UNIVERSITY
OF OULU

FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

**Nikunj Arora**

# AUGMENTING VIRTUAL REALITY TELEPRESENCE EXPERIENCE USING SELF-AVATAR

Master s Thesis
Degree Programme in Computer Science and Engineering
June 2021

**Arora N. (2021) Augmenting Virtual Reality Telepresence Experience Using Self-Avatar.** University of Oulu, Degree Programme in Computer Science and Engineering, 63 p.

# ABSTRACT

**Telepresence refers to a set of techniques that creates the illusion of being present at a remote location to a person. Telepresence may also include the ability to interact with the remote environment, including communication with people physically present at the remote location. In this research, the introduction of a virtual body, which mirrors the user's own movement in real-time, in a telepresence scenario and its effect on the illusion of presence is studied. Earlier research works have shown the effectiveness of having a virtual body in simulated environments, for example, games. In this study, the user embodies a virtual body that is present in a simulated environment, surrounded by a sphere where footage streamed from a 360-degree camera, mounted at a different spot, is being projected. This gives the user a sense of being present in a real location and having a body which they can control.**

**The study is conducted on 20 participants, where the participants put on a Head-Mounted Display showing live footage from a 360-degree camera while having a real-time conversation with a confederate present near the camera. They are then surveyed about their experience, both with and without a virtual body to determine if having a virtual body yielded any improvement on the illusion of presence. Although 18 of the 20 participants preferred the experience with a body, it did not necessarily increase their sense of presence when compared with the scores given when there is no visible body. These results implicate a low sample size, not enough to draw any meaningful conclusions.**

**Keywords: Extended Reality, Telepresent human, Body tracking, Inverse Kinematics, 360 live streaming**

# TABLE OF CONTENTS

# FOREWORD

This thesis is written to fulfill the graduation requirement of the Computer Science and Engineering Master's program at the University of Oulu, Oulu, Finland. It is a continuation of the work done during my role as a research assistant at the university during the summer of 2020. The topic was concocted after several days of brainstorming and skimming through existing research on the topics of virtual reality and telepresence.

The basis of the research stems from a bigger research project currently being conducted at the Perception Engineering lab in Center for Ubiquitous Computing (UBICOMP), University of Oulu. The perception engineering research team focuses on topics such as Robotics and Extended Reality (XR), which naturally includes the project of Virtual Reality Telepresence.

I would like to thank my supervisor, Dr. Markku Suomalainen for his continuous support throughout the thesis process, along with the second examiner, Basak Sakcak, the technical supervisor, Katherine Mimnaugh, and Matti Pouke for their valuable insights on my progress and for providing useful tips and feedback based on their experience on similar studies. I also thank Alexis Chambers for her key role in the study as the one who interacts with the participants. I would also like to thank all the participants of the study without whose cooperation I would not have been able to conduct this analysis. A special thanks to my friend Rohit Mishra for being an early test subject during implementation phase, and also providing me with up-to-date references of studies in the education field which is theoretically related to my study. Above all, I would like to thank the UBICOMP department for the opportunity and the generous funding provided for conducting this thesis. Lastly I would like to thank my friends and family for their motivation and support through the difficult times, especially during a global pandemic.

Oulu, June 10th, 2021

Nikunj Arora

# LIST OF ABBREVIATIONS AND SYMBOLS

| | |
|---|---|
| VR | Virtual Reality |
| AR | Augmented Reality |
| MR | Mixed Reality |
| XR | Extended Reality |
| HMD | Head-Mounted Display |
| VE | Virtual Environment |
| UBICOMP | Ubiquitous Computing |
| FK | Forward Kinematics |
| IK | Inverse Kinematics |
| IPD | Inter Pupillary Distance |
| PI | Place Illusion |
| Psi | Plausibility Illusion |
| CAVE | Cave Automatic Virtual Environment |
| SUS | Slater-Usoh-Steed |

