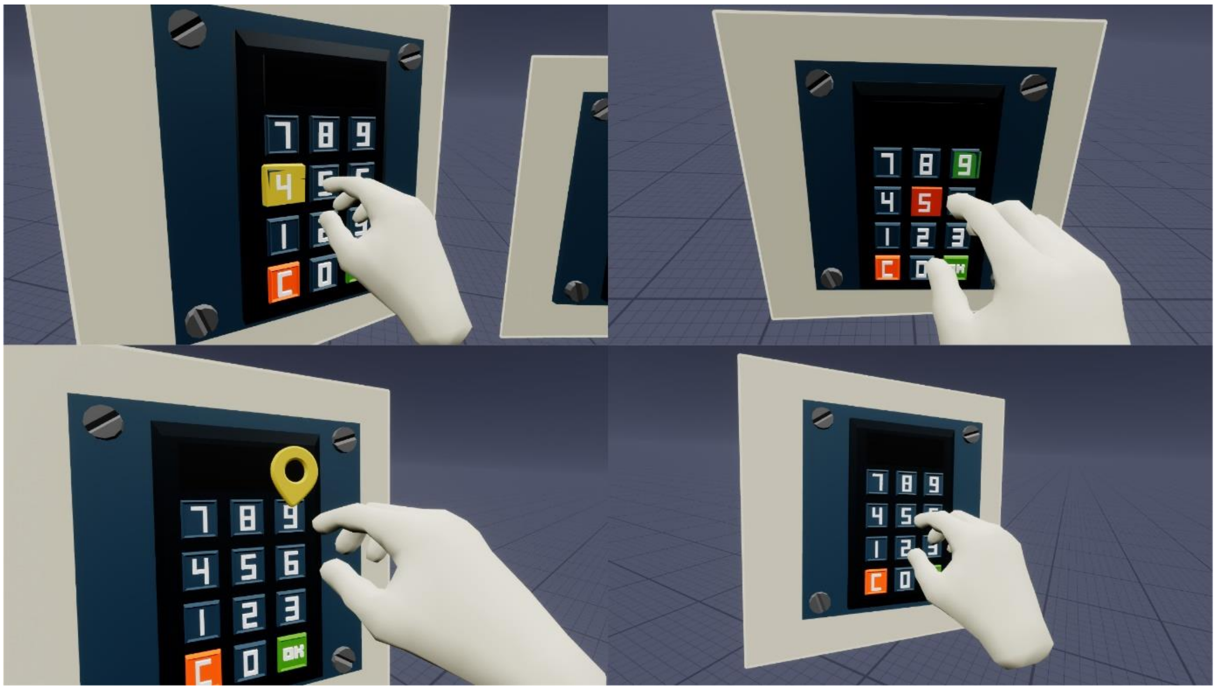


Figure 3: The signifiers sketched in Sketchbox. These images were used as a template for the final prototype. Starting from top left: Keypad 2, 3, 4 and 1/5.

3.4.3 Unity

Unity was recommended by one of the VR experts consulted earlier, unity is described as a free game engine suited for first time developers (Dealessandri, 2020). The contact suggested using Unity's sample case for VR, one of their many online tutorial cases designed to teach the basics of Unity. Using an existing case, one could alter details and assets to make something not completely from scratch. After further talks, it became apparent that the level of transforming needed was within Unity's capabilities, but not within the existing case file. The animation and physically interactive aspects were too complicated and would require additional programming. A prototype would have to be created from the start.

Unity was selected as a suitable solution for the prototype, as only Unity or a similar programming/game engine software would allow for this level of customizability and relative ease of use. But due to a lack of knowledge and experience in unity, outside help was hired, namely the contact who recommended Unity. Talks were had about the nature and goal of the study, and what was needed in terms of programming and assets. A sketch was made in Sketchbox, and was used to visualize the signifiers, and the layout of the room. Even though the facilitator was not involved in the programming directly, there was continuous communication and testing, until both parties were pleased with the result. A simple design document used by the facilitator to communicate the design and layout is offered in the appendices (Appendix B). The prototype was finished after two days, and no serious bugs or problems were found after a couple of test runs. Images from the finished prototype can be seen below in figure 4.

Figure 4: Keypads 1, 3, 2 and 4 as they appear in the final prototype. Notice the highlights based on the hands position, in image 2 and 3.

A fifth keypad was added and differs from the rest as it offers a non-visual form of signifier. It was decided to add another keypad as it was clear that unity would allow for control over the haptic sensors in the Rifts controllers. This last keypad would test out haptic vibrations as a signifier where the vibrating sensors in the controller will be activated upon the hand's proximity to the correct button. This would be done gradually as well, with a stronger and more frequent vibration the closer the subject is to the button.

3.5 Pilot test

Before the study was executed, a pilot test was carried out to check the feasibility and practicality of the test. The HMD and computer were set up in a private area, and five sheets of the RTLX survey were printed, one for each keypad/task. The subject was asked to stand, face towards the screen. This was done to ensure good readings for the sensors. The test went well, as the subject managed to solve all keypads within a reasonable timeframe except for the last one. The subject was convinced that the last pad was defective, as no visual form of help was available. It was decided that going forward with the real study, the subject should be informed that the last keypad involves a non-visual signifier. The subject rated themselves as group 2, having some VR experience prior.

Having the subject face towards the screen did make for good sensor readings but made it more difficult for the facilitator to follow what the subject was doing, as their first-person view could be seen from the screen. There were also instances where the subject's hands got close to the screen, meaning someone could punch the screen if they were eager enough. For these reasons, changes were made to the setup, having the subject face towards the left, with the screen on their right side. The facilitator would then sit to the subjects left, making the subject and facilitator face each other during the test. The screen was simply turned on its side, allowing the facilitator to see what was going on. This can be shown from a bird's eye view in figure 5. No real synchronizing issues were found by having the subject oriented this way. After conducting another pilot test with the same subject, this time focusing on the setup itself. It was found that this orientation worked well and gave both parties more breathing room. The chances of a subject hitting and breaking the screen was also greatly reduced.

