KTH ROYAL INSTITUTE
OF TECHNOLOGY

Degree Project in Technology

30 credits

Comparative Analysis of Spatiotemporal Playback Manipulation:

Evaluating Desktop Environments versus Immersive Head-Mounted Virtual Reality Environments

ANDREAS WRIFE

Stockholm, Sweden, 2023

Comparative Analysis of Spatiotemporal Playback Manipulation:

Evaluating Desktop Environments versus Immersive Head-Mounted Virtual Reality Environments

ANDREAS WRIFE

Master's Programme, Computer Science, 120 credits
Date: December 11, 2023

Supervisor: Alessandro Iop Supervisor
Examiner: Mario Romero Vega
School of Electrical Engineering and Computer Science
Swedish title: Jämförande analys av rumslig och tidsmässig uppspelningsmanipulation:
Swedish subtitle: Utvärdering av datorskärmsmiljöer mot miljöer inom huvudmonterade virtuell verklighet

Abstract

Virtual Reality (VR) is a creative tool that enables immersive learning, planning and training of surgical operations. Extensive research has been conducted in multiple surgical specialities where VR has been utilized, such as spinal neurosurgery. However, cranial neurosurgery remains relatively unexplored in this regard. The thesis project presented here explores the impact of adopting VR, using a headset and controllers, to study the cranial neurosurgical procedure of External Ventricular Drainage (EVD). In this study, pre-recorded Motion Captured (MoCap) data of an EVD procedure is visualised on a desktop monitor as well as through a VR headset. Participants were tasked with identifying and marking one key moment in the recordings. Both objective and subjective metrics were recorded, such as completion time, accuracy, precision, the usage of different interaction controls as well as through the use of a questionnaire. The comparison is done on an objective and subjective scale, analysing user performance and User Experience (UX). The results from the experiment showed that the task was completed on average twice as fast in VR compared to desktop. However, desktop showed more promise in having higher accuracy and precision. Subjective feedback showed a slightly higher preference towards the VR environment concerning system usability. However, the settings were equally comparable in terms of task load. Furthermore, a guidance laser introduced to help with depth perception showed no increase in user performance. In conclusion, VR displays promise as an alternative tool to be used for planning and educational purposes in cranial surgery. Potential future developments could focus on the increased precision in interactive in VR, with the aid of haptic feedback, minor adjustments and scalability.

Keywords

virtual reality, surgical simulations, external ventricular drainage, motion capture, interaction controls

Sammanfattning

Virtuell verklighet (VR) är ett kreativt verktyg som möjliggör ett fördjupat lärande, planering och träning av kirurgiska operationer. Omfattande forskning av VR har utförts inom kirurgiska specialiseringar, såsom spinal neurokirurgi. Den kraniala neurokirurgin är dock relativt outforskad i detta avseende. I det här examensarbetet undersöks effekten av att applicera VR för att kunna studera det kraniala neurokirurgiska ingreppet kallat Externt Ventrikulärt Dränage (EVD). I denna studie visualiserades förinspelad Motion Captured (MoCap) data från en EVD-procedur på en stationär bildskärm såväl som genom ett VR-headset. Deltagarna fick i uppdrag att identifiera och markera ett nyckelmoment i inspelningarna. Både objektiva och subjektiva mått registrerades, såsom slutförandetid, noggrannhet, precision, användningen av olika interaktionskontroller samt genom användning av ett frågeformulär. Jämförelsen gjordes på en objektiv och subjektiv skala, med en analys av användarprestanda (UP) och användarupplevelse (UX). Resultaten från experimentet visade att uppgiften genomfördes i genomsnitt dubbelt så snabbt i VR. Dock visade uppsättningen med vanlig bildskärm mer lovande i form av att vara mer exakt i markörplacering och identifiering av bildrutorna. Användarenkäten visade en något högre preferens för VR-miljön vad gäller system-användbarhet. Miljöerna var dock lika jämförbara när det gäller uppgiftsbelastning. Dessutom visade en styrlaser ingen ökning av UP, fastän den introducerades för att hjälpa till med djupsyn. Sammanfattningsvis visade VR lovande resultat för att kunna användas som ett alternativt verktyg för planering och utbildningsändamål inom kranialkirurgi. Potentiell framtida utveckling kan fokusera på ökad precisionen av svårhanterlighet i VR, med hjälp av haptisk vibration, justeringar för precision och skalbarhet. Dessutom kan det undersökas hur de två systemen jämförs när båda optimeras efter och drar nytta av sina egna styrkor.

Nyckelord

virtuell verklighet, kirurgiska simuleringar, extern ventrikeldränage, motion capture, interaktionskontroller

Acknowledgments

I am grateful to my supervisor, Alessandro Iop, who has provided me with support and guidance throughout the project. He has shown understanding and patience whilst providing me with resourceful feedback and valuable suggestions for the report and the project itself. Thank you for your supervision.

I extend my appreciation to Mario Romero Vega for taking the time to review and evaluate this thesis. Your constructive criticism have helped this thesis achieve greater heights. I would like to thank you for allowing me the opportunity to examine deeper into the field of Virtual Reality and the depths of its application.

To my friend and colleague, Alexander Danielsson, I thank you for your agency and friendly support. It was a pleasure working side by side with you. Good times in Akvariet.

Lastly, I want to express my gratitude to my Astrid, who has shown thorough support throughout the project. Pushing me to greater lengths and giving constructive feedback on my work.

This thesis would not have been possible without the support and encouragement of these remarkable individuals, my family and their belief in my capabilities. Thank you all for being a significant part of this academic milestone.

Stockholm, December 2023
Andreas Wrife

Contents

List of acronyms and abbreviations

AR	Augmented Reality
EVD	Extra Ventricular Drainage
GEQ	Game Experience Questionnaire
HMD	Head Mounted Display
HUD	Heads Up Display
IEQ	Immersive Experience Questionnaire
IQR	Interquartile range
MoCap	Motion Capture
MPE	Marker Placement Error
NASA-TLX	NASA Task Load Index
ROAM	Rotate Around the Marker
SUS	System Usability Scale
UI	User Interface
UX	User Experience

VR	Virtual Reality
XR	Extended Reality

Chapter 1

Introduction

1.1 Background

Since the invention of Virtual Reality (VR) in the 1970s, the technology has been adopted by various development, entertainment and research sectors. One such area is the healthcare industry, where it is used for educational purposes in preoperative planning and training in surgical operations. The technology has shown substantial potential [1]. However, while VR has become a widespread tool in several surgical specialties, its implementation in neurosurgery has been limited [2]. Furthermore, there is a shortage of evidence supporting the effectiveness of VR in this field [3].

In surgical applications, VR has shown a promising role in simulating surgery procedures in a controlled environment where mistakes are non-fatal. Here, haptic and force-based operation* imitate reality and help sell the sense of true touch [4]. Examples of tools in the healthcare system, which utilise these methods are the NeuroVR [1, 5] and the Dextroscope [6]. Nonetheless, the acquisition of such equipment come a substantial cost and is exclusively accessible within select medical educational institutions and healthcare facilities [1].

Extensive research had been deployed in surgical specialities, such as spinal neurosurgery. However, The cranial neurosurgery remains moderately unexplored with respect to VR technology [2]. An important cranial neurosurgical procedure is the Extra Ventricular Drainage (EVD), which includes draining the ventricle in the brain by puncturing the lateral one with a catheter. Moreover, the utilization of Motion Capture (MoCap) technology in

*tactile response to virtual interaction, usually in the form of quick vibrations in the handheld-tool.

visualizing neurosurgical procedures represents an unexplored frontier, which could hold beneficial practise for independent asynchronous learning.

By examining the MoCaped EVD procedure through both VR and Desktop environment, users can view the 3D recording from various angles and perspectives. Thus allowing for a more comprehensive understanding of the device's position and relationship with surrounding anatomical structures. Since there is a lack of research with how people interact with VR and Desktop in neurosurgical procedures, this thesis aims to shine some more light on the issue. This can then aid in the assessment of the drain's effectiveness, potential complications, and may assist in surgical planning or adjustments. Thus improving patient care [2].

1.2 Research question

For the task of manipulating spatial temporal playback of pre-recorded MoCap data, how do Desktop environments compare to immersive head-mounted VR environments? We specify the user, the task, the media, and the evaluation metrics below:

- Experimental Constants
  - Participants: KTH students modeling medical students,
    - Full stereo and color vision
    - 3D competence
    - Previous VR experience
  - Task: manipulation of the spatial temporal playback perspective of the pre-recorded MoCap neurosurgical EVD procedure, with the goal of identifying the point in time and space at which the catheter penetrates the targeted ventricle.
- Experimental Independent Variables
  - Experimental control: simulation on a 2D screen with keyboard and mouse controls
  - Experimental treatment: simulation in an immersive VR environment with Heads Up Display (HUD) and handheld controls
- Experimental Dependent Variables
  - Objective

    * Time to complete the task
    * Space-time Accuracy
    * Space-time Precision
    * Steps taken
  - Subjective
    * USABILITY - System Usability Scale (SUS)
    * IMMERSION - Immersive Experience Questionnaire (IEQ), Game Experience Questionnaire (GEQ)
    * ENGAGEMENT - IEQ, GEQ
    * COGNITIVE LOAD - NASA Task Load Index (NASA-TLX)

1.2.1 Hypothesis

Initially it is expected that the manipulation of the spatial temporal MoCap data will be easier to understand in the Desktop* environment (keyboard and mouse). This is due to that the average user is more accustomed to a desktop computer monitor over a VR Head Mounted Display (HMD) [7, 8].

However, after repeated trials the completion time of each task will be shorter in VR. The immersion and controls should allow for an easier access to the data at hand [8, 9]. Unlike the Desktop environment, the VR controls are detached from the viewers perspective. For instance, the marker can be placed in different spots while the user has the same viewpoint. In desktop, the user has to rotate the camera to change the placement of the marker. Thereby enabling the user to better focus on finding the time and space variables.

A guidance laser will provide the user with visual clues to better judge distances in the virtual world. As it will be projected out from the camera in the Desktop scene, and out from the virtual controller in the VR scene, the viewer will know what the relative distance towards object is better. The perception of distance can be enhanced through several factors, including the relative size of objects in relation to the laser, the potential for interposition between the laser and objects, and the linear perspective provided by the laser.

Thus:

- **H1** The controls to manipulate the spatio-temporal data are easier to understand on a desktop environment.

*Throughout the document, the Desktop setup will be denoted with capital D to increased clarity for the comparison with the VR setup.

- **H2** The task's time to completion is smaller in a virtual environment,
- **H3** The task yields higher accuracy and precision in both space and time in the virtual environment.
- **H4** The guidance laser will help with perceiving the relative distance to objects.

1.3 Purpose and Goal

The objective of this project is to analyze the effectiveness of visualizing and manipulating a pre-recorded MoCap of a surgery using both desktop and VR technologies. The aim is to determine whether there are any disparities between the two settings, and whether a 2D monitor is just as effective as a VR environment. Additionally, the project seeks to gain insight into how to provide a more comprehensive and immersive visualization, specifically regarding the placement and trajectory of the catheters during surgery.

Specifically, the primary goal of the project is to investigate differences in interaction and viewing experience between VR and 2D monitor when working with a pre-recorded MoCap of an EVD. A secondary goal is to develop an environment in Unity that allows for viewing and interacting with the data, enabling users to navigate through the recording and pinpoint a specific time and location. For future analysis, this functionality should be available in both the VR and Desktop environments. It should be able to record the total time taken, the chosen frame, the coordinates of the marker and the different steps taken to get to that point.

1.4 Research Methodology

The research methodology for this thesis project consists for the most part of a user study. Two techniques of visualising and manipulating spatio-temporal data in a Desktop and VR environment are tested and analysed. Each test consists of a sequence of trials in which the participant completes a predefined task.

During each trial, participants is asked to position a spherical marker at the point where they believe the catheter initially makes contact with a purple-coloured ventricle, see Figure 1.1 for more detail. Upon completing the task, the participant moves on to the next trial. The participant completes two sets of six consecutive trials, one in each environment. The task allows for the comparison between VR and Desktop, whether it is easier to identify critical moments in surgical simulations and how controllers are used to interact with the animation. By repeating the same trial multiple times, a threat to validity can be minimized. Repeated trials gives a better overview of the learning process.

The measured data includes the location of the marker placed by the participant together with the frame of where they thought the catheter first made contact. The completion time is also captured as well as the frequency and duration of specific control inputs. Additionally, users will complete a three-part questionnaire, assessing the usability, immersion, engagement, and cognitive load of both techniques.

By examination results from the dependent variables and assimilating user feedback, conclusions can be drawn regarding which setting yielded better results in terms of time, accuracy, precision, and number of steps executed. Specifically, ANOVA tests and in-between subject tests can be conducted, which can yield a comprehensive overview of both environments and their concordance with the formulated hypotheses.

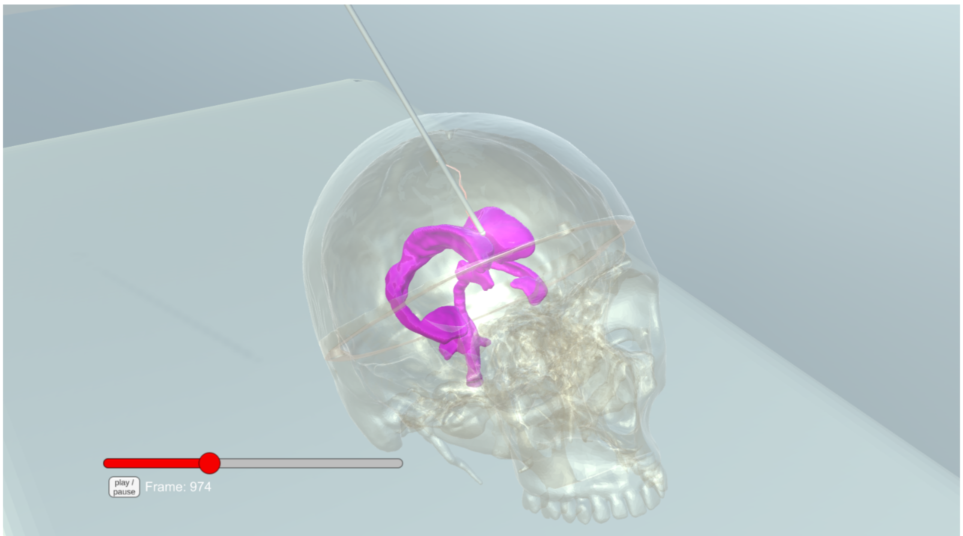

Figure 1.1: Overview of the transparent skull, with the catheter having entered the brain, touching the ventricle (seen in purple), currently the animation is on frame 974, displayed under the red slider in the left hand corner.

1.5 Delimitations

In order to limit the scope of this project, certain restrictions have been put in place. The user study includes one type of VR headset, namely the HTC Vive and no comparison is made with other headsets.

This study will only consider one type of surgical procedure, being the EVD. However, in the user study participants are presented with several different recordings of the operation.

Compared to another study [10], only one type of rendering style will be used since the primary focus is to compare desktop with VR. Furthermore, the task will only include placing the marker and finding the correct frame.

The systems are limited in interaction controls to ensure a standardised comparison between the two systems. Furthermore, the VR controller has a limited amount of triggers and buttons that are available to be mapped to specific key-binds. However, the user should not be overwhelmed with different manipulation controls, which is another reason to limit the amount.

Chapter 2

Background

The structure of the background section can be categorised into four main areas. The technology of VR is described in Section 2.1. Additional topics encompassing VR are also introduced, such as how it is applied in education 2.1.1 and its adherent concept of motion sickness 2.1.2. In order to correctly interpret information visualised on a flat 2D monitor or viewed through VR, different perception cues can be used. These are described in Section 2.2. The following category, see Section 2.3, describes way to visualise the spatio-temporal data, which include using MoCap, defined in Section 2.4. The captured data was of a minimally invasive surgery, which is described in Section 2.5. Specifically, the operation is called EVD, which is explained in Subsection 2.5.1. Lastly, the different questionnaires that were used are detailed in Section 2.6. These include the SUS 2.6.1, the NASA-TLX 2.6.2, IEQ and the GEQ 2.6.3.

In the Related Work Section 2.7, three papers are exclusively discussed. Section 2.7.1 examines the paper which inspired the thesis, which is a systematic review of the application of Extended Reality in neurosurgical education. The following related work, see Section 2.7.2 investigates a previous study in which an operation was recorded using 360° cameras and then shown to participants. The last entry, see Section 2.7.3, describes a developed tool to view and interact with visualised neurons in VR.

2.1 Virtual Reality

The technology of VR aims, at its core, at applying real world perception to an entirely artificially computer generated environment [11]. Since the 1970s, it has been an ever growing field of research that has been iterated upon in a variety of different sectors. The technology started of in the military and automobile divisions, for flight and vehicle simulations [12]. Subsequently, it found adoption within the realms of healthcare, research and educational systems. Eventually, even the gaming industry also adopted it.

In healthcare, the main application for VR is to simulate operations. Popular, yet expensive simulators include the NeuroVR [1, 5] or the Dextroscope [6] to train pre-surgery. The main pushing force that allowed VR to become a more widespread, consumer friendly technology is the interest it gained in the gaming culture [11]. Here, three trends can be identified which allowed VR to gain follower traction in various sectors, such as for the domain of surgical education.

- **Improved Graphic cards.** Better graphics cards means higher refresh rate for a closer to real-life experience. Additionally, this improvement minimizes the occurrence of motion sickness, as the screen remains closely synchronized with real-world motion, reducing the perception of lag.

- **Better visual and ergonomic comfort.** The most recent headsets have also focused on improving ergonomics with face molds and head straps. Additional choices are available, such as adjustable lens distances and external glass integration. Moreover, different headsets on the market come with different controllers, which gives the consumer a wide variety of ergonomic choices to fit their needs.

- **Widespread consumer availability.** As an increasing number of developers choose to create games for the VR market, and as companies continue to manufacture superior headsets, the cost-effectiveness of the typical consumer headset has seen an improvement. Furthermore, besides VR headsets that require a Desktop PC to run, the industry also offers standalone headsets which are even more affordable and mobile.

When compared to traditional 2D monitor viewing, VR offers a unique viewing experience where the user can, with sometimes high spatio-temporal realism, view information like they are present in the same room as it. Usually this can cause a ”wow” moment when experiencing it for the first time [11]. This means that it is valuable to note if the user has any previous knowledge and usage of a VR headset. The focus might otherwise lie on the immersive experience itself and not the prepared task.

2.1.1 VR in Education

Closely related to the subject matter is the application of VR in the realm of education. According to Ros et al. [7] there is an assortment of benefits to using VR in education. These include faster and better understanding, improved retention and longer lasting recall, whilst keeping students more involved and immersed [13, 7, 2]. Additionally, in comparison to Desktop viewing, VR provides an immersive first person viewing experience, for instance when it is applied to surgery [14]. Andersen et al. demonstrated that the utilization of VR prior to a training course, specifically in the context of an surgical operation, led to improved student performance and greater advantages, as evidenced in their study [15]. VR helps users better understand the content at hand, since the immersive experience also reduces outside influences, task-unrelated images or thoughts. *"Indeed, blinding students to the actual environment increases their focus"* [16]. Consequently, content showed in VR should only consists of relevant data. The key is to understand when VR should be applied: prior to, following, or during a lesson. [3].

However, the benefits of employing VR for education is not consistent across all domains or among various authors. According to Van Der Heijden [9], highly immersive environments might not result in higher learning and transfer outcomes compared to more traditional methods. Van Der Heijden showed that applying immersive VR to a science lab simulation leads to more presence but less learning. As mentioned earlier, although the participant is more immersed [7, 11], this also results in increased extraneous cognitive load, which in turn hinders learning and transfer at lower levels [9].

The theory presented introduced the concept that information systems can be perceived as either self-fulfilling or practical. With self-fulfilling, the system is developed for fun and pleasurable experiences, whilst a practical system acts to provide more useful information rather than being attractive. In VR, the distinction between the two information systems is not always clear, which in turn makes users treat them as more entertaining than practical. The consequences of this could result in a disregard of the instrumental value, where the focus of the participant is shifted towards the entertainment value of the system. In other words, rather than expending their cognitive load on the task, their attention is redirected toward extraneous information that is unrelated to the test. Furthermore, simply porting an experience intended for a desktop experience over to a VR environment may either facilitate or impede the learning process.