# 1. INTRODUCTION

Virtual Reality (VR) has gained signi cant limelight since the commencement of the 4th industrial revolution. As the name implies, VR technology simulates an experience that corresponds to a real-life experience. A Virtual Environment (VE) is simulated with the intention of making it as immersive as possible. The most common way to experience an immersive VR is through the use of a display system known as the Head-Mounted Display (HMD). At its root, the HMD contains a headset which is worn over the head, along with two lenses for each of the eye displaying the scene in the VE and internal sensors which track the head movement for updating the view in the headset depending on the head position and orientation. This deceives the human visual system into believing they are present inside the VE. Akin to VR, there are other technologies for mixing the real world with the virtual world, namely Augmented Reality (AR) and Mixed Reality (MR). All these technologies collectively belong to a eld known as Extended Reality (XR).

The ongoing research and development in VR are working towards the end goal of total immersion in a VE. Mel Slater discusses the importance of the illusion of presence to a participant in his study [1] and how participants respond realistically when they are immersed inside a VE. Having the user get a sense of realism inside the VE is useful in social interactions and a number of applications ranging from therapy to rehabilitation as discussed in section 2.1.5. Over the years, several approaches have been researched such as the introduction of a bidirectional treadmill for allowing natural movement in the VE [2], introduction of haptic feedback from the VE onto the real body [3], and nally introduction of a virtual body in the VE. For example, Caserman et al. conducted a study [4] with the objective of measuring the immersiveness by introducing a full-body motion capture system that copies the movements of the user s entire body onto a virtual body, using HTC Vive Trackers and observed that 92.3% of the participants would like to have a visible and controllable body in VR games.

As noted by the previous examples, most studies focus on improving the user s illusion of presence in a VE, overshadowing the potential of using VR for enhancing the eld of telepresence. Telepresence refers to a set of techniques for creating an illusion of being present at a location different from the actual location of the user. This is achieved by deceiving one or more of the human senses such as vision, audio, or touch. The term telepresence was rst coined by Marvin Minsky in 1980. He writes about a scenario where robotic arms are being controlled through your hand movement in real-time at a remote location [5]. One of the approaches for deceiving the senses is by making use of VR technology as mentioned above. Instead of simulating a VE, we show live footage streamed from a 360-degree camera around the user giving them an illusion of being present at the location of the camera. This approach is known as VR Telepresence.

This study focuses on the last aforementioned approach for improving the user s illusion of presence in a VE, probing the aspect of improving VR Telepresence experience by introducing a self-avatar. The self-avatar refers to the process of capturing the motion of the whole or parts of a user s body in the real world ensuing movement of a virtual body, present inside a VE, which the user embodies in a rst-person perspective. Having a self-avatar result in a signi cant improvement in the user s illusion of presence as demonstrated by [6]. This study is based on a similar principle of introducing a self-avatar, but this time in a real environment instead of a simulated one, and the effects it has on the user s sense of presence is studied.

The self-avatar operates by capturing the motion of one or more parts of the user s body. In this case, a three-point motion capture is used which tracks the user s head using the sensors present inside the HMD, and the user s wrists using the two controllers. We can calculate the motion of the entire upper torso by using just the data from head and two wrists by making use of one of the Kinematics techniques. It is common in computer animations and the robotics industry for there to be body parts connected to each other through joints having constrained movement, for example, a human skeleton in a game that needs to be animated. The calculation of the movements is achieved mainly using two techniques, Forward Kinematics (FK) and Inverse Kinematics (IK). FK can be de ned as "The process of obtaining position and velocity of end effector, given the known joint angles and angular velocities" [7 p.12]. IK can be de ned as "The process of obtaining joint angles from known coordinates of end effector" [7 p.13]. Continuing with the example, suppose we want to move the right wrist of the skeleton to a certain point. Using FK, we can only calculate the desired position of the wrist after knowing all the joint rotations, i.e. we need to keep adjusting the shoulder, elbow and wrist rotations until the wrist is at the desired position. IK is a mathematical process that aims to solve this in an opposite manner. We just move the wrist to the desired position and the IK algorithms solve for the rotation of the elbow and shoulder while respecting the joint constraints. In this study, IK is utilized for tracking the user s body.