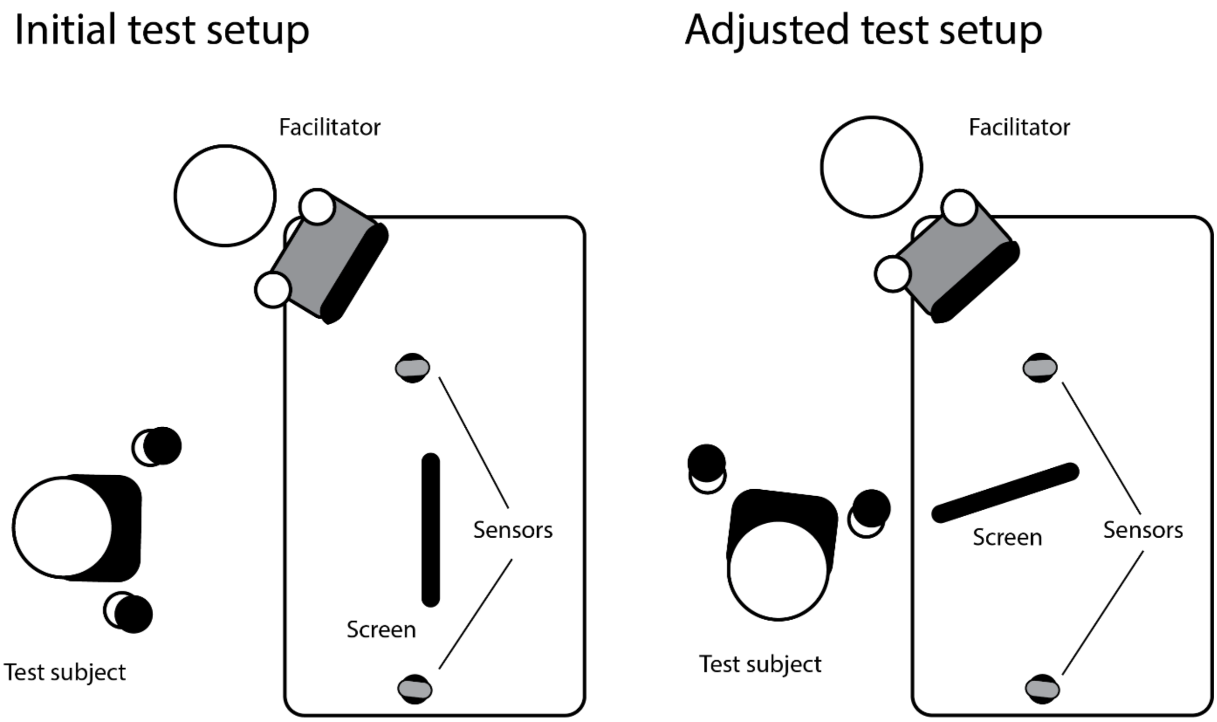


Figure 5: Top-down view illustrating the test setup. Left side shows setup during pilot test, and adjusted setup on the right.

The subject was also a bit confused about Oculus' built-in barrier, believing it to be part of the keypads interface. The barrier is a red mesh that signifies to the user that part of their body is near the edge of the playable area. It was decided that a bigger area was needed for the test, as this private dwelling was too small.

The last change made was the omission of the RTLX survey sheets, having five individual pieces of paper for one test was found to be excessive, an alternative was marking all tests on one sheet, but this made the survey paper difficult to read. It was decided to forego the sheets in favour of a digital table, saving paper as well as time.

3.6 Test setup

To ensure a larger test area and more readily available subjects, the tests were carried out in an open working space for students at NTNU Gjøvik's Department of Design. This was done as the study was executed ~~during~~ from April 12. until April 14., during the third wave of the COVID-19 outbreak in Norway. The location already had existing infrastructure for tracking infection, using check-ins. The location also allowed for readily available subjects, namely students and staff at the department. A vacant table was used, and the computer and HMD were brought in for three consecutive days. With the subjects and screen facing the facilitator. Further precautions were taken, providing subjects with masks, and using NTNU's supply of anti-bac. The HMD was also cleaned with Anti-bac between tests. Home baked cookies were provided as an incentive for participation to those who finished the test (See appendix B for recipe).

4 Practical and ethical considerations

4.1 Ethical issues

As the area of study revolves around usability and human interaction with 3D environments, which means human testing is a must. When dealing with humans in research, there are four categories of challenges according to Leedy and Ormrod (2015, pp. 120-126). These are: Right to privacy, protection from harm, voluntary and informed participation, in addition to general academic honesty.

4.1.1 Right to privacy

For the purposes of this study, the only relevant characteristic will be the user familiarity with XR and virtual environments. No personal information such as names, gender, nationality etc. will be considered important for this study, only confirmation that all test subjects are above the age of 18. This is done to protect all subjects right to privacy. A form will be sent to NSD (Norwegian Centre for Scientific data). Detailing what information will be gathered, and whether it is sufficient in relation to privacy rules. This form will be sent before the start of March, as to be ready for the start of the study planning.

4.1.2 Protection from harm

No user groups or users that may be considered vulnerable, such as disabled or underage people will be included. Reasoning being that these demographics could be more exposed to unforeseen ethical consequences (Leedy and Ormrod, 2015, pp. 120-121). But the nature of the study also requires motor skills, that a child or a person with disabilities may not possess.

4.1.3 Voluntary and informed participation

All test subjects will be given consent forms and informed of the studies' explicit objectives in compliance with guidelines concerning informed participation (Leedy and Ormrod, 2015, pp. 121-123). They will also be given contact information and may withdraw their data at any time during the study. All raw individual data will be deleted after the completion of the study in June 2021.

4.2 Practical challenges

4.2.1 Partnership

As it stands per December, the partnership with MW is very informal and personal. It was considered that a written formal contract/agreement could be written for the sake of both parts. This would have been done to ensures a more stable and reliable working agreement and may allow for involvement in current projects for testing purposes. This

could also supply the study with relevant testers in the target audience. Alternatively, a demo could be made somewhat from scratch, but this would take more time out of the study. But is a possible "plan B". However, this would also mean that the study would revolve around both parties' schedules, and MW has stated that they may not be applicable for a partnership before march.
For these reasons, it was decided to keep the informal soft partnership with one of the employees, this employee would help develop the prototype for monetary reimbursement.
Another option considered was the "VR LAB" at NTNU, who could also provide equipment and locale for the study. These would be contacted, not for a full partnership. But for the lease of equipment.

4.2.2 Borrowing equipment

For the study to take place a HMD and computer is needed, along with a testing place. This can be arranged privately, but it was decided to go through with the VR lab at NTNU Gjøvik, as they had helped previous students with similar tests. The VR lab gave two options, an HTC Vive or and Oculus Rift. The Rift was chosen for the reasons stated in the background chapter, ease of set-up and transport. The computer used to run the VR software was the facilitators personal desktop computer, as the graphics processing unit there was powerful enough to run the software. Thus, borrowing a computer was deemed unnecessary.

The equipment was borrowed from the 25th of March to the 16th of April, allowing for enough time to finish a prototype and do the testing, with a goal of 10-15 participants. This goal was met with time to spare, and the lab offered to prolong the "rental". However, it was decided as the goal was met, this was deemed unnecessary.

4.3 Risk assessment

4.3.1 Coronavirus

As of December 2020, the COVID-19 pandemic is still ongoing, although vaccinations are about to start, it is important to assume that the circumstances will remain the same out the semester. For that reason, it is important to take certain precautions when it comes to testing. Only one test subject will be tested at the time, with both subject and experimenter using masks. All equipment, such as HMD and controllers will be cleaned after use. If the tests take place in a more public setting, like the NTNU VR lab, or MW's offices, existing infection tracing infrastructure will be used. In the case that the testing takes place, contact information from participants must be collected, but not for use in data.

4.3.2 "Plan B's"

In the case that something was to happen to the equipment, borrowing new could be an option if partnered with NTNU or MW. In the case that testing is done privately, borrowing from an acquaintance could be a possible, but unreliable alternative. Data and reports will also be stored on a backup drive, in case of data corruption. Online cloud storage services, like google drive are also options.