Disregarding the cognitive load difference, more immersive environments, containing appealing detail and visually engaging graphics, could allow for more interest in the healthcare sector. It was shown to be more captivating for the user. [9].

Ultimately, VR can not replace traditional education, but can provide value as a supplementary tool [7].

2.1.2 Motion Sickness

Motion sickness, or more specifically Cybersickness is a common issue when considering immersive VR. Linked to a cluster of issues concerning VR, frequently stemming from the disjunction between actual and perceived physical motion, cybersickness can leave the VR user with oculomotor issues, nasuea or disorientation [17]. Oculomotor issues are primarily caused by eyestrain from continued and prolong usage of a HMD. Since the screens of the VR headset are located close to the eyes, yet can display and simulate large distances, the eyes are sometimes forced to focus on virtual objects discomforting close. Meanwhile, it can also overwork the nerve that controls eye movement due to the HMD lenses curving the optical input. These issues can cause the eye to rotate more than in a regular setting.

Disorientation issues adhere to the reality that body orientation is usually detected and understood visually. However, when using an immersive HMD that visual input is lost. Even when using an virtual avatar, traces of disorientation can appear since the body mapping is not always synced 1:1, In other words, that the ingame character's libs are not tracked or bend analogous to reality [18].

The muscular sensory system and the semicircular canals located inside each ear are the two other factors that are connected to disorientation and nausea. The muscles communicate its orientation and perceived force to the brain depending on how gravity affects the internal muscular sensation. The inner ears' semicircular canals measures tilt and acceleration in six principal axes (Roll, Pitch and Yaw), giving access to the sense of both forward and rotational motion [19]

Virtual Headset and game developers have tried to minimise the effect of cybersickness by implementing both software and hardware solutions. The hardware solutions include, among other thing, lens calibration and body tracking sensors. Commonly integrated on more expensive VR HMDs, being able to adjust the lens distances, both relative to each other and also towards the screen can help with eyestrain. Furthermore, allowing users to wear glasses underneath the headset may also contribute to a better experience.

The software solutions consists of visual techniques and tricks to help the brain minimise the perception of faulty motion. From a VR developer's standpoint, the most prevalent strategy for mitigating cybersickness is to minimize the necessity for real-life walking. This is often solved within virtual environments by teleportation. Instead of walking, the user points with a VR controller towards a targeted location, which upon they will be virtually teleported to the selected area [20]. Other motion solutions include walking-in-place and tunneling [21]. Tunneling is the act of which only a portion of the HMD screens are used when navigating using controllers. Whenever the user wants to relocate within the virtual environment, only a small viewing window perceives motion. The outer edges of the display, part of the eyes peripheral vision when staring at the center of the screen, stays fixed at the origin of the motion [18]. This may also increase rendering efficiency since the rest of the screen can be kept a still image or blurred, further strengthening the User Experience (UX) [22].

Other implementation techniques encompasses guidelines and recommendations for virtual environment designs. Avoiding enclosed environments such as narrow winding corridors and instead adopt open landscapes in which the eyes have time to perceive and process. Adding a distant object in the background, such as a mountain, allows the mind to anchor itself to the VR and perceive distance and motion easier. Another approach is to add guidelines such as lasers from the controllers and/or headset. These guides allow the wearer to understand distance towards object more accurately, as the length of the laser decreases when directed towards objects closer to the user. [23].

2.2 Perception

In order to construct an environment easily readable for the human eye, certain techniques can be applied. Monocular cues and binocular cues are helpful perception techniques that allow depths, relative distance and three-dimensional space to be judged. Monocular cues include relative size, interposition, linear perspective, aerial perspective, light and shade and monocular movement parallax [24]. These can be integrated in visualisation on a 2D flat monitor and on a HMD. Lastly there is stereopsis, which holds significant importance among binocular cues. [25], which is present in HMDs.

Relative Size The concept of relative size relates to how objects that are more distant from an observer tend to appear smaller compared to those that are closer to the observer. This phenomenon also applies to the same object, like a car driving away down a street. Whilst the buildings around keep their size, the car becomes smaller and smaller. This is interpreted as the car getting further and further away. Known as size constancy, an object retains its relative size no matter the distance [24].

Interposition The cue of interposition occurs when an object is occluded by another object. The object obscured from sight is then judged as being further away from the observer [24].

Linear perspective Linear perspective details the visual perception of lines vanishing into the distance. As parallel lines from objects such as roads disappear into the distance the lines converge into one point. The perceived angle between the parallel lines also gives some indication on the travel distance [24].

Aerial Perspective Aerial Perspective regards how the human eye judges distance based on the relative colour of objects. When outside and observing objects further away, distant objects appear more blue due to light being scattered in the atmosphere. For instance, mountains further away appear more blue.

Objects can also be perceived as distant or close based on the contrast they create. In cases where the outlines of objects appear blurred due to the scattering of light, they are perceived as distant. Scattering causes the darker parts of an object to have its contrast reduced. For instance, mountains are perceived as closer when the atmosphere is clear [24].

Stereopsis Stereopsis is the term used to describe the capacity to perceive depth. It involves distinguishing the relative distances between objects that seem to have physical displacement. While it is possible to determine the relative position of objects using one eye through monocular cues, it is the horizontal displacement of the eyes that yields two slightly different perspectives of the same object, or disparate images, enabling precise discrimination of depth through stereoscopic vision [25].

2.3 Spatio-Temporal Data

In this thesis, spatio-temporal data are visualised. Spatio-temporal data are data which inherits both space and time properties. In order to display such data Coffey et al. proposed three different design configurations. Each one adopted for both the time and space dimension respectively [26].

In order to maneuver through temporal data, the three main time-designs presented by Coffey et al. are as following [26]:

- The user can interact and control time. This is performed by changing a knob or slider in order to advance the scene forward or backwards in time [27].
- The user sees time pass by as an animation, which can be replayed and/or is on a loop [16].
- Time is shown as static time stamps. Often displayed in multiples within a given scene, giving the viewer an overview of the events [10].

The scene can be manipulated in different ways to facilitate spatial judgments from various perspectives. The three main ways according to Coffey et al. are [26]:

- The user can control the scene interactively by rotating and repositioning it. This enables the user to view the scene from multiple angles [10, 26].
- The scene rotates automatically to create motion parallax. This can be accomplished with a constant rotation or an oscillating movement of presented space. Such movements help provide a sense of depth within the scene. Objects further away move slower and objects in the foreground could cause occlusions over items in the background [16, 26].
- The scene is fixed in a stationary position and orientation, where additional visualization widgets can implemented to bestow certain depth cues [27, 26].

2.4 Motion Capture

MoCap involves capturing a specific type of spatio-temporal data, which is the movement of agents or objects in a confined environment, through the use of sensors. The method encompasses several techniques in order to record motion. The most common approach is to use optical sensors to track infra-red reflective markers in 3D space, which are attached to the objects or subjects of interest, for instance on the joints of a dedicated MoCap suit. Further, the movement of the markers is captured at high rate and high resolution using infrared cameras [28].

It is also crucial to accurately determine the camera positions within the room relative to the subjects and tracked objects. Positioning the cameras to minimize obstruction between them and the markers is essential. Additionally, the distance between the camera and the markers is noteworthy, as a shorter distance results in higher pixel accuracy. [29].

2.5 Minimally Invasive Surgery

Surgical education/training using VR technology has become increasingly important as doctors now perform operations in smaller, complex anatomical areas. These include organs such as the heart, brain, pelvis and spine, each containing critical blood vessels and nerves. Minor errors during surgical procedures can result in significant repercussions. Consequently, surgical navigation systems have been developed and refined over several decades to enhance surgical precision, decrease operation duration and trauma, and enhance overall success rates. Intraoperative navigation using these systems is expected to be an important area of future research in surgery [30].

Li et al. [31] demonstrated the reliability and clinical feasibility of a navigation system based on Mixed Reality technology for sinus and skull base surgery. This new system is more effective than conventional systems in terms of operation time and surgeon workload, especially for less experienced surgeons. The Mixed Reality-based navigation technique offers individualized preoperative planning and real-time 3D navigation during surgery, making it more accurate, minimally invasive, and efficient.

2.5.1 Extra Ventricular Drainage

The minimally invasive surgery visualised and studied in this project is called an EVD. The procedure is one of the most common and valuable lifesaving operations in the neurological intensive care unit. It is executed on patients with raised intracranial pressure or suffering from intracranial hypertension. The cause of the pressure is due to blockage in the outflow of the cerebrospinal fluid in the brain. This increases the pressure, and when the tension becomes too high an EVD may be needed. The hypertension is caused by intraventricular blood, which provokes a disruption in the equilibrium between production and absorption of the cerebrospinal fluid [32].

In order to relieve said pressure in the ventricle, a needle is used to punctuate it. Additionally, a needle is encased inside the catheter, aiding in draining the cerebrospinal fluid. The needle initially penetrates the interior of the cranium after passing through a drilled hole situated on the upper portion of the patient's skull. Upon reaching the designated ventricle, the needle's tip is carefully extracted, thereby retaining solely the catheter injected into the ventricle. The needle is inserted no more than $7cm$ into the cranium. Next, the catheter is connected to an external drainage system in order to extract the cerebrospinal fluid.

Throughout the course of the surgical procedure, it is imperative to acknowledge the potential for complications to manifest. The majority of surgeons predominantly employ unguided, freehand techniques that rely on anatomical landmarks for both alignment and insertion when performing an EVD surgery. In 10% to 40% of operations complications such as hemorrhage and inadvertent placement into brain tissue are reported [32]. Consequently, technological advancements in computed tomography, ultrasound, endoscopy, and navigation in neuroscience are critical. They play a pivotal a role in increasing the accuracy and efficiency of the procedure, thereby mitigating the potential for future complications in EVD placement [32].

2.6 User Experience Questionnaire

In order to evaluate the implemented systems and obtain a subjective evaluation metric of the UX, a variety of assessment approaches can be used. For this project, the metrics of importance for the research question include immersion, engagement, usability and the cognitive load of the experience. One of the most adopted approaches is the use of surveys, where the participants self-report by answering questions and using Likert scale statements. To report on immersion and engagement, VR and Augmented Reality (AR) questionnaires like IEQ and GEQ can be used [8, 33, 34]. For usability, the SUS is often utilized [35]. Lastly, for cognitive load, a load task index survey like the NASA-TLX is generally recommended [36].

2.6.1 System Usability Scale

The System Usability Scale consists of ten 5-point Likert scale questions. Starting at 1 (negative/disagree) to 5 (positive/agree). The issues can be calculated and analysed by themselves, but also together to give an accumulated usability score. Questions alternate in their assessment: every

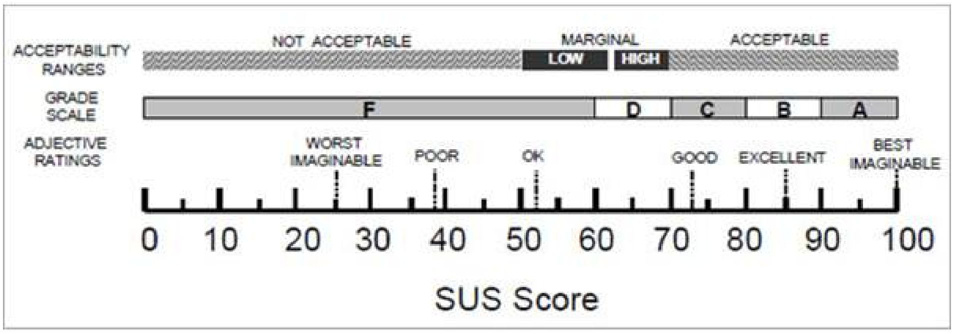

Figure 2.1: Grading scale for SUS scores [39].

other question is regarded as better the higher rated it is, whilst the remaining questions contribute to the final score the lower scores they receive. See Section 3.4 for full list of questions. The accumulated usability score ranges from 0 to 100. It is calculated using Algorithm 1 [37]:

Algorithm 1: Calculate SUS Final Score.

Function `FinalScore` $(Q = \{Q_1 - Q_{10}\})$:
- $Q_{SUS} \leftarrow 0$
- **for** $Q_x \in \{Q : x \pmod 2 \equiv 0\}$ **do**
  - $Q_{SUS} \leftarrow Q_x - 1$
- **for** $Q_x \in \{Q : x \pmod 2 \equiv 1\}$ **do**
  - $Q_{SUS} \leftarrow Q_x - 1$
- $Q_{SUS} = Q_{SUS} \cdot 2.5$
- **return** Q_{SUS}

Based on statistics from Jeff Sauro, the average rating for a system evaluated through SUS is 68 [38]. SUS can also be seen as from a grading descriptive, where an A-F grading scare is derived. A score of 100-90 being A, 89-80 B, etc. See Figure 2.1 for more details.

2.6.2 NASA Task Load Index

NASA-TLX consists of 6 questions, with each rated on a scale between 1 and 20. See Section 3.4 for a full list of issues. In its full version, the average of each scale rating is weighted depending on how the participant rates the importance of each factor in relation to the perceived workload. The individual ratings are accumulated to give an ”integrated measure of overall

Table 2.1: The Interpretation Score of NASA-TLX [42]

Workload	Value
Low	0-9
Medium	10-29
Somewhat high	30-49
High	50-79
Very high	80-100

workload" [40]. However, due to its time-consuming nature, a commonly employed modified version is utilized. Typically known as RAW NASA-TLX, it ignores the the weighting process and proceeds to value all components equally [36]. The pen and paper version of the NASA-TLX features a line with 21 marks for each of the six scales. The digital version can be similar to the SUS with less marks per scale [41]. For the 21 mark version the following algorithm calculates the final RAW NASA-TLX score, in other words without the individual weights:

Algorithm 2: Calculate NASA-TLX Final Score

Function `FinalScore`(Q)**:**
  $Q_{final} \leftarrow 0$
  for $Q_x \in \{Q\}$ **do**
    $Q_{final} \leftarrow (Q_x - 1)$
  return Q_{final}

Based on statistics presented by Hancock et al., the perceived workload of a system in accordance to NASA-TLX follows Table 2.1, [42]. The estimated workload depends on the final score accumulated over the six questions. Ranging from 0 to 100, the workload is interpreted from low to very high.

2.6.3 IEQ and GEQ

The IEQ and GEQ are similarly setup in order to evaluate the participants immersion, engagement and experience. The IEQ consists of 32 questions, on a 5-point Likert scale [33]. The GEQ consists of 19 questions on a 7-point Likert scale [34]. The GEQ is tailored towards gaming experiences, whilst the IEQ aims to capture the engagement and immersion of any UX [33, 34].

2.7 Related Work

2.7.1 Extended Reality in Neurosurgical Education: A Systematic Review

The main research paper that prompted the development of the current Master's thesis was a systematic review by Iop. et al. [2]. The authors analysed 31 recent studies in the application of Extended Reality (XR) technologies to cranial neurosurgical education.

According to the authors, extensive research has been carried out in other surgical specialities, such as spinal neurosurgery. However, cranial neurosurgery remains relatively unexplored when considering the application of XR technologies to educational practices. The systematic review showcases papers highlighting the potential educational benefits of applying XR technology to cranial neurosurgery. Including XR as a complement to regular teaching and guidance, as well as its own stand-alone educational system.

Most papers reviewed, which included an experimental testing setting, involved participants with medical education, residents and expert neurosurgeons. Only a small fraction (2 articles or 6%) relied on participants without medical training. In general, the less experienced the users were, the more they appreciated the XR technology.

The systematic review categorised the studies into three domains of knowledge: the learning of new skills, the practise or improvement of existing skills, and the assessment of skills. The primary outcomes of these studies, at the same time, can be classified as measures of user performance or UX. The user evaluated their performance using objective metrics related to skills in dexterity and hand-eye coordination. In contrast, UX focused on subjective metrics, including usability and appreciation of the User Interface (UI).

The majority of studies were based on stationary and flat monitors, while a few used HMDs. Most of the studies included VR technologies (74%), while only a minority (10%) used AR. The rest (6%) utilised regular monitors with audio.

According to the review, see Table 2.2, the use of HMDs offered participants more freedom of movement and a higher-quality visual experience (larger field of view, head movements with three degrees of freedom, higher pixel density, adjustment for vision limitations, etc.), and a more natural interaction with varying degrees of augmentation (i.e., the amount of virtual imagery superimposed on real imagery). However, the static monitor

Table 2.2: A few of the advantages and disadvantages of static, flat monitors vs. HMDs, in the context of the review [2].

Technology	Advantages	Disadvantages
Fix monitors	Better precision Easier registration More control over experiments	Expensive Limited motion range Not immersive
HMDs	Relatively affordable Enables AR 3 degrees of freedom	Poor research coverage Calibration required

experience provided better control over environmental variables and eased XR system development.

Out of the 31 selected articles, five of them studied user performance in Ventriculostomy placement, an operation closely related to EVD. Four of the five allowed the participant to simulate the operation experience. During these tests, the researchers took objective measurements of the user performance, such as accuracy measures and the completion time of the procedure. The fifth allowed the participants to give a subjective metric by partaking in a survey.

Ultimately, the paper suggests that exploring educational applications on commercial HMDs could enhance the existing literature in the medical field. Particularly since cranial neurosurgery remains unexplored in XR education [2]. Where expensive equipment and devices such as the NeuroVR [5] and ImmersiveVR [6] already have a foundation, research with HMDs in cranial neurosurgery would be beneficial [2].

2.7.2 360° surgery video

Harrington et al. presented a modified depiction of visualising an operating room. A team at the College of Surgeons in Ireland recorded a surgery and then showed it to (n=40) participants [16]. The recording was done by using six GoPros *, rigged to the sides of a plastic cube and then suspended over the operating table during an elective Laparoscopic Cholecystectomy operation.

After the operation, the six recordings from the GoPros were stitched together into a 360° video and edited to include operative educational imaging, patient radiology and an artificial 2D screen alternating between the 2D camera feed and the laparoscopic camera. The video was then viewed through a

*Action Camera, GoPro is a registered trademark of GoPro, Inc. [43]

Samsung GearVR* virtual-reality headsets as a 360° environment and through a 2D experience on a 75-inch television. When the video was shown to the participants, half of them used the VR headset first. The other half was shown the 2D experience first. Whilst the 360° video allowed for camera control for the user, the 2D camera feed provided adjustment of aperture for intense theatre lighting. During 4 intervals, the user was probed with a visual stimulus and were asked for the subjective engagement at each time stamp.

There was no discussion of spatial or temporal controls for the videos, in which the user could control the playback. The video played all the way through with questions for the user to complete throughout, with questions regarded task-unrelated thoughts. However, the videos included sound which in the 360° environment allowed for 3D surround sound, allowing the user to get directional audio from the surgeons. Meanwhile, in the 2D viewing it only allowed for stereo sound.

At the end of the report the authors discussed the limitations of the equipment and technology. For instance the non-consumer friendly use of six GoPros and editing using the Adobe video software. Two expensive production equipments. Furthermore, the time and energy invested into making and editing the 3D videos adds to the limitations. However, in this study the VR headset utilised was the Samsung GearVR, which only requires a smartphone. The headset only provides lenses and a way for the phone to be mounted to the head.

The results of the experiment showed that educational augmented 360° operative videos visualised through the tool of VR yielded improved attentiveness and engagement amongst participants.

The video shown during the experiment was not pure 3D as each direction was only recorded with one camera lens, meaning that the video lacked 3D depth in all directions. Using a GoPro Odyssey† on a camera rig would allow for true 3D 360° video, providing each eye with a different perspective. However, these systems are significantly more expensive than the already expensive GoPro setup used in the experiment.

2.7.3 VR Visualization Tool for Neuron Tracing

In a study conducted by Usher et al. a toolkit to trace visualised neurons in VR was developed [44]. Displaying a model of a neural network in virtual space, participants were able to sketch 3D lines through each neuron using

*Samsung GearVR is a registered trademark of Samsung Electronics Co., Ltd
†GoPro Odyssey is a registered trademark of GoPro, Inc.

VR controllers. One of the datasets used was a large microscopy scan, which had to be scaled down in order to perform well on their GPU. This helped to provide a VR rendering performance for a high quality experience and to prevent software-induced motion sickness. The tools used for exploration of the 3D data were designed in collaboration with expert neuroanatomists. In the report, two fundamental features were of interest, navigation and tracing: The interactions for navigating the data, as well as tracing the neurons were each mapped to separate controllers. When holding down one of the VR controller's trigger button, it enables the user to pan over the data, grabbing the current region in space and translating it through the volume. Meanwhile, the other controller was used for drawing when its trigger button was held in [44].