This thesis aims to study the effect of self-avatar on a user s sense of presence in a VR telepresence experience. The VR telepresence experience is created by live-streaming from a remote 360-degree camera onto a sphere inside a VE surrounding the user s self-avatar. The self-avatar mirrors the user s movement by a three-point tracking system where the user s head and both wrists are tracked and motion of other parts are calculated through an IK algorithm. Having a virtual body in a telepresence scenario can improve experience in certain scenarios such as conferences, lectures, or just general social interactions. Movements of the virtual body can also be transferred to an actual robot body as discussed in section 2.2.2.

# 2. RELATED WORKS AND MOTIVATION

## 2.1. Virtual Reality

### *2.1.1. General*

Virtual Reality (VR) can be de ned as "Inducing targeted behaviour in an organism by using arti cial sensory stimulation, while the organism has little or no awareness of the interference" [8 p.1]. A commercially available VR system generally includes an HMD which is strapped to the user s head with screens viewed through two lenses each positioned in front of the user s eyes. The HMD contains sensors to determine the user s position and orientation in the real world and translates that to the VE to update what part of the scene gets rendered to the screen based on where the user is looking. Rendering refers to the process of projecting a 3D environment onto a 2D screen. Various rendering techniques used for VR are discussed in [9]. VEs are simulated environments that provide an illusion of being present inside the simulated environment. [10] de nes a VE as "Interactive, the virtual image displays enhanced by special processing and by non-visual display modalities, such as auditory and haptic, to convince users that they are immersed in a synthetic space".

The HMD renders separate images inside the two lenses and therefore, requires more computing power than traditional 3D simulations on a single monitor. The reason it needs to render two separate images is to simulate a stereoscopic effect. Most animals, including humans, have a pair of eyes that enable the brain to construct a 3D representation of the world. The HMD displays two separate images of the same scene, viewed from two positions alongside each other in a similar fashion to the human eyes. The brain creating an internal 3D model of the scene, paired with the ability of the HMD to block out the real world from the visual system makes the VE more immersive.

The distance between the centres of our two eyes is known as the Inter-Pupillary Distance (IPD). The IPD of the majority of adults lies between 50-75 mm with a mean of around 63 mm [11]. The lens spacing in the HMD needs to be adjustable to accommodate people with different IPDs. The majority of the modern HMDs have options to adjust the lens spacing for IPDs between 50 mm to 75 mm.

The HMD is the most basic equipment needed for experiencing VR. However, the VR experience can be augmented using additional accessories. The most common accessory used is two controllers for each hand. These are useful for tracking the user s hands as well as provide meaningful input from the user to the VE to make the simulation more interactive. Another common accessory is the base station in the HTC Vive setup. Unlike the Oculus VR setup, which does not need any external sensors to track the user s real-world position and map it to the 3D environment inside the VE, the HTC Vive setup makes use of external sensors (at least two of them) to track the HMD and controllers. This extra step enables more accurate tracking and prevents occlusion of the controllers when they are positioned behind the user. This also enables the use of another accessory called the HTC Vive Tracker, which enables transferring any external objects inside the VE by sticking the marker onto a real object and applying the motion of the real object onto a virtual object inside the VE. There are also several

3rd party accessories, such as the Leap Motion controller for precise nger tracking and VR Treadmills, mainly used in research due to their size and cost.

Figure 1. Oculus Rift S VR system with HMD and two controllers.