In the case that government regulations prohibit testing on large masses, smaller more local samples can be used. Such as fellow students. These may not accurately reflect the real population, and result in a smaller sample size, but may offer some consolidation in the event that the "true population" cannot be sampled.

4.3.3 VR and health effects

Although VR poses no immediate health threat to able-bodied people. There are certain side effects that can happen during prolonged use, or for certain sensitive users. Eye strain, queasiness and motion sickness are the most common. These symptoms are collected under the term Cybersickness, or sometimes VR-sickness. Cybersickness is very similar to motion sickness, caused by a disconnect between the vestibular and visual systems. But Cybersickness is different from typical motion sickness in that only visual input is required for cyber sickness, not vestibular (LaViola, 2000, pp. 47-48). Several different theories surrounding the cause of Cybersickness has been brought, theorizing the human need for up-right orientation, and evolutionary reactions to poisoning, as possible explanations (LaViola, 2000, pp. 51-52). The sensory conflict theory is said to be the most accepted. Which is explained as the discrepancy between visual and vestibular systems, brought forth from movement in VR environments, that does not translate to the real-world senses (LaViola, 2000, pp. 50-51).

The prototype developed for this study was deemed to be stationary enough to not cause much disruption for vestibular systems, as no movement was required. And thus, no real preventive measures were taken to ensure the subjects of this, which can be seen as a breach of informed consent.
The reason for failing to inform the test subjects was simply a lack of forethought and would be ratified in potential future studies. Especially as one test subject became queasy during the test, however, they made a quick recovery and agreed to continue, and they were able to finish.

5 Results

5.1 Heuristic evaluation results

Before and after the pilot test, a quick and simplified heuristic evaluation of the prototype software was conducted after some testing. Only the facilitator committed to the evaluation in full as in comparing the prototype towards all ten heuristics. But the developer and pilot test subject were also willing to give feedback on their impression of the prototype, though they did not go the same depth on all heuristics. Thus, it is unclear whether one could argue that this was a true heuristic evaluation.

Visibility of system status:
All keypads are always visible within the user's panoramic field of view, and text displaying the keypad number are located above the pad itself. Keeping test subjects in the know of their progress in the test. The keypad also displays the number entered. Feedback is provided after a successful button press, visually, audibly and with haptic vibrations. Though the audible feedback is only

Match between system and world:
The concept of a keypad is very familiar to most people, especially as there are similar keypads in the building where the tests were performed. The buttons are pressed manually with a hand in a VR controller and should thus correlate to a real-world action.

User control and freedom:
There are no direct ways to exit, restart or go back within the prototype. The facilitator can restart or quit the application from the desktop, and the built-in options menu in Oculus allows the users to do so. Although users were asked to not press any of the top facing buttons on the controllers. Another option is for the subjects themselves to physically remove themselves from the virtual space by removing the HMD, although this could be considered a brute force solution. These factors could be argued to go against the heuristic of user control, but for the sake of the test at hand it should be fine.

Consistency and standards:
All labelling and design language are the same across all keypads, and there is no text on the buttons themselves (apart from numbers). Colors and symbols indicate the actions of the enter and backspace buttons and should be understandable for most Norwegian/western users, as these symbols appear on standard keyboards. The colors as well give some culturally accepted clues to their purpose, green being go, and red being stop/back.

The only change in labelling would be the signifiers themselves, but they act as the main variable in the test and should therefore be expected to do so.

Error prevention:
The prototype generally accomplishes this heuristic, as the usability of the numpads themselves are very limited. The only uses are typing in numbers and submitting or erasing. These mechanics must be used for the test to progress further.

A potential misuse could be the user's ability to freely move around the space. No locomotive method is included in the prototype as the user is teleported to an appropriate distance after each task. But there is no limit to manual bipedal locomotion. This however would not be practical, as the test area is too small for the subjects to lost. It was also important that no limit of movement was involved, as different users could need different to adjust their proximity to the pads to effectively interact with them.

As mentioned within the pilot test, Oculus's built-in barrier could be confused for part of the interface. Yet this is a setting in the VR driver software itself, so the only precaution that can be taken is ensuring a large enough test area.

Another potential user error in oculus is the controllers top facing buttons of the controllers. The downward buttons open the oculus menu and can be disorientating and easily pressed by mistake. Users were asked to not press the top buttons as mentioned earlier, as they would not be necessary for the test.

Recognition, not recall:
Keeping actions and options visible to mitigate memory load/workload. Instructions should be easily available.

As stated above, the number and complexity of actions within the application are very limited and can be reduced to button pressing. This should be somewhat instinctive, given the real-world correlations.

An obvious defect within the prototype is the lack of available instructions within the virtual space itself. Though the facilitator is available to help the user in case they forget their task or need reminders, reminders within the space could easily be integrated. Either as a static or reactive element, for example mapped to one of the buttons on the controller.

Flexibility and ease of use:
As the prototype was developed for the sake of an experiment, there is little flexibility in its use. And therefore, there are no customizable options for the testers, apart from arguably maybe which hand to dominantly use. The design of the numpad itself is deemed to be true enough to real life that both inexperienced and experienced VR users would recognize it as such, and the means of interaction diegetic enough as well.

Aesthetics and minimalist designs:
Little text or other contextual information other than the keypads themselves are provided. The pads are labelled according to number and could be argued to be unnecessary as the rest of the pads are in the user's field of view. Other than that, the keypads act well as a blanks slate for the signifiers, as the different kinds of visual markings translated well to the buttons in the pads interface.

Help user recognize and recover from error:
The potential errors noted above in *Error prevention*, such as excessive movement and getting lost within the oculus' menus. Although this could be prevented, there is little to be done within the confines of the prototype itself to help the user mitigate this. Therefore, subjects must rely on the facilitator to recover from these errors. A clear flaw in the process.

Help and documentation:
As the prototype is developed for testing purposes, there are no written available instructions within the software. Through the evaluation, it was deemed that the

prototype itself failed to deliver on this heuristic, though the facilitator is available to answer any questions the tester might have.

Following this examination, the prototype was deemed appropriate for testing purposes.

5.2 Test participants

A total of 14 subjects participated in the study, all of whom were students and staff at the institute for design as previously mentioned. The initial scoring system made to categorize the subject's familiarity in VR was deemed too vague, as the scale took regards to both VR- and "traditional" 3D-environment experience. This led to some subjects ranking themselves higher than they should, even though they had no VR experience, even ranking higher than some users with noticeable experience in VR. This led to a very unbalanced distribution, with no one being put in group 0, and eight subjects in group two. While group three had four, and group one and four had one subject each. This means that almost 60% of all subjects fell within one out of the five groups.