A noteworthy item of the design was the decision to remove part of the virtual model of one controller. This decision was made to avoid occlusion between the marker, visualisation and controller.

The authors also implemented a system to track, record and playback the users actions. This allowed them to track the user's actions instead of displaying them on a standard 2D video recording.

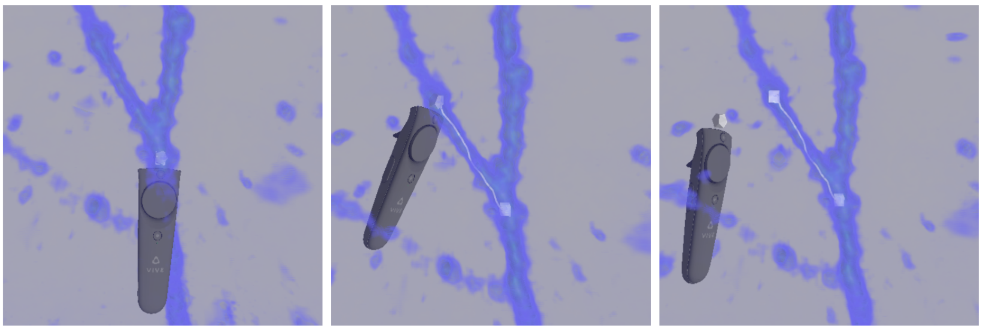

Figure 2.2: The neuron tracing process visualised from left to right. Starting with finding a neuron, the user can hold down the trigger on the VR controller and follow the neuron, tracing it, until released.

Chapter 3

Method

3.1 Research Process

The research process can be categorised into four main steps. A basic implementation of the two environments 3.1.1, its refinement and validation 3.1.2, the user study 3.1.3 and the analysis of its results 3.1.4. After the research process the integration of the content learn consisted of the implementation, pilot and refinement of the VR and Desktop interactive visualisations.

3.1.1 Previous Implementation

One of the earliest interactions of the data visualisation can be found in the course *DM2799*, *Advanced Projects in Interactive Media Technology* at KTH where Holmberg et al. mapped the recorded reflective markers to virtual catheter and brain models. Coupled with this, they also implemented a means to interact with the visualisation by being able to speed up and slow down the animation [45].

The implementation of the data from the *Advanced Projects in Interactive Media Technology* course was subsequently iterated upon in a later course, *DD2470 Advanced Topics in Visualization and Computer Graphics*. There, the controls got iterated upon to, for example, allow for a more convenient interaction with the time controls. Previously, the user could only slow down and speed up the animation, which was inifinitly looping. In order to slow down, the user had to slow down the animation to a stop. In the new implementation, the time controls were converted into a more generally adopted layout, allowing for play, pause, step forward and backward. A short user study of eight participants took place. The task of locating the site at which the catheter made contact with the ventricle was the same as this Master thesis. However, experiment was reduction in scope to fit a smaller time frame Furthermore, the literature research study conducted prior to the study was minimal, only accounting for a few papers.

3.1.2 Refinement & Pilot Phase

The implementation carried out in the present thesis project builds upon the previous work described above, from the MoCap data to the basic implementation from *DM2799* and its upgraded version from *DD2470*. The user study performed in the Advanced Topics course included verbal feedback on the implementation in terms of controls schemes and environment interaction. This feedback was taken into consideration in order to improve upon the visualisation and interaction paradigms. Likewise, the literature study presented here has provided further insight into aspects to change and iterate upon. This phase also involved implementing the items which were excluded from the previous version due to time constraints, such as measuring control interaction and controller visual guide.

One of the more substantial implementations that was later disregarded was the ability to pick up, rotate and scale the animation in VR. Inspired by the neuron pannnig conducted by Usher et al. [10] and SteamVR Home*, the option could have allowed for a more intuitive and efficient navigation in VR. Nonetheless, the implementation had no fair and time efficient counterpart in the Desktop environment and was thus later abandoned. Usher et al. also inspired the decision to further lean into the controller scheme of having two VR controllers for two different purposes, time and space.

After the implementation was completed, tests were performed by the author to evaluate all the implementations, inspecting if all data variables were recorded. A short pilot study was also carried out, using a participant. All aspects of the interactive visualisation were validated. A full review of the trials was conducted to see if there were any unsolved bugs, if all data was collected and stored properly. All keyboard and mouse commands and VR controller abilities were pressed a predefined amount of times or held in for a set amount of time, which were then compared to the data record. The pilot study was also important to check whether a participant would be able to complete the tasks in an allocated time interval. The results from the

*The standard environment shown before entering a game or application using SteamVR [46].

pilot were then evaluated and refined to make the user study go as smoothly as possible. Here, further feedback was taken into consideration and some additional implementation changes were made, such as the ability to increase the speed in Desktop by holding down the CTRL key.

3.1.3 User Study Phase

Following the final refinements from the pilot and implementation, a user study was conducted. At this point, the participants were invited to partake in performing the task defined in the research question; To locate the first contact the catheter makes with the ventricle in space and time. Meanwhile, time taken to perform the task, the precision, accuracy and number of attempts made were recorded as objective variables. The participants were also asked to answer a questionnaire throughout the experiment to collect data on their subjective experience.

3.1.4 Analysis Phase

After the user study, the data collected was analysed in order to answer the research question and evaluate the two alternative systems on a qualitative and quantitative level. The analysis involved qualitative assessments based on measured values, including factors such as time to completion, position accuracy, and precision. It also took quantitative judgement into account in terms of the results from the questionnaire and any additional feedback. Furthermore, the quantitative and qualitative findings were cross examined, compared and correlated to each other.

Besides the 20 participants, two additional participants were excluded from the analysis since they did not fully complete all trials due to personal time constraints.

3.2 Data collection

For each of the two environments and for each trial (i.e. when a new animation was shown), time to complete the task, space-time accuracy, space-time precision and number of steps taken were recorded in a text file. The data was logged every time the user proceeded to the next trial by pressing the *ENTER* button (*Enter* key on keyboard & *Grab* trigger on VR controller).

An important design consideration was to respect the privacy of the participant, to allow user anonymity, where not enough personal information was collected to reverse engineer the user's identity. It was necessary to keep their identities hidden but also to respect their time taken to participate, whilst collecting as much data as possible. More prominent in the user surveys, compared to the VR testing, the anonymization measures applied allowed for valuable information gathering while at the same time avoiding particular questions which could appear offensive or demeaning, for instance.

3.3 Target Population

The objective of the thesis project, in regards to participants, was to recruit KTH students modelling as medical ones. Since the interaction visualisation is ultimately designed for medical learning purposes, it is beneficial if the participants simulate its real intention. KTH students do not have the same knowledge of the operation as educated medical students, professors or surgeons. This will indubitably lead to bias and unrepresentative results. The participants should preferably have full stereo and colour vision, with some level of 3D competence. However, participants without 3D competence is still included to see if it affects the interpretation of the visualisation and if it is reflected in the results. Previous VR experience was also beneficial, since it would ease the onboarding process, in terms of the previously discussed ”wow” factor that sometimes appear the first time interacting with a virtual world through VR. Having previous experience with using a HTC Vive would also be favourable. In such case, the participant would probably be familiar with the headset and hand controllers, or at least know where to find buttons and triggers more easily, while immersed in VR.

The study did not show a preference for any gender. Instead, it would be advantageous to have equal representation of both genders. This gave way to an analysis where gender could be seen as a possible contributing factor. Similarly, there was no targeted age group, but since the experiment was marketed to mostly KTH students, it was expected that the average age would be of young adults.

The target population was reached by posters, emails and hallway interception on KTH campus.

The sample size of the user study involved 20 participants. This means that a semi-high enough significance was met to secure some level of certainty in the analysis. Nonetheless, the ideal sample size would have been 30, given a required significance level of 0.05.

3.4 Questionnaire

A questionnaire was designed to collect the subjective quantitative data from the participants. It was designed to enable the elicitation of enough information, without causing fatigue or boredom, which are potential threats to the validity of the study. The survey also served the purpose of introducing variability into the otherwise routine experiment.The questionnaire was presented on an external laptop, with only the online questionnaire showing in fullscreen, see Figure 3.2.

Before the participant started interacting with either the VR or Desktop experience, he/she was asked to fill out the first part of the questionnaire, collecting various data from the participant, see Section 3.4.1. The data sheet detailed issues such as gender, age, colour vision, previous experience with VR and computer games. This acted as the first part of the online user study questionnaire. After the user has answered the intro sheet, the viewer either enters the Desktop environment or put on the HMD for the VR experience.

The participant would complete the questionnaire questions regarding the present environment after addressing the first environment, and subsequently, at the conclusion of the second environment. This approach would also address another threat to validity, since it ensured that participants' memories of the current experienced environment were recent, prompting relevant and contemporary questions in the study. For each environments, the questions included two immersion and engagement questions from IEQ and GEQ, ten SUS questions, six NASA-TLX workload questions, as well as one open-text feedback section, plus the five introductory questions. Leaving a total of $5 + 2 \cdot (2 + 10 + 6 + 1) = 43$ questions.

The colour blindness test was linked in the questionnaire to a third party site *colorblindnesstest.org* [47]. In the test, 24 colour plates were shown, where each plate depicted a number between 0 and 9. See Figure 3.1 for an example of a colour plate. The colouration of the numbers versus the background colour tested if the user could distinguish a number among the dotted plate. If the user did not have full colour vision the experiment concluded and the participant was free to go.

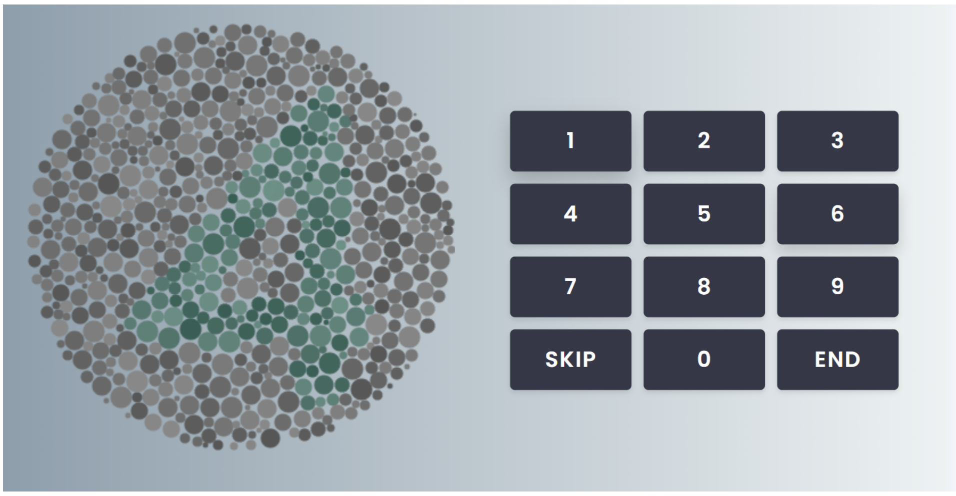

Figure 3.1: Colour blindness test plate, showing a green four on a grey amongst grey dots [47].

3.4.1 Introductory Questions

Before the two environments were tested, five introductory questions collected more personal information from participants. However, the information gathered still kept the anonymity of the person. Table 3.1 shows more details on the five introductory questions. These were meant to get an overview of the participant in order to analyse if any of the five areas could have an impact on user performance.

Table 3.1: Introduction Questions for Questionnaire.

#	Question	Answer Type
1.	Gender	Multiple Choice
2.	Age	Short Answer Text
3.	How often do you use/have you used VR?	Multiple Choice
4.	How often did/do you play video games?	Multiple Choice
5.	What is your experience with video games?	Likert scale [1-5]

The multiple choice question for gender included: *Male*, *Female*, *Prefer not to say and Other*. The multiple choice question for VR experience included: *Never used any type of VR*, *Sometimes*, *Occasionally*, *Regularly* and *Use it extensively*. The Multiple choice question regarding video game frequency included: *Never used any type of video games*, *Tried it once or twice*, *Sometimes*, *Weekly* and *Daily*. The Likert Scale for video game experience had labels *None* and *Advanced* for the extremes of the scale, 1 and 5 respectively.

3.4.2 IEQ and GEQ Questions

From the 32 IEQ questions and the 19 GEQ questions, only two questions were overall kept. Since previous research had shown that VR is more immersive and keep the user more engaged in the experience [8, 7, 13, 2], the focus of the questions was shifted more towards the SUS and NASA-TLX. However, the two questions chosen from the IEQ and GEQ covered most ground over engagement and immersion, their scores were analysed individually:

- **I1. Engagement:** To what extent did you feel you were focused on the game and it held your attention?

  Likert scale from 1 *(Not at all)* to 7 *(A lot)*.

- **I2. Immersion:** At any point did you find yourself become so involved that you were unaware you were even using controls?

  Likert scale from 1 *(Not at all)* to 7 *(Very much so)*.

3.4.3 SUS Questions

All ten SUS questions were included in the questionnaire. S1 was adopted to integrate the target audience in the medical field. The questionnaire provided information about the systems usability by calculated an accumulated score and comparing it to already established score ranges, see Figure 2.1. All questions have labels on the extremes ranging from 1 *Strongly Disagree* to 5 *Strongly Agree*.

- **S1.** I think that a doctor or medical student would like to use this system frequently.
- **S2.** I found the system unnecessarily complex.
- **S3.** I thought the system was easy to use.
- **S4.** I think that I would need the support of a technical person to be able to use this system.
- **S5.** I found the various functions in this system were well integrated.
- **S6.** I thought there was too much inconsistency in this system.

- **S7.** I would imagine that most people would learn to use this system very quickly.
- **S8.** I found the system very cumbersome to use.
- **S9.** I felt very confident using the system.
- **S10.** I needed to learn a lot of things before I could get going with this system.

3.4.4 NASA-TLX Questions

All six NASA-TLX questions were included in the questionnaire. No modification was made to the NASA-TLX questions. The questionnaire provided information about the task workload by calculated an accumulated score and comparing it to already established score ranges, see Table 2.1. All questions, besides NT4. *Own Performance*, have labels on the extremes ranging from 1 *Very Low* to 10 *Very High*. NT4, *Own Performance* is has extreme labels 1 *Perfect* and 10 *Failure*.

- **NT1. Mental Demand:** How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?
- **NT2. Physical Demand:** How much physical activity was required? Was the task easy or demanding, slack or strenuous?
- **NT3. Temporal Demand:** How much time pressure did you feel due to the pace at which the tasks or task elements occurred? Was the pace slow or rapid?
- **NT4. Own Performance:** How successful were you in performing the task? How satisfied were you with your performance?
- **NT5. Effort:** How hard did you have to work (mentally and physically) to accomplish your level of performance?
- **NT6. Frustration Level:** How irritated, stressed, and annoyed versus content, relaxed, and complacent did you feel during the task?

3.5 Experimental Design

3.5.1 Test Environment

During the experiment only one participant at a time was allowed into the test room. The test room included a desk, office chair, computer, monitor, keyboard, mouse, the VR HMD and its controllers. The computer was also connected to a HTC Vive Headset* through a four-metre displayport cable. The headset was wirelessly connected to its two associated controllers. The three were then also individually tracked by three base stations†, mounted in three of the four corners of the ceiling. When the participant was interacting with the Desktop environment, the monitor, keyboard and mouse were in use. When the HMD and VR controllers were in use, the participant did not see the monitor nor interact with the keyboard and mouse. However, the researcher could still use the monitor to observe the participant and the task performed when they were immersed in the VR experience. To reduce motion sickness, the participants were instructed to be seated on the office chair throughout the trials, to minimise injury if the participant experienced any motion sickness. The chair had wheels which allowed for generous movement throughout the designated experimental zone. The zone stretched out 4 metres from the table where the monitor was situated, with a width of 3 metres.

See Figure 3.2 for Desktop setup and Figure 3.3 for the VR setup.

The virtual environment presented in this study is low in detail, only containing the necessary components, the bare minimum for interaction. Besides the catheter, brain and ventricles, the participant can see an empty room with an surgical bed located in the centre of the room, underneath the skull. The scene originally also contained a model of a Computed Tomography (CT) scanner connected to the operation bed. This was deemed unnecessary for the experiment and was removed.

*The HTC Vive Headset is registered under HTC Corporation.

†Also known as *Lighthouses*, they track the exact location of the headset by sweeping the room using wireless pulses of infrared light.

Figure 3.2: Experimental setup for Desktop. The main monitor with keyboard is used to view the virtual environment. The Laptop on the right of the main monitor, seen in the left image, is only used to answer the questionnaire.

Figure 3.3: VR experimental setup for participants. The ventricle is placed in the middle of the room, three metres away from the desk to allow the freedom of movement, without the user feeling afraid to bump into furniture.

Figure 3.4: Introductory scene before the start of the experiment, featuring the operating room, the marker and ice cones. The controller guide for Desktop is located in the bottom right laid on-top of the environment. In the bottom left the red and white slider can be observed.

3.5.2 Test Layout

Upon initial entry into both environments, each user was shown the introductory scene 3.4. The participant was asked to place a red marker sphere on top of an ice cream cone. This required the user to locate the cone, reposition himself/herself and interact with the marker controls. The controls enabled picking up and placing the marker, as well as scaling it up until it resembles a normal-sized ice cream ball. When the user was satisfied with the position and scale, they would continue by pressing the **ENTER** button. This marked the end of the introduction and the start of the actual experiment.

The following scene, which was the first trial of the experiment presented the visualisation of the skull and ventricles, as depicted in Figure 1.1. The user was tasked to locate the point in time and space where the catheter touched the ventricle. The position was marked by the same type of marker used in the introductory scene. When the user was satisfied with their selection of time frame and location in space, the next scene presented the user with another animation of the EVD, which then followed by four more, thus totalling in six animations for that environment. The experiment was then repeated for the other environment, with the same introductory scene, followed by the same six animations. The six animations were generated from three MoCap recordings.

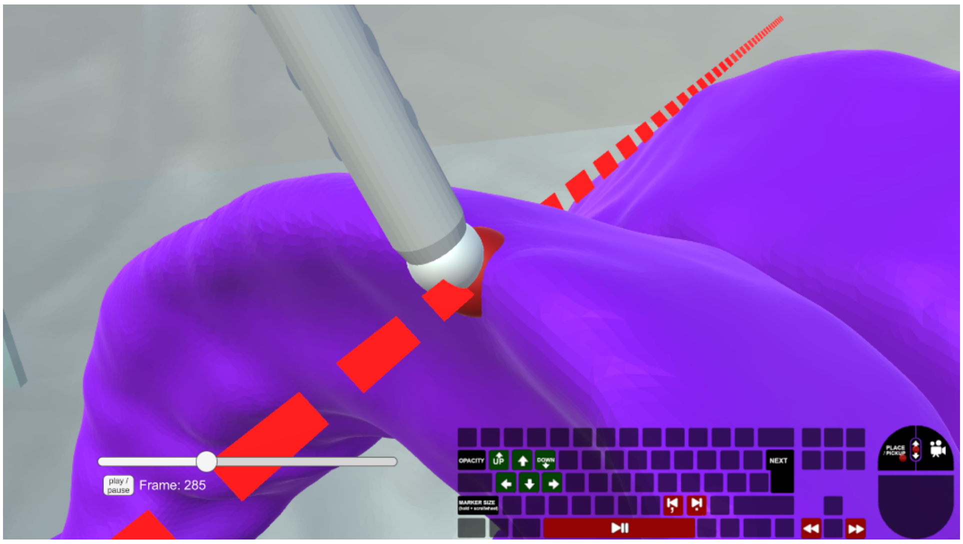

Figure 3.5: Trial scene, marker placed halfway inside ventricle, with the catheter showing in front of it.

The three recordings were duplicated and shifted by 12 frames, in order to not have the same frame of impact, which could cause unwanted memorisation.

The users had access to a guiding laser in every other trial, which displayed a red dashed line extending out from the observer. In Desktop, the guidance laser pointed out from the viewers perspective straight forward, starting from a slight offset from the viewers camera. The laser was included in every other trial to give a good comparison for each of the three recordings.