### *2.1.2. Presence*

Presence refers to the user s sense of being present inside the VE. For evaluating the sense of presence of the user inside the VE, Mel Slater proposed two constructs in his 2009 article [1], namely, Place Illusion (PI) and Plausibility Illusion (Psi). He de nes PI as "the illusion of being in a place despite the sure knowledge that you are not there" and Psi as "the illusion that what is apparently happening is really happening, despite the sure knowledge that it is not". According to Slater, with PI, the user feels like they are present inside the environment, while with Psi, the environment responds to their actions. In this thesis, the PI is measured while the user is inside a real remote environment.

Another aspect studied in this thesis is the sense of Co-Presence. Similar to PI, Co-Presence is the illusion of being present with someone else. [12] proposes a questionnaire for measuring the sense of Co-Presence. Collaborative VE is a major area that bene ts from improving Co-Presence.

### *2.1.3. History*

The study of stereo vision, stereopsis was first conducted in 1838 by Sir Charles Wheatstone [13] which led him to develop a stereoscope that demonstrated the ability of the human brain to combine two 2D images from each eye to construct a 3d image. The Sword of Damocles developed by Ivan Sutherland in 1968 is regarded as the first VR system with an HMD having a stereo vision [14]. He proposes the idea of placing two-dimensional images in front of the user's retina to provide an illusion of a three-dimensional image.

The term "Virtual Reality" was first used by French playwright Antonin Artaud in his 1938 book "The Theater and Its Double", translated from French in 1958 by Mary Caroline Richards [15 p.49]. It was later popularized by Jaron Lanier in 1984 when he founded the company VPL Research, which was one of the first companies that manufactured commercial virtual reality products [16]. Since the early 1990s, VR was mostly used in VR arcades as a pay-and-play option and in government operations such as training NASA astronauts [17]. The modern VR headsets that can be owned by individuals emerged after the prototype of the PC-connected Oculus Rift in 2010. Other VR headsets soon followed such as HTC Vive, Sony PlayStation VR, etc. [18] has a comprehensive literature review on the history of VR.

### *2.1.4. Limitations and Current Research*

Even though VR is readily available in the market on top of being relatively affordable when compared to earlier days, it is still considered a gimmick by many. One of the reasons for this problem is the lack of content for VR. Big game companies are still focusing on developing traditional games rather than embracing VR technology. The major cause of this issue is also a major drawback of the current VR technology, locomotion. Human beings have a vestibular system inside their ears that provides a sense of orientation and balance to the brain. If the visual system and the vestibular system do not agree on certain occasions, the brain assumes that the body is intoxicated and induces nausea in the body to get rid of those toxins through vomiting. If the camera has motion inside the VE, there is a conflict between the visual system, which is seeing movement in the VE, and the vestibular system, which detects that the user is standing still. This phenomenon is commonly known as VR sickness or simply cybersickness.

Extensive research is being done in the field of improving locomotion in VR. One approach is to use software techniques for reducing cybersickness, such as using teleportation techniques inside VE for navigation. This is the most common technique used in games which works by instantly moving the user to a point they choose using controllers. This however does not solve cybersickness in VEs that rely on moving environments such as roller-coasters. For those we use other techniques such as narrowing the user's vision to prevent them from seeing too much of the moving environment which is the main cause of cybersickness, however, that also does not remove the cybersickness completely. Another technique proposed by Chang et al. relies on rest frames to reduce cybersickness in VEs with moving environments [19].

Another approach is to use hardware solutions such as bidirectional treadmills, known as VR treadmills. The VR treadmills are like regular treadmills, but allow movement in two axes instead of one. The user can move in any direction on the treadmill while remaining stationary in the real world and the treadmill will translate their movement inside the VE. This sort of movement maintains harmony between the visual system and the vestibular system, reducing chances of nausea. An example of such a treadmill is presented in [2].