For this reason, a new scale was made, based solely on the users experience in VR, this scale ranged from one to three. Subjects who said they had zero experience in VR, was put in "group 1", users with some in "group 2", and lastly users with reasonable experience in "group 3". Group 3 consisted of users who had ranked themselves at experience level 3 and 4. By solely focusing on VR experience, a lot of the vagueness and grey zones could be ruled out. Meaning no longer could a user with no VR experience be ranked higher because of conventional 3D experience. These new groupings were achievable as the subjects had been quizzed about their experiences, and these experiences had been saved as qualitative data to help more clearly categorize the users.

With this new grouping, the 14 participants were spread more evenly across the groups, in a division of 4/5/5, as opposed to the original 0/1/8/4/1. This made for more consistent samples, and more reliable data as the few participants were spread over fewer groups.

5.2.1 Participant feedback

The participants were also encouraged to share feedback and thoughts around the study themselves. These could be about the administration of the test itself, the prototype software, or their own experience. When it came to the users' thoughts on their own performance, it was mainly based around how they themselves thought they affected their own workload score. Some users admitted to having had some issues with controlling their virtual hand, finding it difficult to make a pointing gesture. They said this may have inflated their workload in the beginning. Some other also admitted to having fun or a good time during the test. Something one subject admitted made them eager, which may have influenced the subscale temporal demand.

Surrounding the test design itself, one participant misjudged the haptics in test 5 to be an error message. While others had different preconceived notions of what the backspace button would do. Some correctly assumed that it would remove only one digit, while other thought it would reset the entire field. Some also had issues figuring out the proximity-based signifiers at first but had a relatively easy time after some closer inspection. While other users managed to learn the controls and systems, helping them predict next steps, such as the proximity-based signifiers and controls. The most apparent problem with the test was the implied time pressure. All participants were told that there was no hard time limit, but that they would be asked to stop after a minute if they were not on the "right track" to a solution. As noted by a participant, since the first test was the most difficult and time consuming, this would make "temporal demand" an overrepresented factor in the first task. As all the other tasks could easily be solved within that timeframe. These factors could all have affected their workload scores.

5.3 Test results

All results were transcribed into a excel spreadsheet, with each subject getting their own table displaying their ratings for each factor and workload for each keypad. The averages for the workload factor ratings and workload scores for each task was calculated within the spreadsheet across each group. An example of a subjects table can be seen below (further anonymized by removing subject number):

Table 1: Data table of an anonymized subject.

Subject X	XP 3	Group 3	20 hrs of VR		
	k 1	k 2	k 3	k 4	k 5
Ment. dem	20,00	50,00	50,00	60,00	60,00
Phys. dem	80,00	10,00	20,00	20,00	50,00
Temp. dem	80,00	10,00	10,00	10,00	10,00
performance	100,00	40,00	30,00	50,00	60,00
Effort	80,00	30,00	20,00	20,00	60,00
frustration	60,00	20,00	10,00	10,00	50,00
workload	70,00	26,67	23,33	28,33	48,33

In order to gain further statistical insights such as correlations and standard deviations, the data was transferred to IBM's SPSS software, and these tables will be used for the remainder of the paper. To see the average workload scores based on skill and in total, a report was calculated.

Table 2: Report from SPSS.

Report

XP		workload1	workload2	workload3	workload4	workload5
1	Mean	38,3350	10,0000	12,5000	3,5400	14,5825
	N	4	4	4	4	4
	Std. Deviation	3,53396	10,29608	18,53445	4,53326	9,34023
2	Mean	41,1660	18,8340	8,1680	3,3340	30,1640
	N	5	5	5	5	5
	Std. Deviation	15,69587	9,29138	9,22966	2,82573	9,36110
3	Mean	47,3340	14,1680	9,4980	15,8320	23,8300
	N	5	5	5	5	5
	Std. Deviation	20,16822	12,70843	8,54995	19,92183	20,57000
Total	Mean	42,5600	14,6436	9,8807	7,8564	23,4500
	N	14	14	14	14	14
	Std. Deviation	14,79348	10,68624	11,45752	12,93755	14,79444

A clear decrease in workload can be seen on average, dropping by over half from keypad number 1 to 2. It continues to decrease for 3 and 4, as the additional and more clear signifiers become available, as expected. The workload juts up again on number 5, as this keypad deals with haptic signifiers as opposed to visual ones, but the workload does not reach near the same level as keypad number 1.

To see how the user differentiated based on skill, their averages were calculated separately as well based on experience level. In this report (table 2) XP represents the variable of experience, with the three groups numbered as 1 2 and 3. The variables named such as workload1, workload 2 and so on, are the workload scores for the keypads with the corresponding number. The mean workloads per group can thus be read.

From a first glance one can tell that the inexperienced group 1 achieved lower workloads on keypads 1, 2 and 5, than the other groups. As their workloads were ranged as 38,33, 10, and 14,58 respectively. The workloads for keypads 4 and 5 were 12,5 and 3,54. With the score for keypad 3 surpassing that of the total mean, while number 4 scored low, but still higher than the mean of group 2.

Group 2 had scores of 41,16, 18,83, 8,16, 3,33 and 30,16 across all keypads. Scoring the lowest out of the three groups on keypads number 3 and 4, and scoring under the means on three, bar tasks number 2 and 5.

The most experienced group, group 3 scored the consistently highest across the board. Where only two out of the five tasks scored lower than the average mean, those being keypad number 2 and keypad number three, although only be decimals. The scores for group 3 were as follows: 47,33 for task 1, 14,16 and 9,49 for tasks 2 and 3, number 4 at 15,84, and finally 23,83 for number 5.

In table 2, one can also see the numbers of participants in each group (N), as well as the standard deviation. Again, these can be seen for each group and the total. In total (all groups) the dispersion is ranked from 10,6 to 14,8. A generally high spread, indicating that the different sets of data are not that consistent with each other. Groupwise the SD ranges from 3,53 to 18,53, 2,82 to 15,70 and 8,55 to 20,57 across groups 1, 2 and 3, respectively.

Table 3: Table of correlations from SPSS.

Correlations

		XP	workload1	workload2	workload3	workload4	workload5
XP	Pearson Correlation	1	,256	,141	-,100	,410	,226
	Sig. (2-tailed)		,377	,631	,733	,145	,436
	N	14	14	14	14	14	14
workload1	Pearson Correlation	,256	1	,279	,404	,121	,375
	Sig. (2-tailed)	,377		,335	,152	,681	,187
	N	14	14	14	14	14	14
workload2	Pearson Correlation	,141	,279	1	,513	,638*	,911**
	Sig. (2-tailed)	,631	,335		,061	,014	,000
	N	14	14	14	14	14	14
workload3	Pearson Correlation	-,100	,404	,513	1	,321	,379
	Sig. (2-tailed)	,733	,152	,061		,263	,181
	N	14	14	14	14	14	14
workload4	Pearson Correlation	,410	,121	,638*	,321	1	,661**
	Sig. (2-tailed)	,145	,681	,014	,263		,010
	N	14	14	14	14	14	14
workload5	Pearson Correlation	,226	,375	,911**	,379	,661**	1
	Sig. (2-tailed)	,436	,187	,000	,181	,010	
	N	14	14	14	14	14	14

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Bivariate correlation can help tell whether two effects, in this case VR experience and workload, occur together. This does not necessary indicate causation between the two variables but can help establish a link. Measuring bivariate correlations between experience and workload in the five tasks within SPSS yields table 3 above. This table represents 15 pairings in total, experience paired against all five workloads, as well as the workloads against each other. For this study, the correlations between XP and workload are the most relevant (top row / leftmost column). The correlation (r) is measured as a value between -1 and +1, with a negative value indicating a negative correlation, and vice versa, a score of zero indicating zero correlation.