When participants were done with one of the two environments, they were asked to answer questions regarding the interaction specific to the environment just used. After that, users were tasked with repeating the introduction environment and the following six recording tasks but for the other, untested, environment. In other words, if a user started with the Desktop experience, they proceeded to answer the questionnaire for the desktop and then enter VR. Which was followed by the VR questionnaire. The order of desktop and VR was swapped after each participant, to guarantee a lack of bias and threat to validity. Upon completion of both environments and the corresponding questionnaires, the participant concluded their involvement in the experiment and was subsequently released.

3.5.3 Desktop Controller Layout

In the Desktop setting, the user had access to the keyboard and mouse controls. The layout of the controls can be seen in Figure 3.6.

Spatial navigation: WASD-keys controlled forward, left, right, and backward camera movement, as a standard in gaming and also used in the Unity editor[48]. CTRL-key was for increasing the magnitude of the camera movements from the WASD-keys. This was a late addition to the control scheme, a better option given by feedback would have been to have the ability to increase speed on shift, a standard in gaming. Q and E are for moving up and down in the room, also from the Unity editor[48]. The camera orientation was controlled by holding in the right mouse-button and moving the mouse, common in many 3D interactions. Alternatively, the user could also achieve the same effect by clicking on the left mouse-button, utilising the Rotate Around the Marker (ROAM) ability. This makes the camera fixate on the marker, if placed down, while keeping the same relative distance. This ability was inspired by video games, where the player can rotate around the in-game character, such as in the Legend of Zelda[49]. By moving the mouse, the participant could rotate around the marker, obtaining a better understanding of its location in the virtual space.

Temporal navigation: The temporal navigation is mainly inspired by the media industry, from sites such as Youtube [50], as well as the TV remote control. The user could play and pause the animation by pressing the *SPACE* button (*PlayPause*). They could fast forward and backward by 5 frames by pressing the LEFT and RIGHT arrow-keys (*SpeedFrame*). By holding those same keys down, the user could also fast forward and backward continuously, adding one frame per 0.5 seconds (*WindFrame*). The user also had the option to step only one frame back or forward by using comma ”,” and period ”.” (*StepFrame*). The user also had access to the slider, which can be seen in Figure 3.4, overlaid on the view of the operating room. It could be selected and dragged left and right with the mouse cursor (*SliderSelect*).

Marker: In order for the participant to indicate its chosen position in space, where the catheter made first impact with the ventricle, a spherical red marker is used. The user can pick up and place the marker with the middle-mouse button (*PickPlaced*). By scrolling the scroll-wheel on the mouse, the user would push the marker away and towards the camera respectively, in a straight line towards or into the monitor screen. If the user instead held down the left SHIFT-key and used the scroll-wheel, the marker would instead increase or decrease in size.

Additional controls: When the user pressed the (*ENTER*)-key their trial would stop and the completion time would be recorded; the scene would then reload with the next animation. This could only take effect if, and only if, the marker was placed down and the animation was paused by the user. The restriction would mitigate people pressing it and moving on to the next trial by mistake.

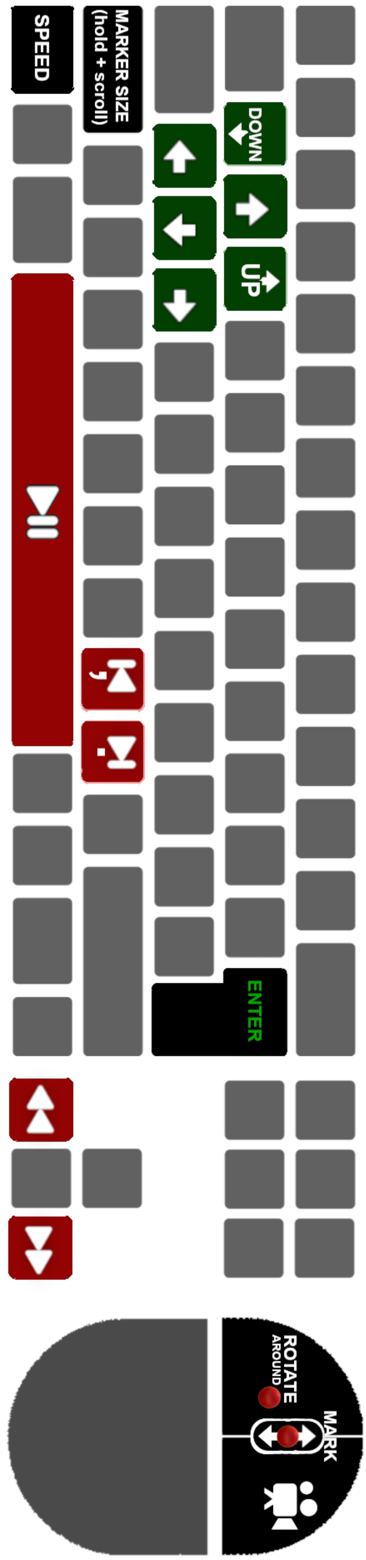

Figure 3.6: Controls for Desktop. Keys with green background denote the spatial navigation controls, those with red background denote temporal navigation controls, those with black background are additional controls.

3.5.4 VR Controllers Layout

In the VR setting, the user had access to the VR HMD and VR controllers as instruments to interact with the environment. The layout of the controls can be seen in Figure 3.7.

Spatial navigation: The user moves through the VR environment by moving their head and body. The user was sitting on a chair throughout the VR part of the experiment in order to mitigate potential motion-sickness, thus anchoring them to the chair. They were allowed to move the chair around and would be warned if the chair got too far out of bounds (this never occurred).

Temporal navigation: In the user's non-dominant hand, the time controller was placed, as depicted to the left of the top image in Figure 3.7. On the time controller, all time temporal controls to interact with the environment were located. The user could play and pause the animation by pressing the middle of its trackpad (*PlayPause*). They could fast forward and backward by 5 frames by pressing the LEFT and RIGHT sides of the trackpad (*SpeedFrame*). By holding down those same sides, the user could also fast forward and backward continuously, adding one frame per 0.5 seconds (*WindFrame*). The user also had the option to step only one frame backward or forward by using the bottom and top sides of the trackpad (*StepFrame*). Hovering above the controller in a fixed position was the time slider. It could be selected by holding down the trigger button on the backside of the time-controller (*SliderSelect*). After pressing and holding the button, the user could move the entire controller left and right to make the current frame number decrease or increase.

Marker: The marker controller was held in users dominant hand, as depicted to the right of the top image in Figure 3.7. The user could pick up and place the marker with the top part of the controllers trackpad (*PickPlaced*). This could also be done by pressing the trigger button on the backside of the marker controller. The marker would get picked up and fixed at the current distance and angle from the marker controller. If the marker was too far away (farther than 0.4 metres) the user would be warned by a red text hovering above the marker controller. If the user pressed the left and right hand side of the trackpad, the marker would decrease or increase in size.

Additional controls: When the user pressed the ENTER button, located on the side of the marker controller, the trial would stop and the completion time would be recorded; the scene would then reload with the next animation. This could only take effect if, and only if, the marker was placed down and the animation was paused by the user. The restriction would mitigate people pressing it and moving on to the next trial by mistake.

Figure 3.7: Controls for VR. The left controller is the time controller, featuring the slider and the playback controls on its trackpad. Above the slider the current frame is presented as a number, currently denoted FrameNr. The right controller is the space controller, featuring the marker controls and the *ENTER*-key, seen in the second image in green. The left controller in the second image shows the slider button.

3.5.5 Hardware/software used

- 1920 x 1080 px 24” 247ELH LED Philips monitor
- Dell OCJ339 Optical Mouse
- Dell L30U Keyboard
- Desk
- Office chair
- HTC Vive Pro 1 VR headset
- 2 HTC Controllers (2018)
- 3 Base stations
- Windows 11
- Unity - 2021.3.9f1
- SteamVR - 1.25.7

3.5.6 Headset Specs

- Screen: Dual AMOLED 3.5” diagonal
- Resolution: 1440 x 1600 px per eye (2880 x 1600 px combined)
- Refresh rate: 90 Hz
- Field of view: 110 degrees horizontal
- Connections: Bluetooth 5.0, USB-C port for peripherals
- Sensors: SteamVR Tracking, G-sensor, gyroscope, proximity, Eye Comfort Setting (IPD)
- Ergonomics: Eye relief with lens distance adjustment
- Adjustable Eye Comfort Setting (IPD)
- Adjustable headphones
- Adjustable headstrap

3.6 Assessing reliability and validity of the data collected

3.6.1 Data validity

The results that follow the method were valid given that the cataloguing of the evaluation metrics were done correctly. Which relied on software, hardware and social aspects.

From a software perspective, it was important that all the metrics were logged correctly. Meaning that time taken or the chosen frame of the animation were recorded properly. The same approach was given for storing position and scale of the marker. It was also important that the correct task data and participant was connected to the correct questionnaire answers. To make sure that the data collected was the intended one, several tests were conducted prior to the user study. During the pilot, all keyboard, mouse commands, and VR controller functionalities were systematically executed either through a predetermined frequency of key presses or by maintaining specific durations of activation. These actions were subsequently compared against the recorded data. The task data also was given a number which served as a unique identifier, allowing for seamless cross-referencing between the task data and respective questionnaire entries, ensuring accurate associations and streamlined analysis, whilst still keeping the participant anonymous.

The pilot study was also important to check whether the participant would be able to complete the tasks in an allocated time interval. The results from the pilot were then evaluated and refined to make the user study go as smoothly as possible. Here, further feedback was taken into consideration and some additional implementation changes were made, such as the ability to increase the movement speed of the camera in Desktop by holding down the CTRL key.

From a hardware perspective, the communication of the VR tracking was important for the validity of the results. If the controller tracking deviated or diverged from its real position, the participant has to counteract the offset. Similarly, if the headset view was distorted, it could cause discomfort and the participant to prematurely exit the environment. It would also shift their focus away from the task of marker placement. Furthermore, it was important that the participant was feeling comfortable wearing the HMD. Informing the user to readjust the lens distance and head straps was beneficial.

Regarding the social aspects, the validity of the method and results relied on trusting the participants. In other words, that the participants did not advance through the trials too hastily, nor extend past the dedicated time. The participants placed the marker with their own intuition and understanding of the situation.

3.6.2 Reliability of data

In order to argue for the reliability of the data, one has to consider the participants taking part in the experiment. Most were KTH students modelling as medical students, which meant there was a difference in knowledge and experience. However, the participants were not expected to have any prior knowledge of an EVD. Regardless of the target audience, there was no intent to allow only experienced VR users to participate. Instead, VR, gaming and 3D competence was recorded in the questionnaire. Before the start of the experiment, the participants were given a written and verbal description of the procedure, where printed A4 pictures of Figure 1.1 and Figure 3.4 were shown.

In order to ensure a consistent basis for comparison, both systems were intentionally controlled to not include extra features that could only be applied or be too beneficial for one of the systems. Features were developed, such as picking up the entire 3D animation in VR, including the skull, ventricle and catheter, using a VR controller. The participant would then be able to rotate and scale it up using the second controller. Nonetheless, this manipulation was proven extremely more easily done in VR than Desktop which caused the feature to be discarded. The purpose of the study was to examine how people interact with VR versus Desktop in neurosurgical procedures, which required an ensured standardised comparison.

3.7 Planned Data Analysis

3.7.1 Data analysis technique & Evaluation Framework

The data yielded by the experiment was both qualitative and quantitative in nature. The means of the Likert scales gave a good understanding on how the users think the usability, immersion, engagement and cognitive load of the two systems were. Similar understanding was obtained from the verbal and written feedback comments. Additionally, qualitative data such as gender, age and previous experiences were compared with the quantitative data, where any significance between how they performed were mirrored in the qualitative inputs. The quantitative data was analysed using group-wise ANOVA. Standard deviations were given for time taken, frame and number of steps taken as well as the three coordinate directions, for both environments. The comparison between having a guiding laser present or not was also done through ANOVA, where the average of all tests for both entries were taken.

3.7.2 Software Tools

Tools used for the analysis included Google Spreadsheet, Google Forms, Microsoft Excel and IBM's SPSS [51]. The data from the questionnaire was imported directly from the Google Forms, whilst the quantitative data logged from each trial was exported from the generated text files, created directly from the Unity program, into Google Spreadsheets. Here the planned data analysis was performed. The Google Spreadsheet was exported into Microsoft Excel in order to perform further analysis together with the creation of graphs. Python was also used for guidance, calculations and the building of figures.

3.8 System Documentation

The entire Unity project is open source and available in Appendix 7. The complete questionnaire as presented to participants in the study can also be found in the Appendix 7.

Chapter 4

Results and Analysis

The purpose of this chapter is to provide the results and analysis relevant to answer the Research Question discussed in the present report. Section 4.1 presents an overview of the four measured variables (Completion Time, Frame Error, Marker Placement Error (MPE) and Marker Size). Section 4.2 describes the correlations between the previously mentioned variables. Section 4.3 shows the results of the Completion Time for the tasks carried out on Desktop and in VR. Section 4.4 details how the participant performed in regards to finding the correct frame of impact. Section 4.5 visualises the MPE in 2D and 3D graphs. Section 4.6 focuses on the laser and its impact on the task performance. Section 4.7 explores the results of the controllers impact. Lastly, Section 4.8 looks into the answers from the questionnaire, featuring the Immersion, Engagement, SUS and NASA-TLX ratings. In total 22 participants took part in the user study, 2 were excluded due to insufficient amount of trials recorded, yielding a total of 144 trials. Amongst the 20 participants, 14 were male and 6 were female.

4.1 Overview

Table 4.1: Mean, standard deviation and number of observations for the four dependent variables (Completion Time, Frame error, MPE and Size) in the two environments. All trials combined.

Type	Environment	Mean	Std. Deviation	N
Time [s]	Desktop	142.597	95.307	114
	VR	77.018	50.202	114
Frame Error	Desktop	2.570	6.299	114
	VR	6.500	14.772	114
MPE [mm]	Desktop	.000794	.000688	114
	VR	.001496	.000936	114
Size [mm]	Desktop	.002333	.001813	114
	VR	.004009	.001736	114

In Table 4.1 an overview of the four main variables can be observed. By examining the mean duration for each type, it becomes evident that the Desktop environment yields lengthier times, yet deliver superior outcomes across the remaining evaluation criteria.

Completion Time is much shorter in VR compared to Desktop, to the point of it being almost twice as short on average. A similar nature can be recognized within its standard deviation. In other words, some participants were able to understand and complete the desktop version quickly, whilst others struggled and took more time. VR demonstrated a reduced time requirement, with comparatively narrower subject variability in the distribution. This implies that participants were on a similar understanding of and approach to VR, although some had previous experience. The average completion times across all trials are $143s$ on Desktop and $77s$ in the VR environment. Nonetheless, as illustrated in Figure 4.3, the completion time diminishes over trials for both the VR and Desktop settings, with the longest time observed in the introductory scene.

Regarding the amount of frames away from the true impact frame, called Frame Error, in the two experiences, VR had more than double the mean and standard deviation. Comparing an average of 2.57 to 6.50 frames on Desktop and in VR respectively. The reason why the standard deviation possesses a substantial magnitude in VR is due to the outliers sometimes being more than 40 frames away, as can be seen in Figure 4.4.

For the MPE, the magnitude from the true point of impact of the catheter with the ventricle in 3D space, once again Desktop is overall twice as good, resulting in $0.000794mm$ versus $0.001495mm$ in VR. Meanwhile, for the standard deviation, the Desktop's superiority over VR is less obvious. The difference is also quite difficult to observe in Figure 4.6, where the spread of placement for Desktop and VR are almost equal. However, by observing each recording individually, as depicted in Figures 4.7, 4.8, 4.9, the variance between Desktop and VR is much more noticeable.

Table 4.2: MANOVA test results for the four dependent variables (Time, Frame error, MPE and Size) depending on the Environment and if the Guidance Laser is present. All trials combined.

Source	**Dependent Variable**	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**
Environment	Time	245139.034	1	245139.034	42.252	<.001
	Frame Error	880.281	1	880.281	6.827	0.010
	MPE	2.815E-5	1	2.815E-5	41.846	<.001
	Size	.000	1	0.000	50.790	<.001
Laser	Size	49.281	1	49.281	.380	.538
	MPE	2.554E-7	1	2.554-5	.377	.540

In Table 4.2 the significance of the dependent experimental metrics can be observed. All differences observed between the dependent variables are significant. The higher completion time in VR is apparent, however the same is true for a bigger Frame Error, MPE and Size.

The significance level of having a guidance laser was also measured. There is no significant impact on the results including a laser or not.

4.2 Correlation

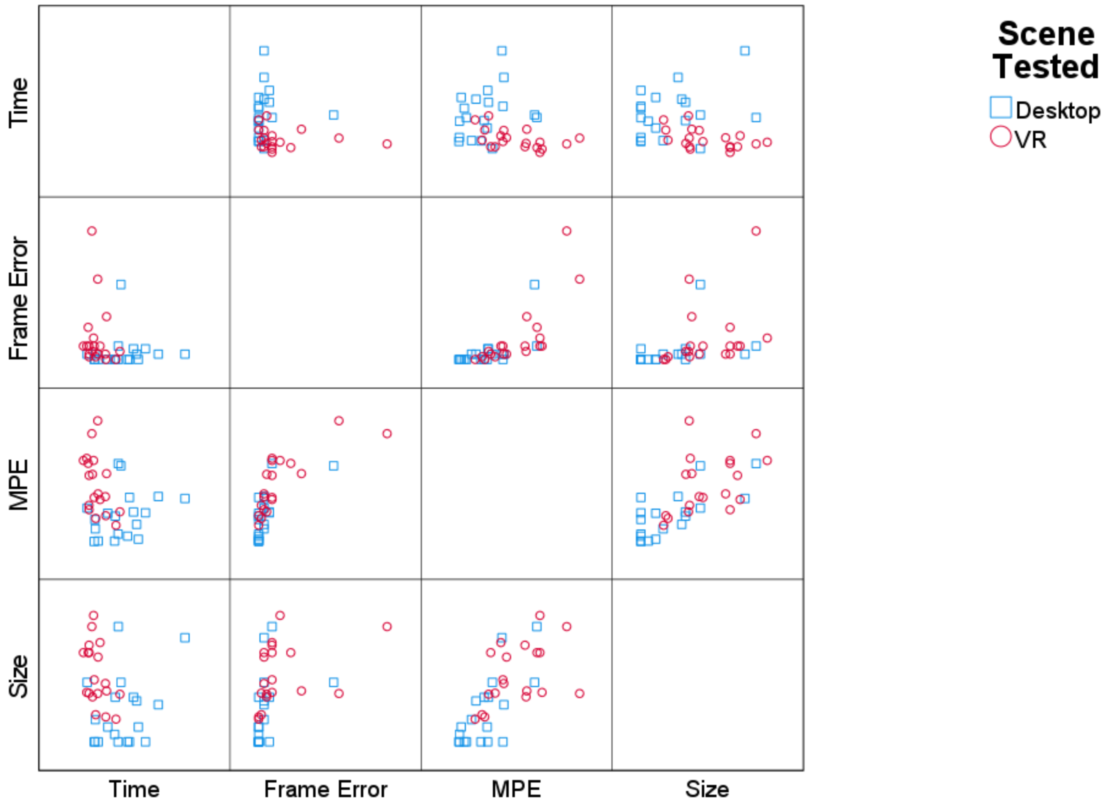

Figure 4.1: Scatter plot over correlation between Time, Frame Error, MPE and Marker Size. Box (1,1) is at the top left, box (4,1) is at the bottom left, box (4,4) is at the bottom right.

Table 4.3: Correlation Significance between Time, Frame Error, MPE and Marker Size. Significant values marked with green.

		Time		Frame		MPE		Marker Size	
		VR	DESK	VR	DESK	VR	DESK	VR	DESK
Time	Pearson			0.158	-0.061	0.172	-0.013	.272	-0.048
	p-value			0.094	0.520	0.067	0.893	0.003	0.612
Frame Error	Pearson	0.158	-0.061			.342	.311	.189	0.080
	p-value	0.094	0.520			0.000	0.001	0.044	0.397
MPE	Pearson	0.172	-0.013	.342	.311			.617	.384
	p-value	0.067	0.893	0.000	0.001			0.000	0.000
Marker Size	Pearson	.272	-0.048	.189	0.080	.617	.384		
	p-value	0.003	0.612	0.044	0.397	0.000	0.000		

Figure 4.1 shows a scatter plot matrix over the four experimental evaluation metrics and their relative correlations. It can be observed that for some of the cases, signs of correlation can be observed. Their significance level is also given in Table 4.3.