There are also systems that get rid of the HMD altogether and rely on physical movement for interactions inside the VE. One such example is Cave Automatic Virtual Environment (CAVE) proposed back in 1992 [20]. It works by projecting the VE onto the walls of an enclosed room. The user s position and orientation are tracked using a motion capture system and the VE is updated with respect to that. This is in principle similar to an HMD but on a larger scale. Stereoscopy is achieved through the use of one of many techniques for displaying 3d images such as active shutter glasses [21]. A major drawback of CAVE is that it requires an entire room for it to function and is really costly to set up.

There have even been studies done on the effects of food and drinks on cybersickness in VR. One such example is [22] where the effect of alcohol is studied on cybersickness. In this study, a controlled experiment was performed on two groups of participants, one made to consume alcohol and the others made to consume a placebo. When asked to play VR games, results show that even 0.07% of alcohol level in blood was enough to signi cantly reduce cybersickness.

### *2.1.5. Future Prospects*

So far, the only applications of VR discussed in this thesis are games however, there are a lot of elds where VR has signi cant potential. Since VR focuses on simulating reality, there is a possibility of using VR for therapies. Rovira et al. demonstrate that people respond realistically to events happening in the VE when they conducted a study to measure people s responses to violent incidents inside the VE [23], which proves that VR is a viable platform for rehabilitation and therapies. [24] conducts a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis for VR rehabilitation and therapy elds and concludes that there are a lot of strengths and opportunities in using VR in therapies and the eld is not held back from the few weaknesses and threats.

On a similar note, VR can also be used for improving mental health by simulating nature as demonstrated by Browning et al. in their study [25] where a group of people is exposed to real nature and another group to a 360-degree video of nature inside VR. Results show that both cases instigate positive mood levels when compared to results of another group that stayed indoors without any exposure to nature. Not only mental health but VR can also be used to alleviate the pains of physical injuries which is demonstrated by Das et al. in their study [26] where children between the age of 5-18 years with acute burn injuries take part in a controlled experiment for studying the effect of VR games on the pain due to their burn injuries. Results show that VR games do in fact ease the pains and anxiety from the wounds without any side effects.

VR is not limited to humans only. Studies have been conducted where VR is used on other non-human entities. [27] studies the brain activity of mice running on a spherical treadmill constituting a VR system. Desert Locusts are known to be unpredictable and fatal for the agriculture sector. Virtual Reality technology is being used to subject locusts to different VE and their behaviour is observed [28]. This enables researchers to predict the locusts behaviour in advance and prevent disastrous food shortages. There are also claims of cows producing more milk when they are tted with VR goggles [29]. It does not have to be living organisms either as demonstrated by Suomalainen et al. in their study [30] where the idea of tethering a robot with a VR headset to fool the robot s sensors is discussed which enables testing of robots in targeted environments.

## 2.2. Telepresence

### *2.2.1. General and History*

The recent SARS-CoV-2 pandemic has forced in person meetings among people to dwindle, while remote communication methods and work from home techniques are ourishing. The eld of telepresence is gaining traction as one of the candidates for aiding the vision of visiting any place in the world without physically going there.

Telepresence does not have a well-structured de nition yet. It can be roughly described as "A technique for emphasizing the interaction between people and remote sites and enabling people who are immersed in telepresence systems to act and receive input as if they were at the remote site" [31]. It was rst introduced in an Omni magazine article in 1980 by Minsky [5]. The basic idea of telepresence is that a user feels like they are present at a remote location different from their physical location. Telepresence is usually coupled with a teleoperation system. A teleoperation system performs certain tasks at the remote location after getting interacted with by the user. [32] discusses measures to determine whether a teleoperation system can be considered a telepresence system and whether the user can be labelled as a telepresent human. It also presents a concept for comparing two telepresence systems and determining whether one user is more telepresent than another.