SPSS also gives the significance for each correlation, indicating the certainty of the correlations not being influenced by other factors or variables.

Of the five tasks, all workload scores have a somewhat positive correlation, apart from task 3, which has a slight negative. The scores are as following: 0,256, 0,141, -0,100, 0,410 and 0,226, which neither indicates strong positives or negatives. Their significance however was more fluctuating, going from as low as 0,145 for task 4, to 0,733 for task 3. This would indicate that the most significant correlation of experience was found in the third task, which had the least correlation at a whole. And the one with the highest Pearson score (task 4), had the lowest significance.

6 Discussion

6.1 Research questions and hypothesises answered.

6.1.1 How do different kinds of signifiers impact user's mental workload in VR experiences?

From the mean scores across all experience groups, users clearly have an easier time finding the correct buttons when they are signified apart from the rest. Visual signification clearly outdid haptic signification, as the mean score for task 5 was 23, and the highest of the visual ones was task 2 at 14,6. Of the three visual tasks, the proximity-based positive signifier (task 2) scored 14,6. While proximity-based positive and negative (task 3) and constant positive (task 5) scored 9,9 and 7,9 respectively.

As argued earlier, the reasoning for the drop between task 2 and 3 can be attributed to the users learning about proximity-based signifiers in task 2. Which then would have made the same effect easier to negate in the next task.

With this in mind, the results in workload score for the three visual-based tasks are very similar. Thus, one could argue that all three are as good as equal when it comes to signifying action. Psotka wrote that VR helps learning by emulating real world scenarios through immersion (Psotka, 1995, p. 410). An argument for proximity-based signifiers could be that they can be conceived as less intrusive, only appearing if the user nears the object in question, and therefore more "lifelike" or immersive.

6.1.2 What kind of signifiers offer the least mental workload toward the user?

As mentioned above, based solely on mean scores, the constant visual signifier offered the user the least mental workload. But some users may perceive it as more intrusive to the whole experience, than others. The differences in workload among tasks 2, 3 and 4 were very small, and with a larger sample size the results may vary.

6.1.3 How does different kinds of signifiers impact mental workload in users of different experience levels?

As the results are, one cannot draw confident conclusions as to how users of different experience levels react to signifiers. The results seem to imply that the less experience a user has, the less impacted their mental workload is. Whether this is simply because of a low sample size, or some other factors such as a lack of motivation or some other psychological/physical reason is unclear.

Another reason for these unexpected and inconsistent results may be the groupings themselves. The groupings were changed during the data analysis, which lead to their grouping being based partly on their own accounts. And partly on the facilitator's interpretation of their explanations and recounts.

6.1.4 Improved visual signifiers will greatly reduce workload, especially concerning Mental demand.

The results seem to indicate that visual signifiers really do help reduce workload, as discussed prior. However, as for the subfactor of mental demand, it does not seem to be more affected than any of the other subfactors. As seen in table 4 below, which show all mean subscales across all tasks and experience groups.

Table 4: Table showing average subscales for all subjects.

Averages	all xp levels				
	k 1	k 2	k 3	k 4	k 5
Mental dem	51,07	23,57	17,86	10,71	33,21
physical dem	21,07	10,36	7,86	7,50	16,43
temporal dem	46,79	18,21	7,14	3,93	13,57
performance	68,21	15,71	8,21	8,93	22,50
effort	43,57	17,14	10,36	9,64	36,07
frustration	24,64	7,14	7,86	6,43	18,93

6.1.5 Inexperienced users will respond with a lower mental workload better to highlighted and color-based stimuli/signifiers.

True, inexperienced users respond better, but so does everyone else. But the inexperienced users (Group 1) responded with a lower workload on almost all keypads. Except for keypad numpad 3, which uses negative and positive color-based signification. Again, the results are not reliable due to the small sample size, and further testing is necessary to gain more insights.

6.2 Test results

The results indicate that visual signifiers greatly help reduce mental workload, as the mean scores decrease from task to task. Though, the results indicate no real correlation between experience and mental workload rating. Some correlations are stronger than others, but none can be considered strong enough to be used as solid evidence. A weakness with the test is the low number of participants, at 14. As they were again divided and spread into three sub-samples, this made the effective sample sizes for comparing experience even lower. A more substantial goal would have been around 30, something that could have been easily achievable, as the facilitator developed a solid grasp on the routine. And the locale offered plenty of potential subjects. The main issue was that the testing took place during "the 3rd wave" of the COVID-19 outbreak, encouraging testing to be done quickly.

The VR lab offered to extend the borrowing of the equipment, but it was decided that the time would be better spent analyzing results. It was during this time it became clear that the sample size was too small, but it was decided that the best course of action was to continue with analysis and discussion.

6.2.1 Skewed results in group 1

Although no strong correlations were found between experience and workload scores, the inexperienced (group 1) still scored the lowest overall on task 1, 2 and 5. And scored lower than the most experienced (group 3) at task 4. Though it is likely this is a cause of the low participation, with only four participants in group 1, other factors may contribute. A potential explanation might be that one or more of the participants were less engaged than the others, and thus did not put in as much effort.

Another factor may be that they overestimate their abilities, while the more experienced are humbler about their results. This may be a result of the Dunning Krueger effect, which proposes that the less knowledge/expertise a person has in a subject, the more likely they are to be overconfident in their abilities/knowledge surrounding it (Dunning, 2017). As truly experienced persons realize how much they do not know. Whether this could be the case here is uncertain, as again, there are no clear correlations. These factors may have something do to with unexpected results, but they are assumptive and do not offer any concrete proof.

6.3 Prototype and test design

There are questions about whether the prototype design, test setup and facilitation had any impact on the results of the study. Such as the location the study was performed, or the order of tasks and codes. These factors are as follows:

6.3.1 Study layout

As mentioned in chapter 3.5, the tests took place in a shared working space at the department of design's offices. As the tests were conducted, the participants were shielded from other students by a modular half wall, to minimize distractions. However, noise from other students and staff may have served as distractions for the test subjects. It is also possible that subjects overheard others before their own tests, which could have given them biases and "unfair" advantages.

Nevertheless, no participants admitted to being influenced by the earlier tests, though not everyone was asked about this, as it was not part of the standard facilitation. A few subjects were asked if they went in quick succession after another. More steps could potentially have been taken to negate this; however, the layout of the room offered few options than the walls already used.

6.3.2 Keypad design

A problem with the keypads could be the codes themselves. The codes were predetermined and were selected from a random number generator for it to be as random as possible. The theory being that this would help negate any sort of emerging pattern. These codes were given to the developer, who coded them into the application.