For instance, in Box (4,1) Marker Size vs. Time, it seems the marker Size decreases as more time passes. Not necessarily meaning that the accuracy increases as time passes, but the user might become more confident after trying to get it right for a longer time. In Table 4.3 the p-value for VR in Marker Size vs. Time is 0.003, while for Desktop, it is 0.612, meaning that it is only true for VR.

Longer time does not necessarily mean higher accuracy, which Frame Error vs Time Box (2,1) and MPE vs Time Box (3,1) demonstrates. They exhibit no significance in Table 4.3 and resemble two irregular clusters in Figure 4.1.

Regarding Box (3,2) MPE vs. Frame Error, MPE seems to be growing as the Frame Error increases. This seems to be the case for both VR and Desktop. During the frames which give the Frame Error, the catheter travels a distance from the correct marker placement. This means that the user may mark a position further away from the correct time and space, when the catheter's tip is at a different position. With a significance level of 0.000 and 0.001 for VR and Desktop respectively, see Table 4.3, a correlation between MPE and Frame Error can be assessed.

Just like how the MPE increased with a higher Frame Error, so does the Size in Box (4,2). However, this time it is only significant for VR, according to Table 4.3. As Frame Error decreases, the user shows more confidence and decreases its MPE as well as the size.

In MPE vs. Marker Size Box (4,3), Marker Size seems to be increasing as the MPE increases. This occurrence appears in both Desktop and VR. Such rationale aligns with the fact that accuracy can be compensated with precision. By placing the marker slightly offset, the user can still include the exact point of impact by increasing the Marker Size. Thereby decreasing his/her precision, but gaining accuracy. From Table 4.3 one can observe that these are indeed significant, with a p-value of 0.000 for both VR and Desktop.

4.3 Time to Completion

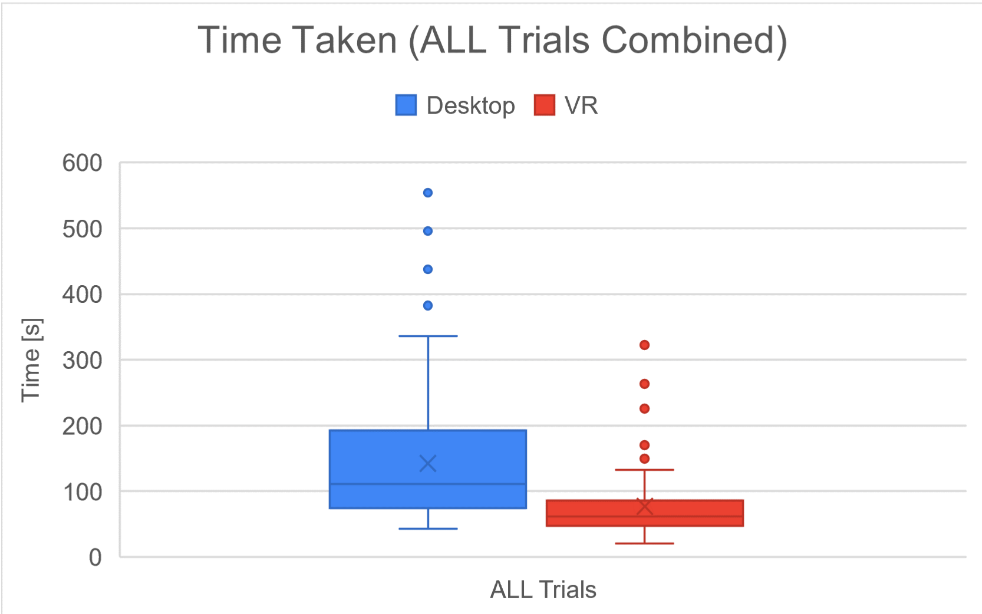

Figure 4.2: Time taken for all trials [1 - 6] combined for Desktop and VR respectively.

In Figure 4.2 a box and whiskers plot over Completion Time for all trials 1-6 can be observed. This box plot as well as all consecutive ones display the respective median as a line passing through the box, splitting the interquartile range (IQR) in two. The average is marked with a cross. The minimum is marked as the $lowerquartile - 1.5IQR$, whilst the maximum bar is the $upperquartile + 1.5IQR$. The dots represent outliers, which are values measured beyond the range defined by the minimum and maximum thresholds.

In 4.2, the outliers originate from distinct participants in Desktop and VR, signifying entirely separate subjects within each environment. Furthermore, the completion times of the introductory scene is not included in the box plot, but can instead be observed alone in Figure 4.3.

It should be noted that the completion time is almost always shorter in VR. The median, mean and the overall Interquartile range (IQR) body of VR are below the correlated one for Desktop. Moreover, the outliers from VR are in line with the upper whisker from Desktop.

In Figure 4.3, the Completion time is segregated into individual trials, this time also including the introductory scene. An overall descending trend can be detected. However, some fluctuations are present as, for example, the third trial in Desktop is on average longer than the preceding one, trial 2. However, a general trend is that the Completion time is decreasing.

It should be noted that the participants were recommended to stay in the introductory scene until they felt accustomed to the controls. The controls were novel to all users but some were more used to gaming than others, meaning that some grasp the controls quicker. This aspect is examined further in section 4.8.

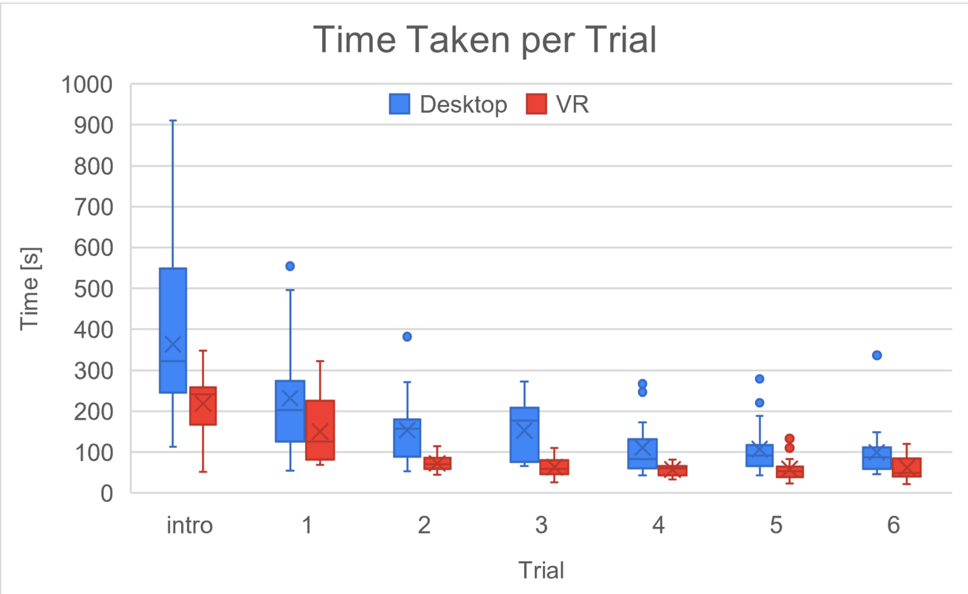

Figure 4.3: Time taken for the introductory scene and for each trial [1 - 6], in Desktop and VR respectively.

4.4 Frame Error

The animations were were on average 765 frames long, the impact between the catheter and ventricle happened during frames 224, 236, 284, 296, 310 and 322 for the six animations respectively. For each of the three recordings an additional 12 frames were added to duplicate the recordings, yet not cause the participants to answer the same frame just by observing the frame numbers. For example, the animations with impact at frame 224 and 236 were the same recording. Besides the different frames of impact, the main difference was where on the ventricle the catheter hit. During the first recording, the catheter hit the ventricle on the side. In the second recording, the catheter enter the ventricle on the top. During the third recording to this was closer to the front of the ventricle.

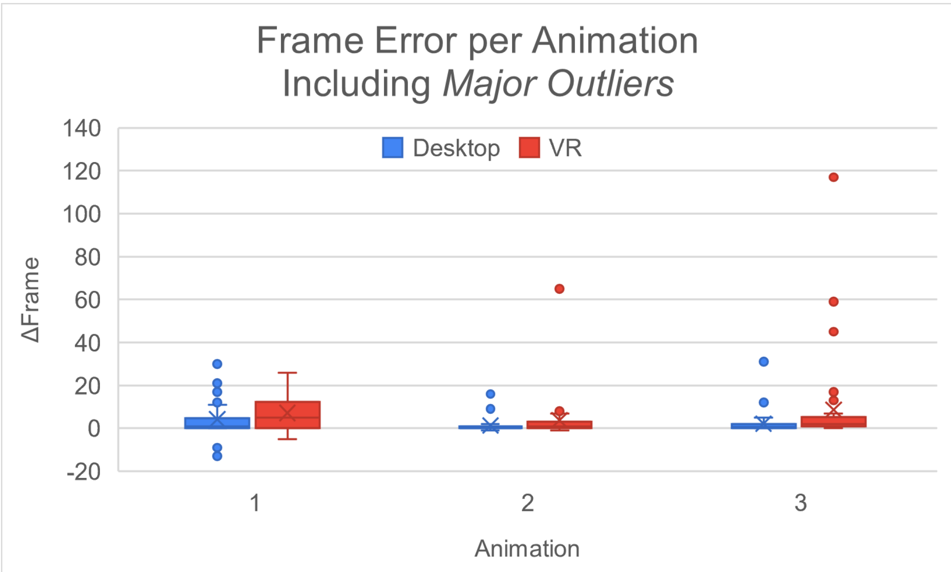

Figure 4.4: Number of frames off from correct frame of catheter impact for Desktop and VR. Including Outliers outside 1.5IQR.

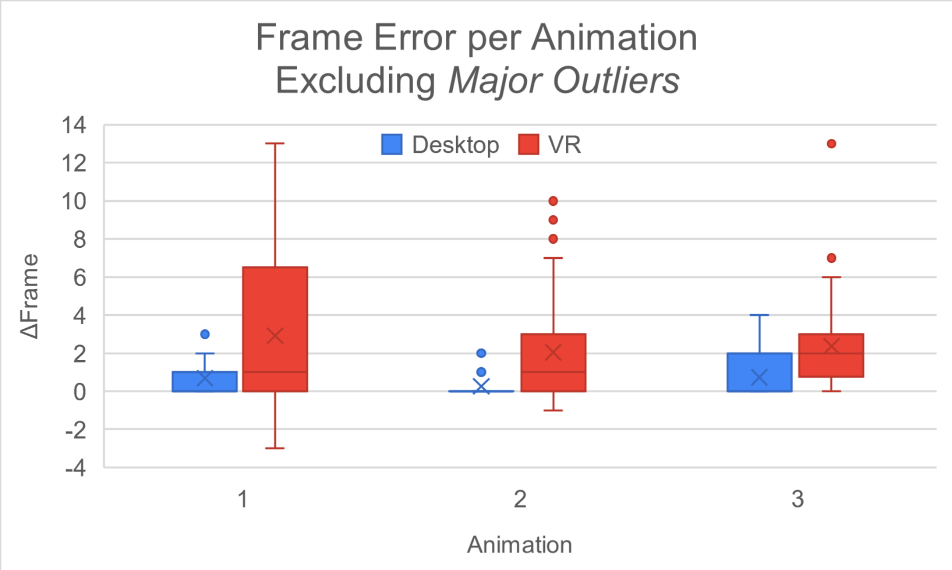

Figure 4.5: Number of frames off from correct frame of catheter impact for Desktop and VR. Excluding Outliers outside 1.5IQR of Figure 4.4.

In Figures 4.4 and Figure 4.5 the Frame Error is visualised in box and whiskers plots. The Frame Error plot is subdivided into each animation for VR and Desktop, thus combining all trials for each. Figure 4.4 shows all measurements, while in Figure 4.5 the major outliers have been omitted. The outliers are measurements outside ± 1.5 of the interquartile range. Through the omission of significant outliers, a more precise analysis of the Frame Error can be undertaken, fostering a heightened level of clarity.

Overall, the second animation was the most straightforward, exhibiting the lowest median, mean and the interquartile covering zero. Within the VR environment, the third animation exhibits heightened precision albeit diminished accuracy when juxtaposed against the second animation. Furthermore, the third animation stands out as the most intricate, marked by an exceptional abundance of outliers, with one trial exceeding 110 frames away from the ground truth.

The first animation displays the most number of negative Frame Errors, shown as outliers in Desktop and part of the minimum quartile bar in VR. A negative Frame Error implies that the participant selected a frame before the true impact of the catheter with the ventricle. A possible explanation for such behaviour could be that the catheter is moving slower between frames just before its impact. Thus, causing the subject to think that it has contact, when in reality there is still some distance left to the ventricle. Another reason could be because the view of the impact is not easily accessible and requires a lot of movement and readjustment of the camera.

4.5 Marker Placement Error (MPE)

The following section includes the results from examining the MPEs. The first Figure 4.6 includes all three animations, and the subsequent, Figures 4.7, 4.8, 4.9, show each recording by themselves. The MPE are visualised through 3D plots where the centre of each marker placed by the participants are noted with BLUE for Desktop and RED for VR. Below each 3D visualisation plot, three 2D renders can be seen. The renders display the 3D cube projected onto three different planes (XY-axis, XZ-axis and YZ axis). To enhance clarity, each coordinate axis is represented by a coloured, dashed line. RED line for the X-axis, GREEN line denotes the Y-axis and LIGHT BLUE for the Z-axis. These axes are depicted in both the 3D and 2D plots. In order to illustrate the range of the MPE in the two environments, convex hulls* are also included. For instance, see the 2D plots of Figure 4.7.

In all nine 2D plots, illustrated in Figures 4.7, 4.8, 4.9, the convex hull is considerably larger for VR than its Desktop equivalent. Sometimes the VR hull fully encloses the Desktops one. In other words, the precision in marker placement were found to be better in Desktop than in VR.

*The smallest convex polygon, that encloses all of the points in the set

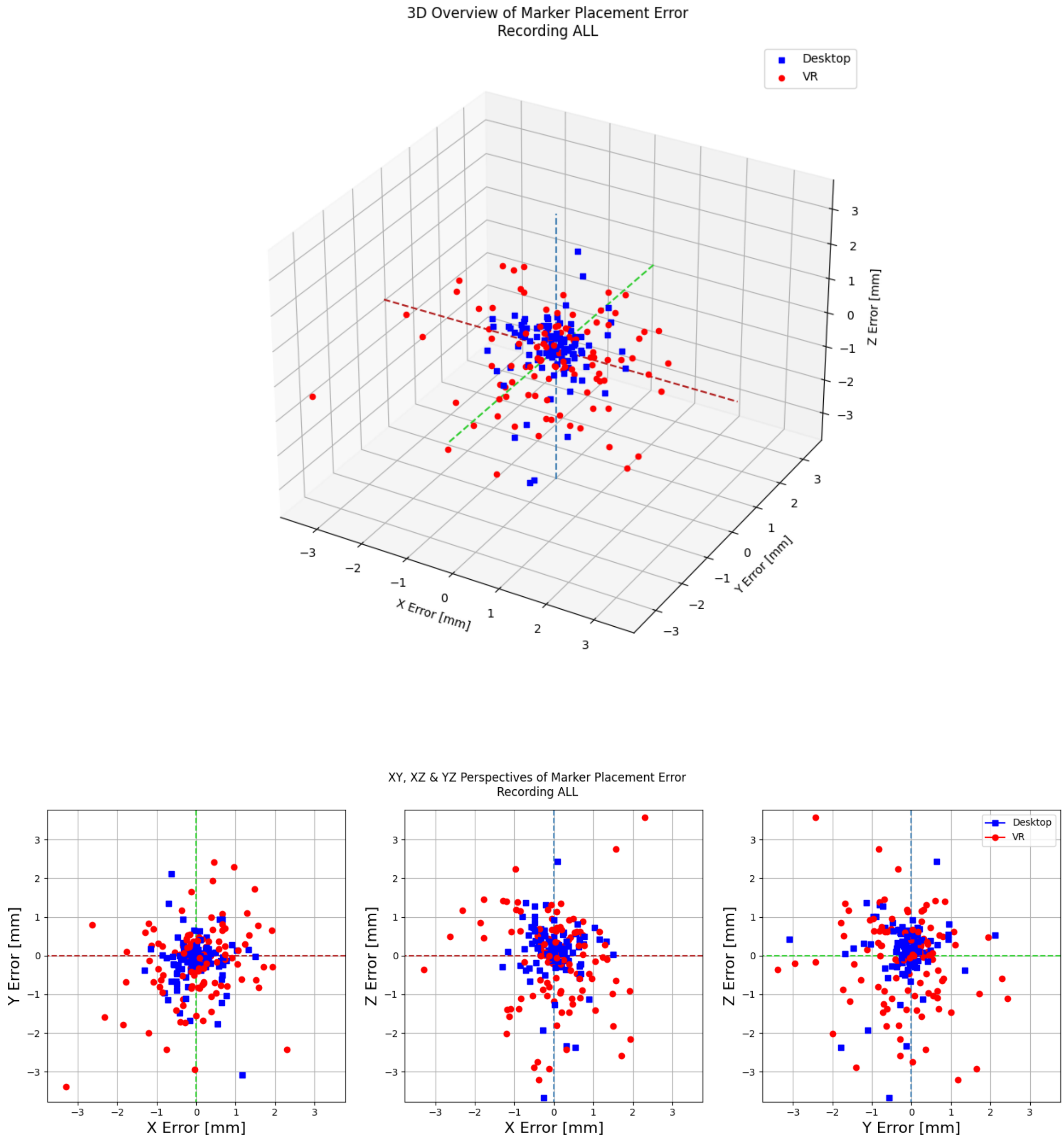

Figure 4.6: Visualisation of MPE in 3D (TOP) and 2D subplots (BOTTOM) for all recordings [1,2,3] combined. 2D plot features orthogonal viewing perspective from XY, XZ and YZ directions.

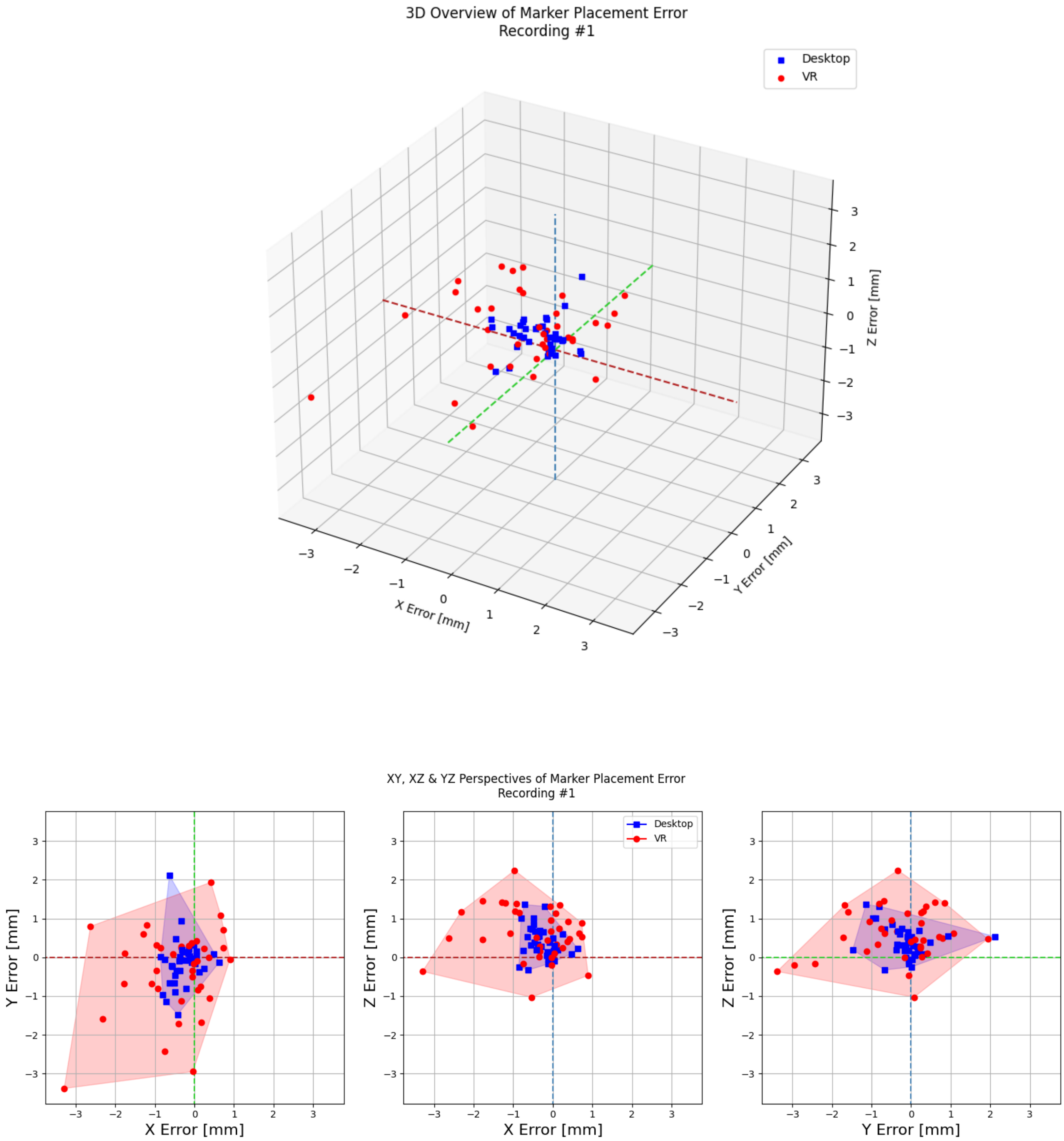

Figure 4.7: Visualisation of MPE in 3D (TOP) and 2D subplots (BOTTOM) for **Recording #1** combined. 2D plot features orthogonal viewing perspective from XY, XZ and YZ directions including the convex hull for Desktop and VR MPEs.