Telepresence can be considered analogous to experiencing a VE through VR technology, however it does not necessarily require a VR like system. Similar to how we make the user feel like they are present inside a VE, we make the user feel like they are present in a real remote location in telepresence. For this purpose, telepresence robots are used [33] which are mobile robots capable of being controlled remotely and having the ability to stream your video to the remote screen and stream video from the robot s camera to your device. This technique uses a regular camera mounted on the robot streaming to a device present at the user s original location that can also be used to control the robot. If we replace the mounted camera with a 360-degree camera and the device with a VR headset, we get a VR telepresence system.

In VR telepresence, the illusion of being present at a remote location is achieved by utilizing the already well-established techniques of VR headsets. A 360-degree camera streams a video feed after performing appropriate stitching operations (see section 2.3.1) to a VE which projects it onto a sphere around the virtual cameras which renders the scene to the HMD. Now the user can see the remote location all around them, however, it s still not enough to create a strong Place Illusion (PI). Good spatial audio plays a crucial role as a vision in creating a strong PI. This effect is illustrated by

Jessica J. Baldis in a study on the effects of spatial audio on memory, focal assurance, perceived comprehension, and listener preferences [34]. Participants were made to listen to conference recordings with and without spatial audio. Results prove that having spatial audio improves all metrics which is further corroborated by participants preferring spatial audio over non-spatial one.

An example of a VR telepresence setup is presented in [35]. The authors make use of a commercially available 360-degree camera for streaming video feed and a full-duplex audio system for spatial audio to create a fully functioning VR telepresence experience.

#### *2.2.2. Current Research and Future Prospects*

At the core of telepresence is the idea of performing some remote operations. Therefore, there are myriad applications it can be applied to. The advancements in these systems go hand in hand with the advancements in the field of robotics. Mechanized robot parts are needed which can be controlled by the user to perform the remote activities. The best example of this process is in the field of Telemedicine, which is the delivery of health services remotely, such as remote surgery. There are times when people are not able to receive proper surgery due to a specialist being too far from their location and die. Thus, there has been extensive research done to improve this process and thus, save a lot of lives. For example, digitizing medical data and creating a 3d visualization inside virtual reality [36].

Other applications which can also save lives are disaster rescue, bomb disposals, exploring toxic atmospheres, and even hostage situations. Each of these fields requires a separate kind of robot built specifically for that purpose. For example, a drone might work in scouting from a high ground but might fail in an indoor situation. Leaving behind immediate dangers, teleoperation systems can also be used to prevent damages such as in surveillance and fire prevention systems [37].

Telepresence systems can also improve day-to-day life with applications in the tourism and entertainment industry. A person can visit a live concert or visit their dream place without leaving the comfort of their home. Work can also be made easier with Teleconferencing. With the current pandemic situation and work from home on the rise, teleconferencing systems can be quite auspicious to be back in the office environment without physically being there. For teleconferencing and similar social events, VR Telepresence is a capable candidate. Microsoft's Hololens is an MR technology and is primarily marketed towards teleconferencing operations. Another use case would be people visiting their relatives far away.

### 2.3. 360 Degree Cameras

#### *2.3.1. General*

The difference between a regular telepresence experience and a VR telepresence experience is the type of camera present at the remote location that performs the live streaming. In regular telepresence, a low field of view camera suffices for the required operation, however, in VR telepresence, we require a panoramic camera to be able to justify the use of VR. A panoramic image refers to an image that has a higher field of view than what the camera is capable of capturing. Panoramic cameras capture more than one image at the same time and combine them using various stitching techniques. A commonly used panoramic camera is the 360-degree camera capable of capturing 360 degrees of the surrounding, i.e. the entire sphere of view surrounding the camera.