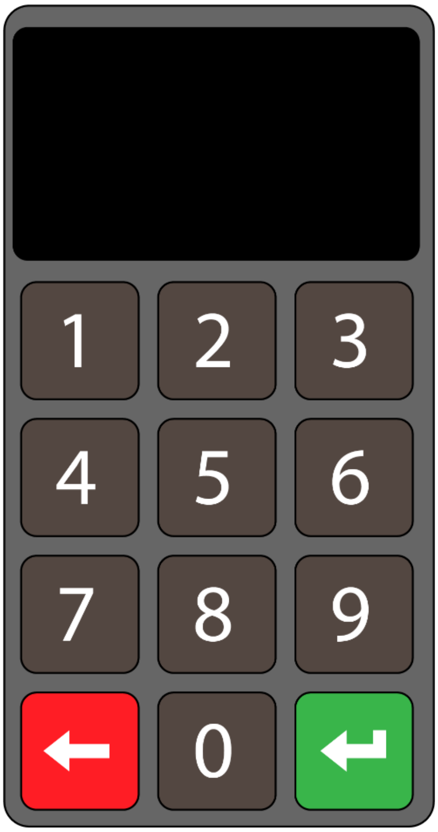


Figure 6: Illustration of keypad used in prototype.

The codes were as follows: 241, 132, 945, 736 and 075. This led to some code-combinations being more "spread out" than others.

For example, task 2 has the code "132", and uses a proximity based visual signifier. Since the correct buttons are clustered together on only one row, it would make it easier to locate them all in quick succession (see figure 6 for reference). While more users had a harder time finding the first digit on the last test, as the code was "075". The zero button was located on the bottom row, between the "enter"- and "backspace"-keys. This coupled with the fact that this signifier (haptic vibrations), differed so much from the others, led to more users having difficulties initially figuring it out. The distance between the keys could therefore also be a contributing factor to the search time, and therefore workload scores for each task.

6.3.3 Keypad order and learning

The order of which the tasks were presented may have had some effect on the workload scores as well. After the control test with no added signifiers, the order went from proximity-based positive signaling to proximity-based positive/negative signaling, and then constant positive signals. Haptic signaling was the last task, as it differed from the rest. Barring the last one, the four first tasks could be considered to ascend from most difficult to least. Meaning that the tasks became easier as they went. This is also supported from the overall test results, as workload decreased noticeably from task 1 to task 2, and so on.

During a study where medical students were tasked in repeating simulation of a complex medical operation, Takashige Abe (et.al., 2019, pp. 1-4) found that repetitions lead to lower workload (Abe et.al, 2019, pp. 5-9). As subjects reiterate a task, they will become accustomed to it, which in turn will make the procedure easier in the future. Although this study was conducted in a medical, non-VR setting, the findings can certainly be applied here. As the users conducted five similar tasks in quick succession, they may have become accustomed to the setting, giving another reason for decreased mental workload in task 2, 3 and 4.

In order to negate the factor of learning from repetition, the test could alternatively be ordered in the opposite order, going from "most easy" to "most difficult". This could have helped negate any eventual learning-based decrease in workload. As the rising workload from more difficult tasks could outweigh the potential benefits of repetition. Randomizing the order could also have been a potential solution, and randomizing the order was considered during development. But this was decided against, as it would be deemed more consistent result-wise, if all subjects were presented the tasks in the same order.

As admitted by one subject, another form of learning that may have impacted the study is the concept of proximity. Both tests 2, 3 and 5 utilized proximity-based signifiers, where the signification would only become clear once the users' hands were close enough. The drop in mean workload from 14,6 to 9,9 here may also have been affected by the users already retaining the knowledge that the signifiers can be proximity-based.

6.4 Prototype validity

During the study, it could be argued that the focus on training somewhat shifted. The field on VR in training simulations did not always stay relevant and was never explained as a use case during the tests themselves. This is mainly due to the change in prototype, as it was changed during development. This change came about as the original test of interacting with different objects in a small "office-like" space was deemed too simple, there was not any real challenge to it. And thus, the prototype became the numpad. as it was reasoned that the users would have to truly rely on the signifiers. This is because the keypad test would have the subjects interact with the signified objects (buttons) at least three times per task, one for each digit in the combination. It also helped shorten development time as less assets were required, as all the keypads used the same pre-existing model. If the prototype was developed as originally planned and built in an application like Sketchbox or Tvori, development time would have been much longer.

But although it was imagined that the original test concept would be too easy, it seems that most people had no issues completing the tasks in the final prototype either. Whether this is because this test was too simple as well, the signifiers worked as intended, or a combination of the two, is up for debate. Therefore, one can still be confident that applying the signifiers from the buttons onto other objects in other learning applications could still yield a good result. Even if the prototype changed away from the intended area, its findings and methods should still hold true.

Another option, related to the validity of the prototype, would be that with more development, both prototype concepts could be used. Seeing the effects of the signifiers in both "puzzles", like the keypad as well as natural human-like environments.

7 Conclusion

The aim for this study was to examine if users with differing degrees of experience within VR and VR-environments reacted to visual signifiers in diegetic interfaces differently. And if so, how they did. The goal for this study was set in part by three VR developers, and the goal was for it to help inexperienced users distinguish important elements in VR-environments. Especially, those made for training purposes.

A test was set up, and 14 participants divided into three experience groups participated. But the study has some clear weaknesses, mainly attributed to limitations and precautions set in place by the COVID-19 outbreak, and its 3rd wave in Norway. As well as lack of experience when it comes to programming and application development. As well as rectifications during the study, such as the change in test design and participant grouping scales after the tests were conducted.

The effects of these limitations were long development times, and limited testing capacity and timeframe. It is clear that some broad conclusions can be drawn from this study, there is no denying that the signifiers helped the users in finishing their tasks easier. But whether there are any clear differences between users of different abilities is unclear as there is little statistical correlation. For there to be clear reliable data, more participants, and clearer divisions between the subjects with regards to prior experiences.

In conclusion, one could say that this study could serve well as a pilot study to a larger test, as it has some shortcomings, such as low participation, uncontrolled factors in the environment, and vague groupings. If further studies are to take place, these things should be improved upon. Such as a separate room, clear requirements when it comes to experience within the groups, potentially even selective recruiting. Post-Covid, getting more participants should also be easier.

References

Abe, T. et. al. (2019). The effect of repeated full immersion simulation training in ureterorenoscopy on mental workload of novice operators. *BMC Medical Education*. 19 (318). https://doi.org/10.1186/s12909-019-1752-2

Byers, J. C., Bittner, A. C. and Hill, S. G. (1989) Traditional and Raw Task Load Index (TLX) Correlations: Are Paired Comparisons Necessary? in Mital, A. (ed.) *Advances in Industrial Ergonomics and Safety I*, pp. 481-85.