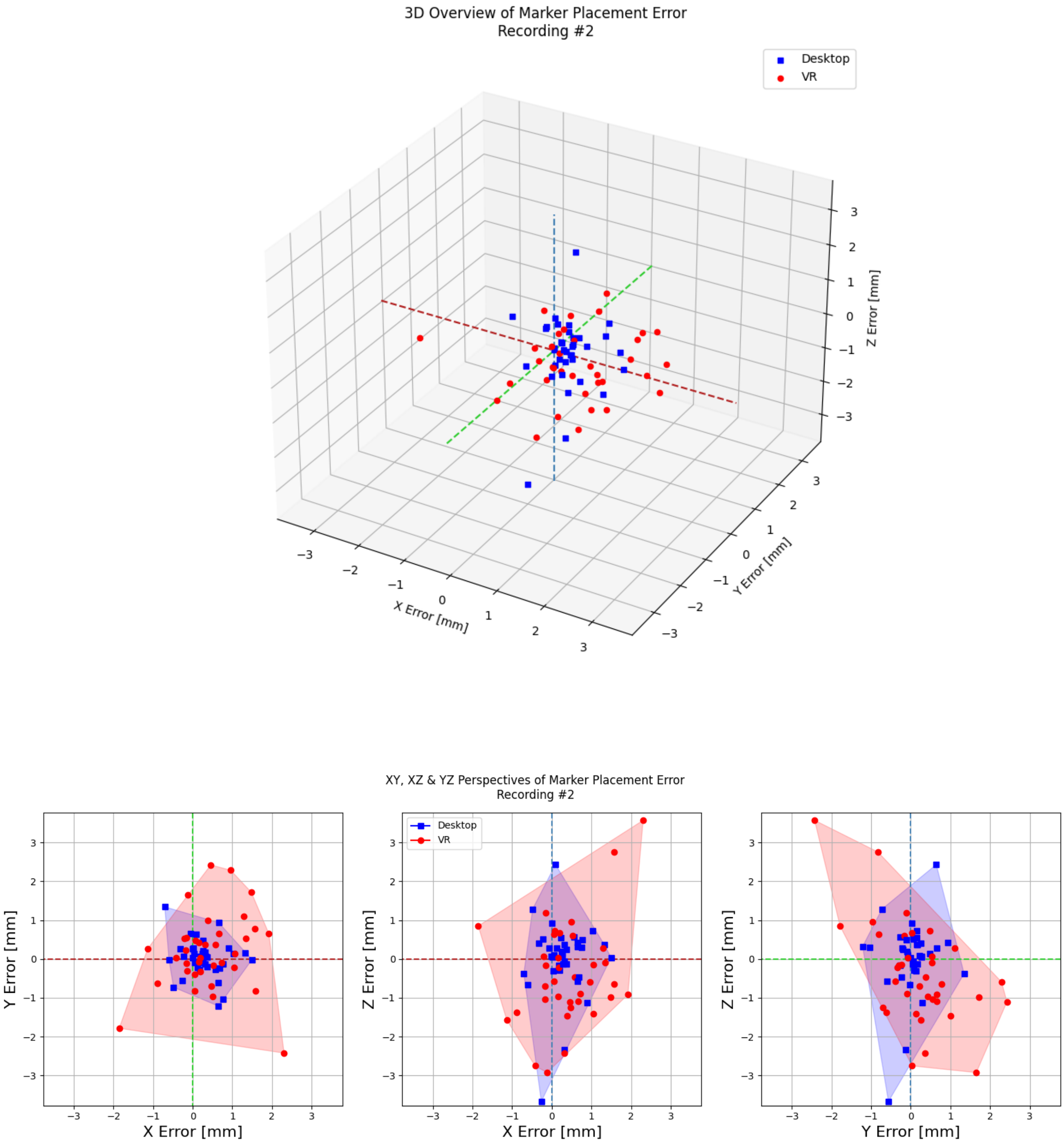

Figure 4.8: Visualisation of MPE in 3D (TOP) and 2D subplots (BOTTOM) for **Recording #2** combined. 2D plot features orthogonal viewing perspective from XY, XZ and YZ directions including the convex hull for Desktop and VR MPEs.

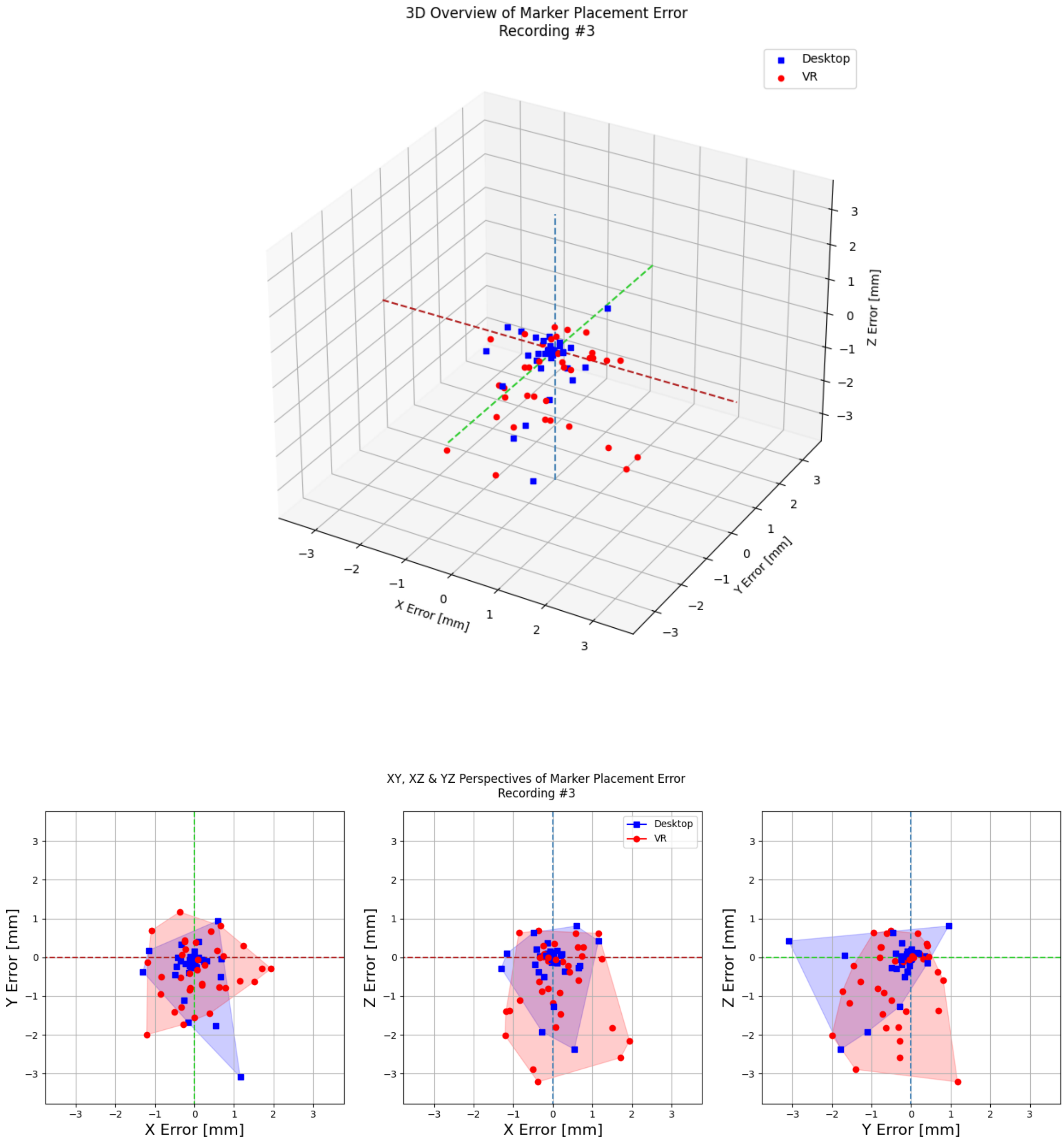

Figure 4.9: Visualisation of MPE in 3D (TOP) and 2D subplots (BOTTOM) for **Recording #3** combined. 2D plot features orthogonal viewing perspective from XY, XZ and YZ directions including the convex hull for Desktop and VR MPEs.

4.6 Guidance Laser

The guidance laser did not have any significant impact on the results, as seen in Table 4.2. Figure 4.10 and Figure 4.11 show visualisations of the laser's impact in Box and Whiskers plots. Comparing Desktop with VR, the participants displayed no increased level of accuracy or precision having access to the guidance laser.

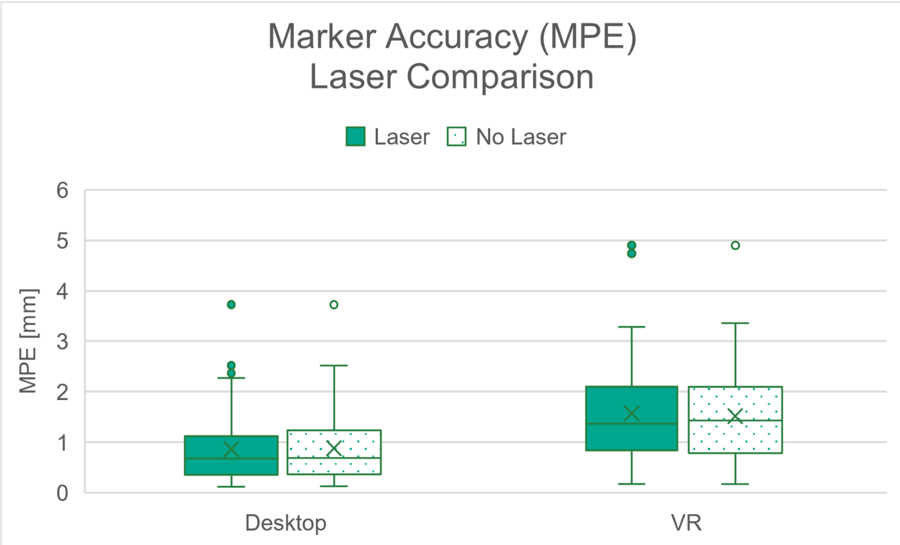

Figure 4.10: Impact of laser (full green) and without (dotted green) on MPE for Desktop and VR respectively. Median marked with cross, mean marked with horizontal line through the boxes.

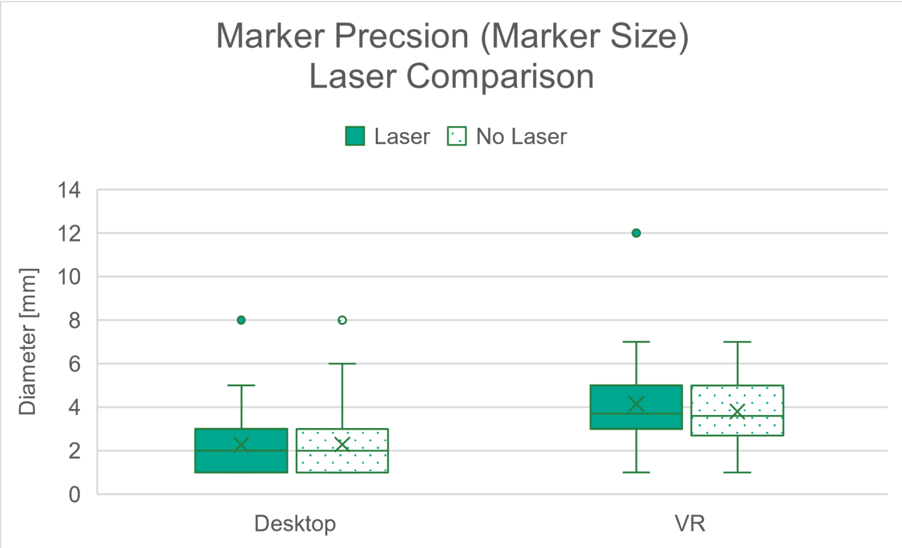

Figure 4.11: Impact of laser (full green) and without (dotted green) on Marker Size for Desktop and VR respectively. Median marked with cross, mean marked with horizontal line through the boxes.

4.7 Recorded Human-Computer Interaction

A record was maintained of the participant's level of interaction with the various controls during the task execution. Whenever a button mapped to one of the controls was pressed in or held down, its related index was incremented. In Table 4.4 the average, median and standard deviation of the different control frequencies are showcased.

Table 4.4: Mean, Median and Std. of the controls (PickPlace, PlayPause, RotateCAM, ROAM, SpeedFrame, StepFrame, Windframe and SliderSelect) for Desktop and VR.

	Environ.	*Pick-Placed*	*Play-Pause*	Rotate-CAM	ROAM	*Speed-Frame*	*Step-Frame*	*Wind-Frame*	Slider-Select
Mean	Desktop	7.96	3.18	5418	1454	11.58	15.89	172.40	123.82
	VR	6.35	1.69			13.89	10.60	133.92	61.91
Median	Desktop	5.00	1.00	3305	697	6.00	9.00	0.00	0.00
	VR	5.00	1.00			12.00	6.00	121.50	0.00
Std. dev.	Desktop	7.28	4.00	6082	1998	15.98	19.30	554.30	316.92
	VR	6.41	2.26			13.19	13.79	140.88	138.24

PickPlaced denotes the frequency of the marker placement and retrieval by the user. A wide spread of placement can be derived from the standard deviation in Desktop. Furthermore, a higher number of markers were placed in Desktop with an average of almost eight times per trial, whereas the mean was closer to six in the VR setting. Nevertheless, the median remained consistent in both scenarios with the marker being picked and placed a total of five times. Additionally, it is important to remember that the *PickPlaced* index was invariably an odd number, as the user was required to conclude the task by positioning a marker. The marker was picked up and placed a maximum of 29 times across both settings. Meanwhile, in 35% of trials the marker was only placed once, in 31% trials the marker was placed twice, in 26% trials it was placed three times. Continuously retrieving the marker increased the average, increasing the average to approximately seven instances. However, it was most frequently placed once and increases in rarity as the placement number is incremented.

The *PlayPause* index denotes the number of times the user paused and resumed the animation. Most trials, 68 and 83 in Desktop and VR respectively, the user paused the animation and never played it again. During the first trials the participant used the play/pause function until they understood the catheter behaviour. After that point, the user paused the animation once and then proceeded to either use the *StepFrame* function using the ±5 frames controls or hold them in order to utilise the *WindFrame* functionality.

The indices for rotate camera and rotate around markers were only registered in Desktop. In VR the user is constantly moving their head, making it difficult to know when they are making minor adjustments, looking down at the controls or looking around the skull. In Desktop the participants shifted back and forth between operating the camera and navigating through recording using the temporal controls. The time the participants spent using the camera controls were added to their designated time counter.

The *SpeedFrame* index specifies the number of times the user pressed the ±5 frames controls. Further, the *StepFrame* specifies the number of times the user pressed the ±1 frame controls. The participants used the *SpeedFrame* controls an average of 11.6 times in Desktop and 13.9 times in VR. The *StepFrame* function was used an average of 15.9 times in Desktop and 10.5 times per trial in VR. In 29% of the Desktop trials, and 23% of the VR trials, the *SpeedFrame* control was not utilised. The *StepFrame* controls were only ignored in 11% and 9% of Trials for Desktop and VR respectively. In other words, in 13 trials in Desktop and 10 trials in VR, the participant did not micro-adjust frame by frame, to look for the exact frame of impact. In 5 of the 13 cases and 6 of the 10 cases, neither the *SpeedFrame* nor the *StepFrame* was utilised. Instead, the participant opted to solely use the slider.

Every frame at which a user fast forwards or fast backwards the recording the *WindFrame* index is incremented. The averages in Desktop are fairly deceiving since in 50% of the trials, the participants never utilised this feature. The shortage of its utilisation is likewise mirrored within the median. However, there were only 17 reported cases of the controllers inactivity in VR. With an average of 134, median of 121 and a standard deviation of 140, it is a fairly frequently-utilised controller scheme. It allows participants to quickly speed through the animation, identifying key moments in time.

Contrary to the step-frame controls, the slider saw little use. *SliderSelect* represents the total duration for which the slider was held in seconds. Across both environments, the medians of how many frames the slider was actively engaged with were 0. In 89 and 78 out of 114 cases for Desktop and VR respectively the slider remained inactive. Desktop had double the average amount of frames in use, yet it also possessed a higher standard deviation. Interestingly, the slider control in VR was interacted with more times than desktop.

Considering that users are used to click and drag sliders from any type of interactive video, it is intriguing that the invented slider control in VR gained some traction. One reason for the dormancy could be because the screen felt only as an instrument to interact with the camera, with the slider only being a visual indicator. Another explanation could be that the participants found the keyboard controls more intuitive than dragging the slider. Nevertheless, the slider provides greater control compared to the ± 5 frames key but falls short when it comes to stepping from frame to frame.

4.8 Questionnaire Answers

In the following section the results and analysis of the questionnaire are presented. It illustrates the significance of correlation between the participants' subjective experiences in Table 4.5. Furthermore, the results for Engagement and Immersion, SUS and NASA-TLX questionnaires are shown in box and whiskers plots, see Figures 4.12, 4.13, 4.14. The overall SUS and NASA-TLX scores for Desktop and VR are also presented in their respective sections.

Table 4.5: Tests of Between-Subjects Effects on the direct impact of VR and gaming experience on main measurements.

Previous Experience	**Dependent Variable**	**Desktop p-value**	**VR p-value**
VR Experience	Time	0.004	0.162
	Frame Error	0.691	0.721
	MPE	0.013	0.809
	Marker Size	<0.001	0.162
Gaming Experience	Time	<0.001	0.398
	Frame Error	0.093	0.026
	MPE	<0.001	0.463
	Marker Size	<0.001	<0.001

Before any tasks were performed in the two environments, participants answered some introductory questions, such as their gender, their previous experience and knowledge of VR as well as how much they were used to computer and/or console gaming. The possible answers ranged from "never used it" to "use it extensively". These answers were then converted into a range from 0 to 4, which were used in the analysis of the effect of the participants' previous experience on their performance in the tasks. In Table 4.5 the significance of the correlations are showcased.

The first obvious observation that can be made is the fact that the participants' previous VR experience had no significant influence on their performance in VR. However, it seemed to affect their proficiency in Desktop. Interestingly, the between-subject test on the users VR experience gave a positive judgement for Completion Time, MPE and the Marker Size for Desktop. The subjects with a richer VR experience displayed shorter completion times, MPE and higher precision in terms of the chosen Marker Size in the Desktop environment.

The second observation is in regards to the gaming experience. Here the previous knowledge contributed to a rather substantial increase in efficiency across VR and Desktop settings. Gaming experience yielded significant and better results in Completion Time, MPE and Marker Size for Desktop, with a p-value of <0.001 for all three respectively. In VR, the Frame Error and Marker Size were significant with a p-value of 0.026 for Frame Error and <0.001 in Marker Size. Considering the significant values observed on Desktop as a result of the VR experience, the question of why this is the case emerges. With a p-value <0.001, VR and gaming experience are correlated with a factor of 0.34. This means that the participant with VR experience also possessed a high to medium level of gaming knowledge. Which explains why the previous experience in VR reported better performance in the Desktop environment. This can be explained by the fact that entertainment is tailored for the individual and as VR has high presence in the gaming industry, having experience with a VR headset could insinuate that the person is into video games.