Image stitching is the process of combining multiple images taken from different points but having overlapping segments to create a single panoramic image which feels like it was taken by a camera having a wide field of view. Since a single lens is incapable of capturing the image behind the lens, at least two lenses are required for creating a panoramic 360-degree image. A comprehensive survey of techniques used for achieving image and video stitching is presented in [38]. The most common commercially available 360-degree cameras make use of two fisheye lenses placed back to back capable of capturing images slightly wider than 180 degrees. Figure 2 shows one side of a 360 degree camera with fisheye lens. The other side of this camera is exactly the same with another fisheye lens. Having two lenses with wide field of view provides the necessary overlap required by the stitching algorithm to create a 360-degree panorama. Since there is only one of each lens, they are incapable of producing stereo images.

Figure 2. A Ricoh Theta V 360 degree Camera.

The stitched image is then converted into an equirectangular format by applying various image transformations. This type of image projection is well defined and is a standard across major manufacturers. Equirectangular projection is a projection technique in which a spherical image is converted into a rectangular image. This technique is commonly used in world maps. Figure 3 shows the map of the earth, which has a spherical shape, in equirectangular projection for viewing in 2D.

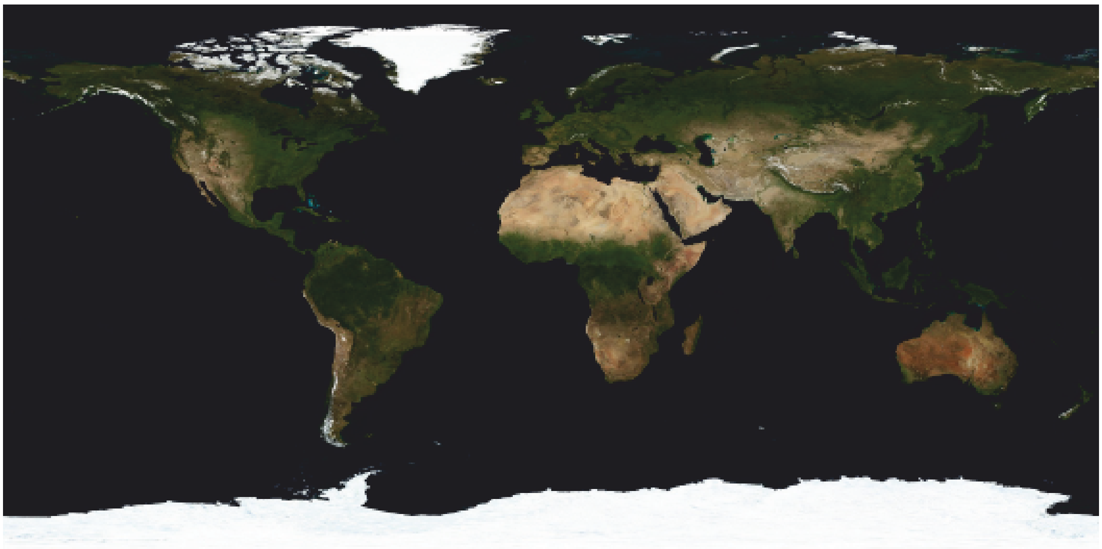

Figure 3. Equirectangular Projection of the Earth. Image under Pixabay License

A drawback of equirectangular projection is that objects near the top and bottom of the image are squashed and appear the wrong size. For example, in Figure 3, Greenland appears to be really small in size, however in reality that is not the case. This limitation does not apply to VR as the equirectangular image is projected back onto the sphere and the original scales are restored. The virtual camera inside the VE, which renders the scene onto the HMD, sits at the centre of the sphere and is surrounded by the 360-degree image or video surrounding them. This gives the illusion of being present at the location of the image or the video, which can be greatly improved if the image is stereoscopic.

#### *2.3.2. Limitations and Future Prospects*

One major drawback of panoramic images is the extra step required for stitching the images. This is ne if the video is not going to be live-streamed or if latency is acceptable in the live stream. However, in the case of telepresence, any sort of latency is not acceptable, especially in a teleoperation system where operations need to be performed in real-time. If stereoscopy is introduced, that would mean two images that need to be stitched instead of one and thus, twice the time required. Therefore, in commercially available cameras, there is an option to reduce video resolution to enable latency-free live streaming.