Dealessandri, M. (2020) What is the best game engine: is Unity right for you?, *Gamesindustry.biz.* Available from: https://www.gamesindustry.biz/articles/2020-01-16-what-is-the-best-game-engine-is-unity-the-right-game-engine-for-you (Accessed 30. Apr. 2021)

Diegesis. (2021) *Merriam-Webster*. Available from https://www.merriam-webster.com/dictionary/diegesis (Accessed 22. May 2021)

Dunning, D. (2017) We are all confident idiots, *Pacific Standard*. Available from: https://psmag.com/social-justice/confident-idiots-92793 (Accessed: 4. Jun. 2021)

Gibson, J, J. (1986) *The Ecological Approach to Visual Perception*. 2nd edition. Psychology Press.

Hart, S. G. and Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*. 52, pp. 139-183. North-Holland.

Human Performance Research Group (n.d.). TASK LOAD INDEX NASA-TLX Paper and Pencil Package. NASA Ames Research Group. Available from: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000021488.pdf (Accessed 22. Oct. 2020)

Immersion. (2021) *Merriam-Webster*. Available from https://www.merriam-webster.com/dictionary/immersion (Accessed 22. May 2021)

Jerald, J. (2015) *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.

LaViola, J. J. (2000) "A discussion of cybersickness in virtual environments." *ACM Sigchi Bulletin 32.1*: 47-56. Available at: https://dl.acm.org/doi/10.1145/333329.333344

Leedy, P. D. and Ormrod, J. E. (2015) *Practical Research: Planning and Design*. Harlow: Pearson.

Milgram, P., Takemura, H., Utsumi, A. and Kishino, F. (1994) Augmented Reality: A class of displays on the reality-virtuality continuum, *Proceedings of Telemanipulator and Telepresence technologies.* pp. 2351-34.

Nielsen, J. and Molich, R. (1990) Heuristic evaluation of user interfaces, *Proc. ACM CHI'90 Conf.* Seattle, WA, 1.-5. April. pp. 249-256

Nielsen, J. (2007) *How to Conduct a Heuristic Evaluation.* Available from: https://www.semanticscholar.org/paper/Heuristic-Evaluation-How-to-conduct-a-heuristic-How/15d167c3c55b036513f8bff3ff7a1d394536a65e?p2df (Accessed 23. Oct. 2020)

Nielsen, J. (2020) 10 Usability Heuristics for User Interface Design, *Nielsen Norman Group*. Available from: https://www.nngroup.com/articles/ten-usability-heuristics/ (Accessed 14. May 2021)

Norman, D. (2013) *The Design of Everyday Things.* Revised and Expanded edition. New York: Basic Books.

Psotka, J. (1995) Immersive training systems: Virtual reality and education and training, *Instructional Science.* 23 pp. 405-431. Doi: https://doi.org/10.1007/BF00896880

Salomoni, P. et.al (2017) Deiegetic user interfaces for virtual environments with HMDs: a user experience study with oculus rift. *Journal on Mu*ltimodal User Interfaces. 11 pp. 173-184. Doi: https://doi.org/10.1007/s12193-016-0236-5

Torenvliet, G. (2003) We can't afford it: the devaluation of a usability term, *Interactions*, 10(4), pp. 12–17. doi: https://doi.org/10.1145/838830.838857

Urke, E. H. (2018) VR og AR – en norsk introduksjon til virtual og augmented reality. 1st edition. Oslo: Cappelen Damm.

Wohlgennant, I., Simons, A. and Stieglitz, S. (2020) Virtual Reality, *Business & Information Systems Engineering,* (62) pp. 455-461. Doi: https://doi.org/10.1007/s12599-020-00658-9

Appendices

Appendix A: Consent/Information form for participants

Appendix B: Design Document

Appendix C: Cookie Recipe

Appendix D: ANOVA and association tables

Appendix A: Consent/information form for participants

Vil du delta i forskningsprosjektet

Effektiv bruk av signifiers i VR?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å *vurdere hvordan ulike typer grensesnitt påvirker brukere i VR*. I dette skrivet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Formålet med prosjektet er å teste ulike brukergrensesnitt i VR-miljøer for å undersøke hvordan de påvirker brukere av ulike ekspertisenivåer. Dvs. hva slags hint som er mest åpenbare for nye brukere, og om det er en forskjell i hvordan mer erfarne brukere sanser dette.

Denne studien utføres for en Masteroppgave i Interaksjonsdesign ved NTNU Gjøvik.

Dataene som blir samlet inn vil ikke være identifiserbare, og all rå data vil bli slettet etter den 14. Juni.

Hvem er ansvarlig for forskningsprosjektet?

NTNU Gjøvik, fakultet for design er ansvarlig for prosjektet.

Studien utføres av Vegard Wilson Dahl, masterstudent i Interaksjonsdesign, og veiledes av Ole E. Wattne.

Hvorfor får du spørsmål om å delta?

Utvalget for denne studien er funksjonsfriske myndige personer. Som en følge av pandemien, vil også kun personer fra studentens nærmeste sirkler bli bedt om å delta, samt andre studenter og ansatte fra fakultetet for design på Gjøvik.

Du har blitt spurt om å delta fordi du fyller disse kriteriene.

Hva innebærer det for deg å delta?

Å delta i prosjektet innebærer at du utfører en kort test, som tar omtrent 10-20 minutter. Denne testen består av en kort serie oppgaver. Etter hver oppgave vil du bli bedt om å rangere seks faktorer på en skala fra 0 til 100, disse faktorene beskriver din mentale arbeidslast.
Du vil også bli bedt om å rangere din tidligere erfaring med VR og 3D-miljøer, rangerende fra ingen til erfaren.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Dersom du ønsker å trekke deg fra prosjektet, send e-post eller SMS til Vegard Wilson Dahl:
Telefon: +47 40 55 40 23
E-post: vegardwd@stud.ntnu.no

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- Kun student og veileder vil ha tilgang på de rå opplysningene.
- Navn og kontaktopplysninger vil bli kodet, og disse kodene vil bli holdt separat fra resten av data.
- All data vill bli anonymisert i publikasjonen, ingen deltakere vil være gjenkjennelige.

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Opplysningene anonymiseres når prosjektet avsluttes/oppgaven er godkjent, noe som etter planen er 14. Juni. Alle direkte opplysninger og identitetskoder vil bli slettet etter prosjektets fullførelse.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke personopplysninger som er registrert om deg, og å få utlevert en kopi av opplysningene,
- å få rettet personopplysninger om deg,
- å få slettet personopplysninger om deg, og
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra NTNU har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?
Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:
- Fakultet for design, NTNU Gjøvik ved:
  - *Ole E. Wattne*

        - *+47 93 44 58 85*
        - *ole.wattne@ntnu.no*
    - *Vegard Wilson Dahl*
        - *+47 40 55 40 23*
        - *vegardwd@stud.ntnu.no*
- Vårt personvernombud:
    - *Thomas Helgesen*
        - Thomas.helgesen@ntnu.no

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:
- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 55 58 21 17.