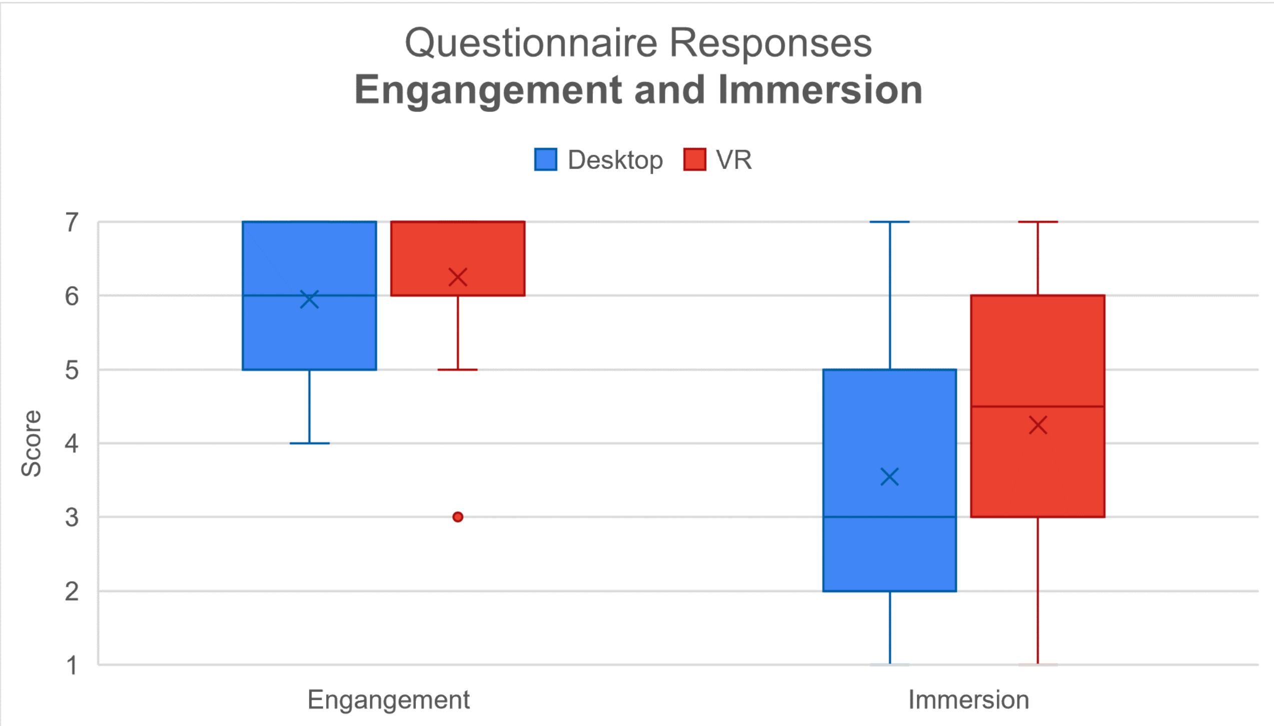

Figure 4.12: Questionnaire ratings of Engagement and Immersion for Desktop and VR respectively. Detailed descriptions of asked questions are found in Section 3.4.2.

Questionnaire ratings for engagement and immersion are showcased in Figure 4.12. The main observation from the graph is that VR received higher praise for reaching a higher engagement and immersion with the viewer. However the results are not significant, meaning that Desktop was still considered highly immersive and engaging, reaching top scores from a considerable number of participants.

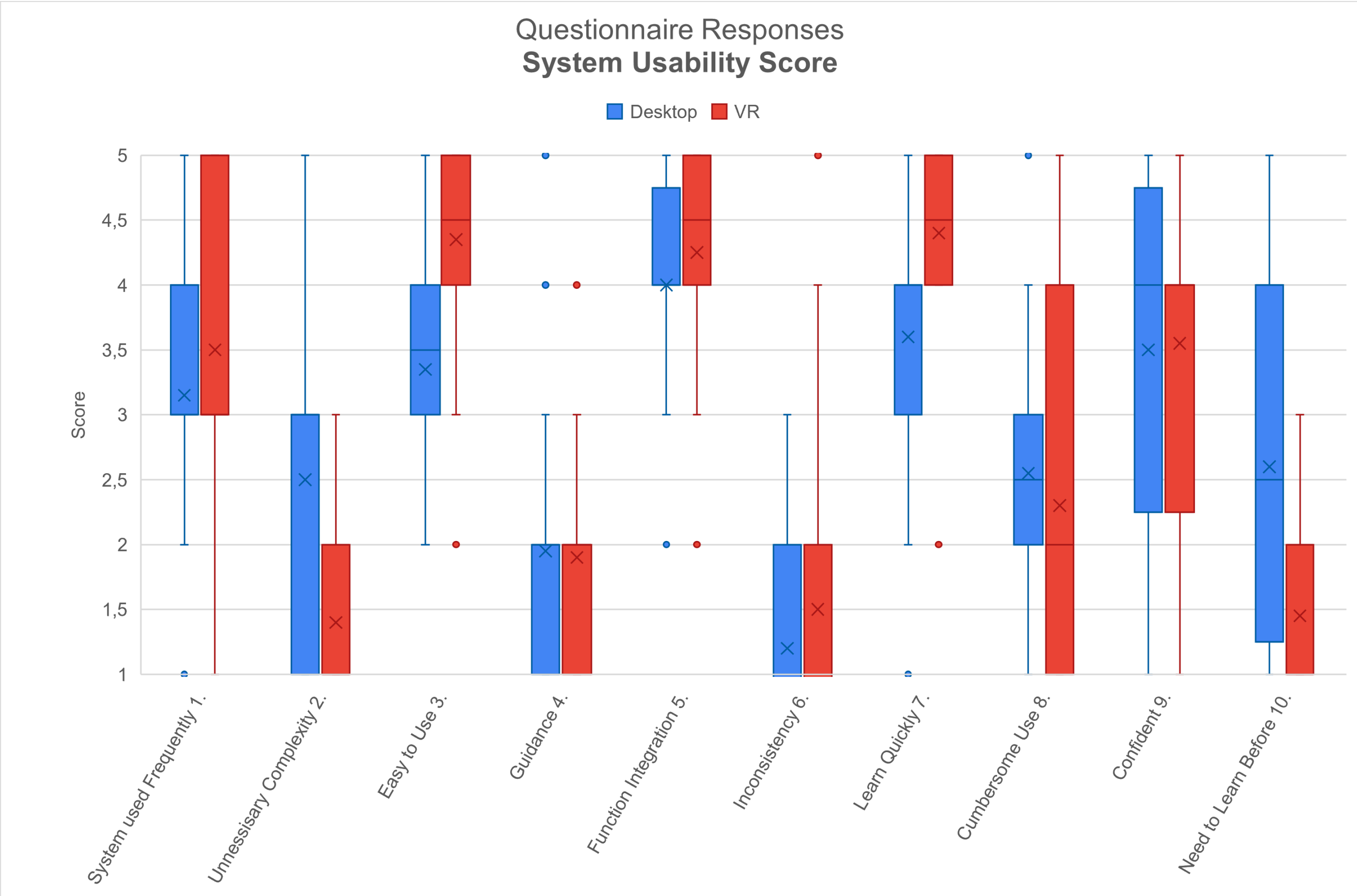

Figure 4.13: Questionnaire ratings of SUS for Desktop and VR respectively. Detailed descriptions of asked SUS questions are found in Section 3.4.3.

In Figure 4.13 an overview of the ratings from the SUS questions is showcased. Alternate questions within the SUS receive higher ratings as their scores increase, whereas the remaining questions contribute more to the cumulative score when rated lower. In other words, questions 1, 3, 5, 7 and 9 should have a score of 5 if the usability is perfect. Meanwhile, questions 2, 4, 6, 8, 10 should get scores of 1 in order for the system to count as the most useful.

At first glance, one can notice the oscillating pattern in the ratings, for both environments. It is easier to observe in the VR case, being the most obvious in questions 1 through 7 and 10. Question 8 had extremely varied results where some participants found the system not cumbersome at all whilst others had difficulties navigating their surroundings. A similar state can be observed with how confident the subject valued themselves in question 9. However, one should keep in mind that it is a subjective statement about themselves and not directed towards the environment itself.

Another interesting observation is how the participants valued themselves towards the environment in question 10. Every participant had more experience with computers than VR, but the Desktop Environment required more learning before they could efficiently use it. The handheld controllers in VR were split into time and marker controls which could have had a significant impact. Furthermore, the participant had to steer themselves around in the environment using the WASD-keys in Desktop, whilst head movement provided the necessary movement to navigate their surroundings in VR.

The final accumulated scores from the SUS questionnaire were 67 and 78.75 for Desktop and VR respectively. A score of 68 is considered average, meaning that the Desktop setup can be considered close to a regularly useful program. Meanwhile, the VR setup seems to display a quite substantial benefit in usability with a score of 78.75.

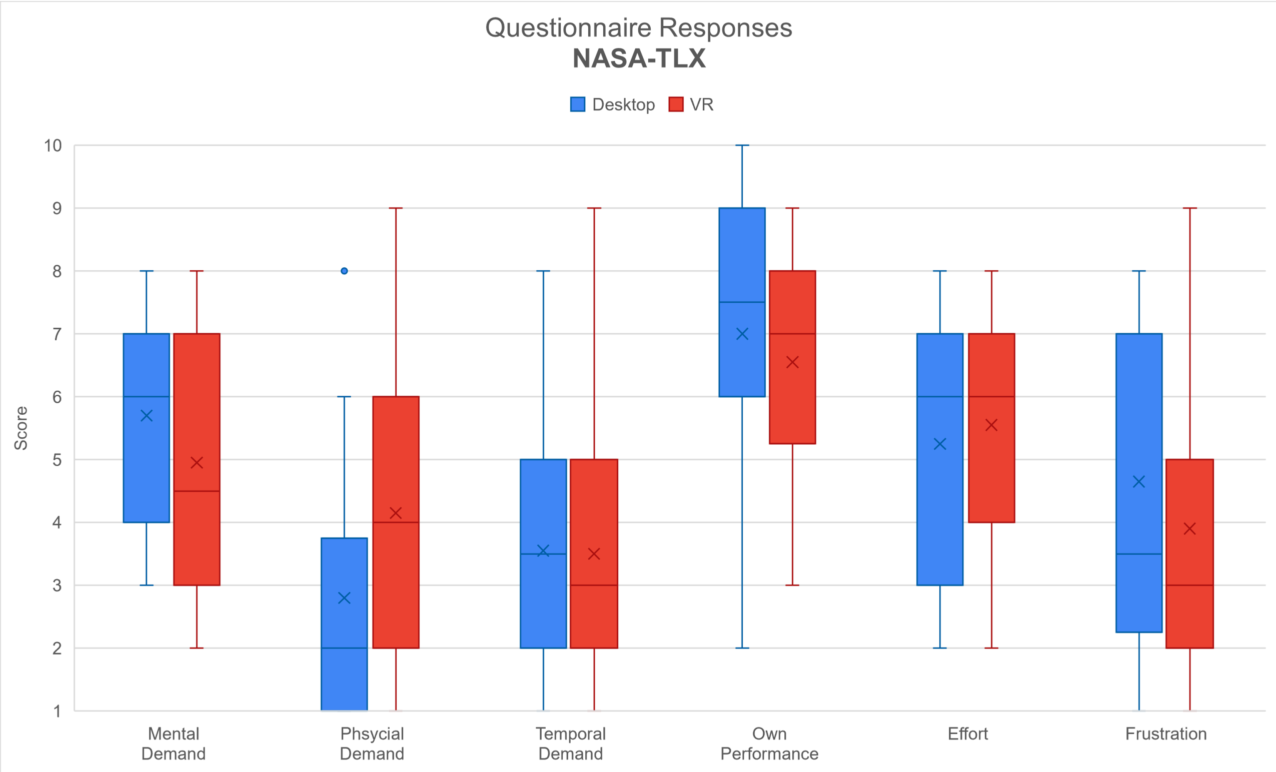

Figure 4.14: Questionnaire ratings of NASA-TLX for Desktop and VR respectively. Detailed descriptions of asked NASA-TLX questions are found in Table 3.4.4.

In Figure 4.14 an overview of the ratings from the NASA-TLX questions is showcased. No significant difference between Desktop and VR can be found within the NASA-TLX ratings. While a minor hypothesis within the NASA-TLX questions would be that Desktop requires more mental demand and VR requires more physical demand, only the latter can be observed to some extent.

The final accumulated scores for the NASA-TLX are extremely similar. Desktop received a NASA-TLX score of 43.3 and VR a score of 44.2 identifying both as having a somewhat high workload. It suggests that the task required a noticeable amount of mental effort and concentration, but it was not excessively demanding. The participants might have experienced some level of stress or cognitive load, but not to an overwhelming extent. Overall, this score indicates that the task was manageable, but it did require a moderate degree of mental resources.

Chapter 5

Discussion

5.1 Hypothesis

The research question for this Master's thesis *How do Desktop environments compare to immersive HMD VR environments, in regards to manipulating spatio-temporal playback of pre-recorded MoCap data?"*, gave way to four hypotheses {**H1, H2, H3, H4**}.

- **H1** postulated that the Desktop controls to manipulate the spatio-temporal data would be easier to understand compared to the VR controls.
- **H2** proposed that the Completion Time would be faster in the VR setting.
- **H3** expected that the task would have higher accuracy and precision in both spatial and temporal assessments in VR.
- Lastly, **H4** envisioned that the guidance laser would help the user determine distances in the virtual world and thus minimise the spatial errors.

The objective evaluation metrics were pertinent to the hypotheses of the experiment. **H1** related to the steps taken, use of controls and Completion Time. **H2** was directly connected to the Completion Time metric. Lastly, **H3** and **H4** were related to the MPE and Marker Size measurements. Additionally, **H1** could also be investigated using the more subjective metrics acquired from the questionnaire.

The initial hypothesis, **H1**, posited that the average user, who is more familiar with keyboard and mouse controls, would find desktop controls to be more readily accessible. However, this proved to be incorrect. The average participant in the study displayed more comfort using the VR controls. Comments made about the Desktop controls, such placement of keys such as the rotating the camera or the key to increase speed were made. Besides the shaking of VR controllers and size of the brain model feeling too small, VR receive less visual, verbal and written frustration. This was also observed in Table 4.14 where the last question regarding frustration was lower. It was also easier to use the VR controllers from how participants maneuvered towards and around the ventricle. In Desktop, going straight up and down with Q and R was almost never utilised. Some participants gave up on using the main camera ability and instead proceeded to place down the marker, utilise the ROAM ability, pick up the marker and continue forward. Meanwhile, in VR, participants understood the controls and use them as intended. However, in four cases, after the participants found their frame of impact, they placed the time controller in their knee and proceeded to use both hands to obtain better precision in placing the marker. From the analysis in Section 4.1, a conclusion could be drawn that the Completion Time was almost twice as fast in VR. Being significant between the two environments with a p-value of < 0.001, the Completion Time showed that participants could operate in the VR environment quicker. After the introductory scene and the first true trial, the Completion Time levelled out for VR whilst still decreasing for Desktop, continuously decreasing until and including the last trial. This also promotes that the learning curve for the controllers used in Desktop takes longer time. Thus, the **H1** hypothesis is rejected. However, the results do confirm the **H2** hypothesis, as it was hypothesised that the task would be completed faster in VR after repeated trials. Incidentally, the user did not only become faster in VR than Desktop after repeated trial, but the user was also faster in VR to begin with. Regardless whether the user had been instructed in Desktop already or not, or whether the user had previous VR experience, the on-boarding process in the VR environment also took considerably less time. In a few cases observed by the experimenter, the *"wow-factor"* of VR was palpably present, but not to the point where it hindered the user from accepting the new reality and completing the task.

The VR controls were more easily accessible. However, this did not imply that they also gave way to better user performance or also to some extent UX. VR performed worse in the other evaluation metrics and parts of the questionnaire. As observed at first in 4.1, Desktop performed better in terms of reduced Frame Error, lower MPE and smaller Marker Size; all being significant with a p-value ≤ 0.01. In terms of the mean and standard deviation, besides the standard deviation for Marker Size, Desktop showed to be the superior setting. This contradicts the **H3** hypothesis, which can therefore be rejected.

Next, the results regarding the UX could highlight important factors in user performance. Here, comments given on the VR questionnaire included experienced trouble with the VR shaking. Due to the limitation of the VR tracking system, the controllers showed some minor shaking which came across as an inconvenience when placing the marker. This is a possible cause of the discrepancy between VR and Desktop in regards to the MPE.

The guidance laser proved to have a weak impact on user performance. As described in Section 4.6, the comparison between having the laser active and not having it, in both VR and Desktop, showed little difference on precision or accuracy. The magnitude of the MPE with the laser was close to equal in the two settings. This contradicts the **H4** hypothesis, which can thus be rejected. Moreover, the laser guidance failed to enhance precision. Given H4, the marker would perhaps also improve the users precision, meaning that they feel more secure in their placement, decreasing the Marker Size. This was shown not to be the case. In accordance with the monocular and binocular cues, the marker should hypothetically give heightened depth perception, by increasing the perception with relative size differences, interposition and linear perspective. When entering the room it could also be argued that aerial perspective could play a role given the intrinsic fog present in the unity environment, thus changing the colour of the laser. The aerial perspective is always present in the environment and does not work on the distance at which the user is often located in relation to the skull. The fog is instead introduced after about two metres. During the implementation phase, experimentation was conducted to assess the viability of incorporating fog. However, due to the dynamic movements of the player within the environment, it was concluded that fog would not be a feasible option. Throughout the trials, the user was involved in observing the skull from a distance as well as investigating the ventricle and catheter movements in detail. Thus, if the fog started further away from the camera, it would have no impact on the details when viewing them up-close. If objects would only change within an arm's length of the user's viewing perspective, they would not be able to see the skull or room until they were adjacent to the objects. Ultimately, no fog or render style based on distance seemed appropriate.

5.2 Controls

In order to compare the Desktop setting to the VR one, both environments had to contain similar or identical controls options. This meant that if one action, restraint or manipulation over the data was implemented in one environment, a comparable one had to be developed for the other setting. Although this gave way to an impartial and objective comparison of the two systems, it did not allow for both environments to play to their own benefits. One such implementation made in the early stages of the development process was the ability to pick up, rotate and scale up the skull-catheter animation in the VR environment. The utility was later discarded since a fair and equal one was not applicable in the Desktop setting. While it would be quite beneficial to pan, zoom in and rotate the animation with the VR controllers, the correlation analysis would be biased and flawed. Another implementation that could have improved the VR system was a smooth pull" of the marker after placing it [44], to allow for minor readjustment of its position in space. This also came across in the comments, where some participants suggested controls for fine-tuning already-placed markers or to scale up the entire scene to enhance placement precision. One constraint and limitation in the two environments was that the skull was kept the same size as it would appear in real life, arguing that it would make it more authentic. For instance, simulations in XR equipment such as the NeuroVR [10] and ImmersiveVR [6] simulates with the same proportions as real life, which allows the brain to train muscle and photographic memory in a controlled environment. However, it can also be argued that this experience could have been aided by more control over the environment. Since the user is supposed to understand the operation in detail, allowing users to see the operation at a different scale might facilitate learning.

The control placement on the keyboard received a few comments from participants with highly rated gaming experience. They were used to the ability to move forward with increased speed being located to the SHIFT key, which caused some issues with UX. Those participants showed verbal or visible frustration with changing the sprint ability with the ability to resize the marker. The sprint ability was introduced at the final stages of the implementation, after feedback had been given during the pilot tests. Before that, the Marker Size control had already been implemented and been assigned the SHIFT button. Thus, the CTRL-key was given the *SPEED* ability.

Moreover, the brain model in Unity came with a toggle-translucency option. This ability allowed the user to toggle the transparency of the skull, changing its texture from a fully opaque skull-material, to a cloudy translucent one. This allowed the user not only to see the catheter go through the hole in the skull, but also see what happens inside as it heads towards and punctures the ventricle. However, it was observed during the pilot phase that the included brain transparency option was only used once in the beginning and then never again. Therefore, the option to change the transparency of the model was excluded from the model viewer, minimising the amount of additional steps the user had to perform at the beginning of each trial. If it was only required to do once, at the start of trial one, the trials would not be considered equal.

Following the experiment, there might be a debate regarding the efficacy of the introductory scene as a successful introduction to the experiment. The user had access to all the controls that were available during the trials. They could pick up the maker and were supposed to place it in a specific spot together with scaling it up to the appropriate size. However, there was no animation which changed as the participant used the temporal controls. The subject could select the slider and see it move as well as control it with the *StepFrame, PlayPause* and *WindFrame* buttons, on both keyboard and the VR controller. Every participant was tasked to use all controls in the introductory scene, which included using every temporal control. However, giving the participant feedback on their actions, in terms of more than the slider, such as an animation, might have been beneficial.

5.3 Errors

A reason for the discrepancy in Completion Time in Desktop compared to VR is the difference in gaming experience. Most participants had the same amount of VR expertise. This meant that everyone was going into the VR environment with a similar amount of practice. This effect can be observed in Table 4.5 where the previous experience in VR did not have a significance on the VR performance.

A notable observation when analysing Frame Error in different animations is the infrequent occurrence of participants choosing a frame before the initial contact point. In most trials, participants selected a frame where the catheter was clearly inside the ventricle. However, in a few exceptional cases, participants chose a frame corresponding to a second impact or when the catheter exited the ventricle.

One significant factor contributing to the occasionally large Frame Error in certain cases was how the participants utilised the temporal controls. As observed in a previous study, some participants tended to overshoot" the animation and navigate backward. Consequently, they might have viewed the second instance when the catheter entered the ventricle or when it was withdrawn from it.

The notable difference in standard deviation of the MPE within animations compared to the overall MPE might be attributed to how each animation shows the catheter's entry into the ventricle. In the first and third recording, the catheter made contact with the top of the ventricle, while in the second recording, the catheter entered the ventricle from the side. Since the ventricle system model includes both lateral ventricles, it means that the relative size of them makes the closest one bigger which helps with perception. Although only one of the lateral ventricles are being operated on, if only one of the ventricles was displayed there would be a lack of relative size, since no other brain structures are shown inside the transparent skull, and thus potentially hinder the perception and distance judgement.

The experiments did not impose a specific limit for Completion Time. However, participants were advised to prioritize precision over speed while performing the tasks, being rewarded with sweets at the end of the experiment. Given the subjective nature of assessing the effort exerted by each individual during the experiment, only motivational encouragements were provided as a form of treatment. During opposition, a suggestion arose to introduce a visual scoring system upon completion of each trial. This scoring mechanism would offer participants feedback on their performance based on the accuracy of their spatial and temporal placement within the true frame and impact point in space. While displaying the score continuously throughout the trials could potentially compromise the integrity of the entire experiment, presenting it at the end of each trial might incentivise participants to strive for better performance. Nonetheless, implementing a visual guide would be necessary to effectively communicate the degree of spatial-temporal accuracy achieved. However, there were considerations regarding the potential misleading interpretation of a MPE if displayed directly. For instance, a $5mm$ error might inappropriately suggest closeness to the target when, in reality, it represents a substantial deviation. To mitigate this, a transfer function translating numerical scores into colors could be employed. This colour scheme could indicate a poor score with red and a good score with green, providing a more intuitive representation. However, determining the appropriate mapping from numerical values to colors would require careful evaluation, potentially through a larger pilot study or similar assessments.

A threat to validity and error was the tracking of the VR system. The tracking precision of both the VR HMD and VR controllers achieves a professional-grade level of fidelity; however, there remains a margin for potential error. While the normalization and regular updates of the base towers' tracking were maintained, considering a measurement of the tracking offset could have proven advantageous. Though challenging to conduct throughout each trial, assessing the offset at the onset of every trial could have significantly enhanced the test's validity. Several precautions were implemented to mitigate tracking issues, such as using curtains to minimize variations in ambient light between and during trials. Additionally, ensuring unobstructed visibility of all base towers from the participant's perspective, without any interference from, for instance, the experimenter, was a priority. Given the participant's freedom of movement within the room, even though restricted to an office chair, situations occasionally arose where one of the three base tracking stations might momentarily have lost sight of a controller. To mitigate this, the virtual animation was positioned at chest height while seated, minimizing potential disruptions. Moreover, the arrangement of the three base stations made it challenging, if not impossible, to obstruct more than one station at a time during the experiment.

One argument to consider is the inherent advantage of the Desktop experience, given that individuals spend increasingly more time in front of a computer or laptop equipped with familiar instruments such as keyboards, mice, and monitors, compared to their exposure to VR. On the other hand, while the individual might have more time spent in front of a 2D monitor, their interaction with a 3D world using such a system is most likely minimal, only relating to flat 2D images of documents and web pages. Furthermore, VR offers a simulation closer to reality, where human gestures and head movements in the virtual world mimic real-world interactions. Therefore, one might contend that individuals possess a substantial amount of experiential knowledge within the VR realm, even if the participants had not previously used a HMD. The experiences with both environments might then be considered more similar than not. Future studies could delve into exploring the nuances and distinctions between VR experiences and reality, particularly with advancements in graphics, detail, and interaction, to unravel how individuals engage and perceive these distinct domains.

5.4 Future Work

The analysis of the results from the NASA-TLX questionnaire pointed towards both systems having a somewhat high workload on the participant for this task. This means that both environments could be further improved by better interactions and visualisations. Better, in terms of controller layout, different controls and tailoring the experiences more towards the two systems. In other words, having each environment exploit more into the benefits of their interaction paradigm. However, as mentioned, this would make comparing the two environments more difficult since they are less equivalent in interaction. Similarly, the visualisations could also be different in the two environments, such as the improvements proposed by Hombeck et al. in [10].

A further investigation could explore the training aspect of neurosurgical visualisation, whether the current setup is too distracting in VR for educational purposes. The participant might be more immersed and engaged, as seen in Table 4.12 and observed in research [7, 11], but it might not mean that they are learning more. Future work could evaluate if the participant understands the EVD procedure better from only interacting and observing the operation in VR or only in Desktop, given the same environment. The same applies to testing the two environments when more details are present.

Perhaps an overlooked issue with the MoCap recording is that the surgeon is completely excluded from the recording. Only the skull and catheter are tracked and recorded in space. The translation into the VR environment means that a model for the catheter, brain, as well as the ventricles inside it can be placed. However, with the lack of hand placement, arm tracking and body position of the surgeon, education on the full procedure is incomplete. Future work could include these details and study the difference between including a virtual-agent proxy the surgeon and the current implementation.

In VR a better depiction of reality can be experienced [7, 25], given that both VR and the Desktop environments are the same. Ultimately, reality is what the surgeon has to contend with. With the current setup, the user is not evaluated based on learning. Perhaps the user learn better in a Desktop environment and can practise and simulate the operation in a VR setting, such as the NeuroVR or ImmersiveVR [5, 8]. The results from this project show that VR is faster for the current task with pre-MoCaped EVD visualisation, however not that it is better in precision or accuracy.

This Master's thesis study was executed with the help of engineering students from the Royal Institute of Technology. The students exhibited varying amounts of gaming and VR experience. Repeating the study with medical students would give an arguably better representation of the impact of VR. Moreover, experimenting with medical students or surgeons proficient in, or having knowledge of, cranial neurosurgical procedures (including EVD) would provide an ideal estimate of VR's utility in this field.

Chapter 6

Conclusion

To recapitulate, the primary motivation behind the project was to assess the effectiveness of using VR technology compared to traditional 2D monitors for visualising and manipulating pre-recorded MoCap spatio-temporal playback of surgical data, particularly focusing on EVD. Additionally, a goal was to develop a navigation environment in Unity that supported both VR and desktop settings for visualising the given data. The environment was able to record the essential evaluation parameters including completion time, accuracy, precision and the usage of different interaction controls. From analysis of the collected data, the controls to manipulate the spatio-temporal data were easier to understand in the Virtual Reality setting. It was also considerable shorter to complete the task in the VR environment. However, the desktop environment yielded higher accuracy and precision in both time and space. The integrated guidance laser did not help with perceiving the relative distance to objects. Previous VR experience did not help with user performance, yet gaming experience did have a positive impact throughout desktop and VR. Whilst VR was faster, featuring more engagement and immersion for the user, VR compared worse to desktop where precision, accuracy and the stable controlling of the camera allowed for better manipulation of the spatio-temporal data. Thus, desktop appears to be the superior option for data manipulation, yet VR should still be considered to be included as an alternative tool to be used for planning and educational purposes in cranial surgery. Future work and developments could focus on the increased precision in interactive in VR, with the aid of haptic feedback, minor adjustments and scalability.

Bibliography

[1] T. Mazur, T. R. Mansour, L. Mugge, and A. Medhkour, "Virtual reality-based simulators for cranial tumor surgery: A systematic review." *World neurosurgery*, vol. 110, pp. 414–422, 2018. [Pages 1 and 8.]

[2] A. Iop, V. G. El-Hajj, M. Gharios, A. de Giorgio, F. M. Monetti, E. Edström, A. Elmi-Terander, and M. Romero, "Extended reality in neurosurgical education: A systematic review," *Sensors (Basel, Switzerland)*, vol. 22, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:251638669 [Pages 1, 2, 9, 18, 19, and 29.]

[3] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," *Comput. Educ.*, vol. 147, p. 103778, 2020. [Pages 1 and 9.]

[4] A. M. Okamura, "Haptic feedback in robot-assisted minimally invasive surgery," *Current Opinion in Urology*, vol. 19, p. 102–107, 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:18121219 [Page 1.]

[5] G. Riva, A. Gaggioli, D. Villani, A. Preziosa, F. Morganti, R. Corsi, G. Faletti, and L. Vezzadini, "Neurovr: An open source virtual reality platform for clinical psychology and behavioral neurosciences," *Studies in health technology and informatics*, vol. 125, pp. 394–9, 2007. [Online]. Available: https://api.semanticscholar.org/CorpusID:18825251 [Pages 1, 8, 19, and 76.]

[6] R. A. Kockro, A. T. Stadie, E. S. Schwandt, R. Reisch, C. Charalampaki, I. H. B. Ng, T. T. Yeo, P. Y. K. Hwang, L. Serra, and A. Perneczky, "A collaborative virtual reality environment for neurosurgical planning and training," *Operative Neurosurgery*, vol. 61, p. 379–391, 2007. [Pages 1, 8, 19, and 72.]

[7] M. Ros, B. Debien, C. Cyteval, N. Molinari, F. Gatto, and N. Lonjon, "Applying an immersive tutorial in virtual reality to learning a new technique." *Neuro-Chirurgie*, 2020. [Pages 3, 9, 10, 29, and 76.]

[8] D. A. Bowman and R. P. McMahan, "Virtual reality: How much immersion is enough?" *Computer*, vol. 40, 2007. [Pages 3, 15, 29, and 76.]

[9] G. Makransky, T. Terkildsen, and R. E. Mayer, "Adding immersive virtual reality to a science lab simulation causes more presence but less learning," *Learning and Instruction*, 2017. [Pages 3, 9, and 10.]

[10] J. N. Hombeck, M. Meuschke, L. Zyla, A.-J. Heuser, J. F. Toader, F. Popp, C. J. Bruns, C. Hansen, R. R. Datta, and K. Lawonn, "Evaluating perceptional tasks for medicine: A comparative user study between a virtual reality and a desktop application," *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 514–523, 2022. [Pages 6, 13, 24, 72, and 76.]

[11] M. E. Beheiry, S. Doutreligne, C. Caporal, Ć. Ostertag, M. Dahan, and J.-B. Masson, "Virtual reality: Beyond visualization." *Journal of molecular biology*, vol. 431 7, pp. 1315–1321, 2019. [Pages 7, 8, 9, and 76.]

[12] M. Oberhauser and D. Dreyer, "A virtual reality flight simulator for human factors engineering," *Cognition, Technology & Work*, vol. 19, pp. 263 – 277, 2017. [Page 8.]

[13] S. G. Izard and J. A. J. Méndez, "Virtual reality medical training system," *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, 2016. [Pages 9 and 29.]

[14] L. Fiorella, T. van Gog, V. Hoogerheide, and R. E. Mayer, "It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task," *Journal of Educational Psychology*, vol. 109, p. 653–665, 2017. [Page 9.]

[15] S. A. W. Andersen, S. Foghsgaard, P. Cayé-Thomasen, and M. S. Sørensen, "The effect of a distributed virtual reality simulation training program on dissection mastoidectomy performance." *Otology & Neurotology*, 2018. [Page 9.]

[16] C. M. Harrington, D. O. Kavanagh, G. Wright Ballester, A. Wright Ballester, P. Dicker, O. Traynor, A. Hill, and

S. Tierney, “360° Operative Videos: A Randomised Cross-Over Study Evaluating Attentiveness and Information Retention,” *Journal of Surgical Education*, vol. 75, no. 4, pp. 993–1000, Jul. 2018. doi: 10.1016/j.jsurg.2017.10.010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1931720417302945 [Pages 9, 13, and 19.]

[17] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and L. Mg, “Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness.” *The International Journal of Aviation Psychology*, vol. 3, pp. 203–220, 1993. [Page 10.]

[18] M. Dolezal, J. Chmelík, and F. Liarokapis, “An immersive virtual environment for collaborative geovisualization,” *2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pp. 272–275, 2017. [Pages 10 and 11.]

[19] M. Potel, “Motion sick in cyberspace,” *IEEE Computer Graphics and Applications*, vol. 18, pp. 16–21, 1998. [Online]. Available: https://api.semanticscholar.org/CorpusID:39354058 [Page 10.]

[20] E. Bozgeyikli, A. Raij, S. Katkoori, and R. V. Dubey, “Point & teleport locomotion technique for virtual reality,” *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, 2016. [Page 11.]

[21] S. Tregillus and E. Folmer, “Vr-step: Walking-in-place using inertial sensing for hands free navigation in mobile vr environments,” *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016. [Page 11.]

[22] L. Fu, J. Zhu, W. Li, Q. Zhu, B. Xu, Y. Xie, Y. Zhang, Y. Hu, J. Lu, P. Dang, and J. You, “Tunnel vision optimization method for vr flood scenes based on gaussian blur,” *International Journal of Digital Earth*, vol. 14, pp. 821 – 835, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:233975498 [Page 11.]

[23] K. Choros and P. Nippe, “Software techniques to reduce cybersickness among users of immersive virtual reality environments,” 2019. [Page 11.]

[24] M. Kalloniatis and C. Luu, “The perception of depth,” 2007. [Pages 11 and 12.]

[25] M. Kersten-Oertel, S. J.-S. Chen, and D. L. Collins, "An evaluation of depth enhancing perceptual cues for vascular volume visualization in neurosurgery," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, pp. 391–403, 2014. [Pages 11, 12, and 76.]

[26] D. M. Coffey, F. Korsakov, M. Ewert, H. Hagh-Shenas, L. Thorson, A. M. Ellingson, D. J. Nuckley, and D. F. Keefe, "Visualizing motion data in virtual reality: Understanding the roles of animation, interaction, and static presentation," *Computer Graphics Forum*, vol. 31, 2012. [Page 13.]

[27] J. Dobos, N. J. Mitra, and A. Steed, "3d timeline: Reverse engineering of a part-based provenance from consecutive 3d models," *Computer Graphics Forum*, vol. 33, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:2196547 [Page 13.]

[28] S. Sharma, S. Verma, M. Kumar, and L. Sharma, "Use of motion capture in 3d animation: Motion capture systems, challenges, and recent trends," *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, pp. 289–294, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:204231221 [Page 14.]

[29] D. P. Noonan, P. Mountney, D. S. Elson, A. Darzi, and G.-Z. Yang, "A stereoscopic fibroscope for camera motion and 3d depth recovery during minimally invasive surgery," *2009 IEEE International Conference on Robotics and Automation*, pp. 4463–4468, 2009. [Page 14.]

[30] H. zhi Hu, X. bo Feng, Z.-W. Shao, M. Xie, S. Xu, X. huo Wu, and Z. Ye, "Application and prospect of mixed reality technology in medical field," *Current Medical Science*, vol. 39, pp. 1–6, 2019. [Page 14.]

[31] L. yu Li, J. Yang, Y. Chu, W. Wu, J. Xue, P. Liang, and L. Chen, "A novel augmented reality navigation system for endoscopic sinus and skull base surgery: A feasibility study," *PLoS ONE*, vol. 11, 2016. [Page 14.]

[32] R. Muralidharan, "External ventricular drains: Management and complications," *Surgical Neurology International*, vol. 6, pp. S271 – S274, 2015. [Pages 14 and 15.]

[33] C. I. Jennett, A. L. Cox, P. A. Cairns, S. Dhoparee, A. Epps, T. J. W. Tijs, and A. Walton, "Measuring and defining the experience of immersion

in games," *Int. J. Hum. Comput. Stud.*, vol. 66, pp. 641–661, 2008. [Online]. Available: https://api.semanticscholar.org/CorpusID:1462839 [Pages 15 and 17.]

[34] W. W. IJsselsteijn, de Yaw Yvonne Kort, and K. Poels, "The game experience questionnaire," 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:261292894 [Pages 15 and 17.]

[35] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *International Journal of Human–Computer Interaction*, vol. 24, pp. 574 – 594, 2008. [Page 15.]

[36] S. G. Hart, "Nasa-task load index (nasa-tlx); 20 years later," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, pp. 904 – 908, 2006. [Pages 15 and 17.]

[37] J. B. Brooke, "Sus: A 'quick and dirty' usability scale," 1996. [Online]. Available: https://api.semanticscholar.org/CorpusID:107686571 [Page 16.]

[38] J. Sauro, "Measuring usability with the system usability scale (sus)," *measuringu*, February 2011. [Page 16.]

[39] J. B. Brooke, "Sus: a retrospective," *Journal of Usability Studies archive*, vol. 8, pp. 29–40, 2013. [Page 16.]

[40] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," *Advances in psychology*, vol. 52, pp. 139–183, 1988. [Online]. Available: https://api.semanticscholar.org/CorpusID:15252590 [Page 17.]

[41] J. Sauro, "10 things to know about the nasa tlx," *measuringU*, August 2019. [Page 17.]

[42] P. A. Hancock and N. Meshkati, "Human mental workload," 1988. [Online]. Available: https://api.semanticscholar.org/CorpusID:61029959 [Page 17.]

[43] "Gopro | world's most versatile cameras," *Gopro*, 2023. [Online]. Available: https://gopro.com/en/se/ [Page 19.]

[44] W. Usher, P. Klacansky, F. Federer, P.-T. Bremer, A. Knoll, J. Yarch, A. Angelucci, and V. Pascucci, "A virtual reality visualization tool

for neuron tracing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, pp. 994–1003, 2018. [Pages 20, 21, and 72.]

[45] M. Holmberg, M. Nyman, L. Prémont, and M. Sherzad, "Teamdreammocapproject," *DM2799 Advanced Projects in Interactive Media Technology*, 2022. [Online]. Available: https://github.com/MarcusHolmbrg/TeamDreamMocapProject [Page 23.]

[46] V. Corporation, "Steamvr," 2023. [Online]. Available: https://store.steampowered.com/app/250820/SteamVR/ [Page 24.]

[47] "Ishihara test | color test | ishihara chart," 2023. [Online]. Available: https://www.colorblindnesstest.org/ishihara-test/ [Pages 27 and 28.]

[48] U. Technologies. (2023) Unity game engine. [Page 35.]

[49] Nintendo, "The-legend-of-zelda-ocarina-of-time," *Nintendo 64*, November 1998. [Online]. Available: https://www.nintendo.co.uk/Games/Nintendo-64/The-Legend-of-Zelda-Ocarina-of-Time-269536.html [Page 35.]

[50] A. Inc., "Youtube," February 2005. [Online]. Available: https://www.youtube.com/ [Page 35.]

[51] "Ibm spss statistics," *International Business Machines*, 2023. [Online]. Available: https://www.ibm.com/products/spss-statistics [Page 43.]

Chapter 7

Appendix

Github repository for the project.

https://github.com/glas444/Neuro

User Study Questionnaire

https://docs.google.com/forms/d/e/1FAIpQLScSrqos-RDYUSujDo0JrVnJ6SMojhaAq7fpQ1h5QY5rXEwwdQ/viewform?usp=sf_link