Med vennlig hilsen

Ole E. Wattne

Vegard Wilson Dahl

Prosjektansvarlig

Student og fasilitator

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet *Effektiv Bruk av Signifiers i VR*, og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ å delta i eksperimentet
- ☐ *å gi opplysninger om min erfaring med VR- / 3D-miljøer*

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

----

(Signert av prosjektdeltaker, dato)

Appendix B: Design Document

Setting

Applikasjonen vil bestå av et lite til middels-stort rom, hvor deltakeren står i midten. På veggene rundt vil det være fem keypads, nummerert fra 1 til 5. disse numrene kan gjerne være plassert over keypadene, sånn at de er tydelig nummerert for brukerne. Brukerne må også ha en måte å komme seg rundt i rommet på, dersom knappene er langt nok unna hverandre.

Målet med testen er å teste ulike typer «signifiers» eller «feedback». Dette gjøres ved at brukeren skal taste inn den riktige koden på hver numpad, med hjelp av ulike typer feedback. De riktige kodene er tilfeldig generert, og oppgitt under.

Keypad 1

Keypad nummer én vil være kontrollpunktet, denne keypaden har ingen hjelp i form av feedback eller signifiers, dvs, ingen haptisk eller visuell feedback på hvilke knapper som er riktig.

Kombinasjon: 241

Keypad 2

Denne keypaden vil ha visuell feedback, i den form av at de riktige knappene vil lyse når fingeren «svever» over den, som illustrert i bildet nedenfor. Knappene som ikke er riktige vil ikke ha noen form for feedback.

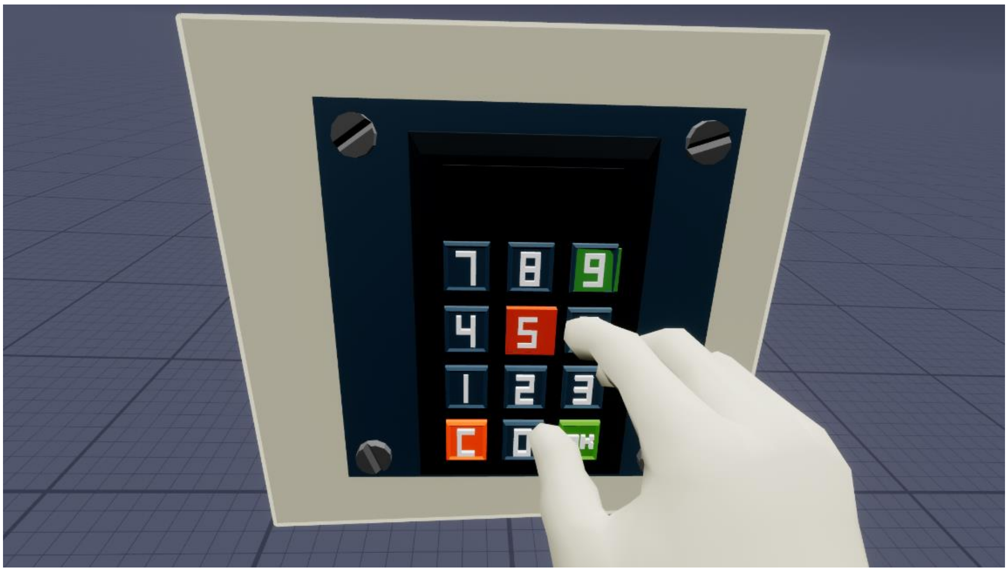
Kombinasjon: 132

Keypad 3

Denne keypaden viser feedback på alle knapper, da hver eneste knapp lyser opp når de «sveves» over, akkurat som i forrige eksempel. Forskjellen her er at den riktige knappen lyser opp i en blå- eller grønnaktig farge, mens de andre lyser opp i en rødaktig farge. Merk at knappene lyser KUN når de sveves over med fingeren.

I eksempelet nedenfor lyser to knapper samtidig, dette er kun for å vise de forskjellige fargene.

Kobinasjon: 945



Keypad 4

På denne keypaden vil det alltid være en markør det riktige valget. Markøren trenger ikke være som vist på bildet nedenfor, et alternativ kan være at de andre knappene kan være konstant grået ut.

Kombinasjon: 736

Keypad 5

Denne siste keypaden vil bruke haptisk feedback i motsetning til visuell feedback. Brukeren vil føle en liten vibrasjon i kontrolleren når de svever over det riktige alternatiet.

Kombinasjon: 075

Appendix C: Cookie Recipe

Ingredients:

- 200 g flour
- ½ teaspoon salt
- ½ teapoon baking soda
- 112 g white sugar
- 200 g brown sugar
- 1 whole egg
- 1 egg yolk
- 2 teaspoons vanilla sugar
- 112 g butter
- 200 g chocloate

Step by step

1. Mix flour, salt, and baking soda to a mix.
2. Mix the brown and white sugar to a mix, melt the butter in a pan and add it and the eggs.
3. Mix until it turns smooth and dark.
4. Add the flour mix in increments, making sure it forms a soft, heavy brown batter.
5. Cut the chocolate into thick chunks and add it to the batter.
6. Cover the batter in wrapping and refrigerate for at least an hour.

Baking

1. Preheat oven to 180 degrees Celsius.
2. Roll the batter into small clumps and spread on a tray.
3. Bake for 6 minutes, with additional increments of 2 minutes depending on size.
4. Let cookies cool.

Appendix D: ANOVA and Association tables

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
workload1 * XP	Between Groups	(Combined)	195,074	2	97,537	,405	,677
		Linearity	186,166	1	186,166	,773	,398
		Deviation from Linearity	8,908	1	8,908	,037	,851
	Within Groups		2649,937	11	240,903		
	Total		2845,011	13			
workload2 * XP	Between Groups	(Combined)	175,180	2	87,590	,736	,501
		Linearity	29,380	1	29,380	,247	,629
		Deviation from Linearity	145,800	1	145,800	1,225	,292
	Within Groups		1309,363	11	119,033		
	Total		1484,544	13			
workload3 * XP	Between Groups	(Combined)	42,842	2	21,421	,142	,870
		Linearity	17,195	1	17,195	,114	,742
		Deviation from Linearity	25,647	1	25,647	,170	,688
	Within Groups		1663,731	11	151,248		
	Total		1706,573	13			
workload4 * XP	Between Groups	(Combined)	494,837	2	247,418	1,619	,242
		Linearity	365,723	1	365,723	2,393	,150
		Deviation from Linearity	129,113	1	129,113	,845	,378
	Within Groups		1681,107	11	152,828		
	Total		2175,944	13			
workload5 * XP	Between Groups	(Combined)	540,641	2	270,321	1,290	,314
		Linearity	156,410	1	156,410	,747	,406
		Deviation from Linearity	384,231	1	384,231	1,834	,203
	Within Groups		2304,741	11	209,522		
	Total		2845,382	13			

Measures of Association

	R	R Squared	Eta	Eta Squared
workload1 * XP	,256	,065	,262	,069
workload2 * XP	,141	,020	,344	,118
workload3 * XP	-,100	,010	,158	,025
workload4 * XP	,410	,168	,477	,227
workload5 * XP	,234	,055	,436	,190

NTNU

Norwegian University of
Science and Technology