Research and developments are currently in progress to overcome this limitation. The most obvious solution to this problem is a fast CPU to stitch the images faster. Even on today s hardware, this limitation can be overcome by improving upon the stitching algorithm used. [39] proposes a fast stitching algorithm that can reduce the time spent on that operation by a lot and enable live high-quality video streaming with low latency. Insta360 even managed to live stream an NFL match in 8k using Verizon s

5G technology with low enough latency to not notice any lag between the actual event and the video stream [40].

## 2.4. Body Tracking

### *2.4.1. General and History*

Body tracking, more commonly known as motion capture technology is generally used in computer animations to give realistic movements to the character. The roots of motion capture can be traced back to the invention of rotoscoping by Max Fleischer for which he got a patent in 1917 [41]. In rotoscoping, a live action footage was recorded and then animation frames were drawn frame by frame by tracing over the live action frames. As the computing power increased and new graphics technologies emerged, modern motion capture techniques were introduced in the 1980s. It was rst used in medical eld when Calvert et al. attached potentiometers to a body which was mirrored in a computer animation [42]. This animation was then used for clinical assessment of any movement abnormalities. Before long, a system called "Graphical marionette" was developed at the MIT labs [43] which introduced the idea of full-body suits for motion capture that is still a popular technique to this day. LED lights were attached to an actor which was tracked by cameras and reconstructed in the form of 2D animations.

There are two main methods for body tracking, marker-based, and markerless motion capture. With marker-based techniques, an actor usually wears a full-body suit with markers attached that are tracked to create an animation that mirrors the actor s actions. This setup is time-consuming and requires a large space to set up the tracking cameras, thus it is mainly used by large movie studios for creating Computer Generated Imagery (CGI) in live-action movies or used entirely for fully animated movies. However, this technique produces quite accurate and realistic animations and with advancements in computer vision techniques, it can be achieved with far fewer cameras than what it used to require, thus reducing the required space.

Small scale marker-based techniques are also available for general consumers thanks to gadgets like the HTC Vive trackers [44]. The VIVE trackers do not require an intricate body suit, but only a few trackers, at the cost of relying on IK techniques for solving the movements for rest of the body, hence they are less accurate. The trackers can be strapped onto the user s waist and both feet to create a six-point tracking system. The six points refer to 6 items being tracked, the HMD tracking the user s head, the two controllers tracking the hands, the waist tracker for tracking the user s waist, and the two trackers on the feet for tracking the same. This technique makes use of Inverse Kinematics (IK) to generate the animation therefore, we can also use only one tracker on the waist and let the IK algorithm gure out the leg movement, or remove the trackers altogether and let the IK algorithm gure out the entire body movement from just the HMD and the two controllers, known as a three-point tracking.

The IK technique can not only be used for generating animations but can also be used to control an avatar in real-time for use inside a VE. Here, an avatar refers to a virtual body that the user can control as if it was their own body. Avatars are generally used in a VE where having a body improves the overall experience, such as VR games with social interactions.

While marker-based techniques have dominated the industry since the inception of body tracking, recently there have been techniques for motion capture without using any markers, called markerless tracking. Instead of having physical markers on the user, depth cameras make use of infrared light to project invisible markers onto the user and track the user s movement by detecting any changes in the distance of the infrared markers. The most commonly used device with this capability is the Microsoft Kinect. It was originally released as an accessory to Microsoft s Xbox 360 game console, and later Kinect v2 for the Xbox One console, to enable the ability to control game elements by making use of hand movements and gestures. However, this idea did not thrive as it required developers to add special functionality in the games to be able to make use of the gestures, and thus, many consumers only saw the Kinect as a gimmick. It, however, caught the attention of researchers for its cheap price and really accurate body tracking capabilities, and thus the Kinect v2 became really popular in researches requiring a depth camera.

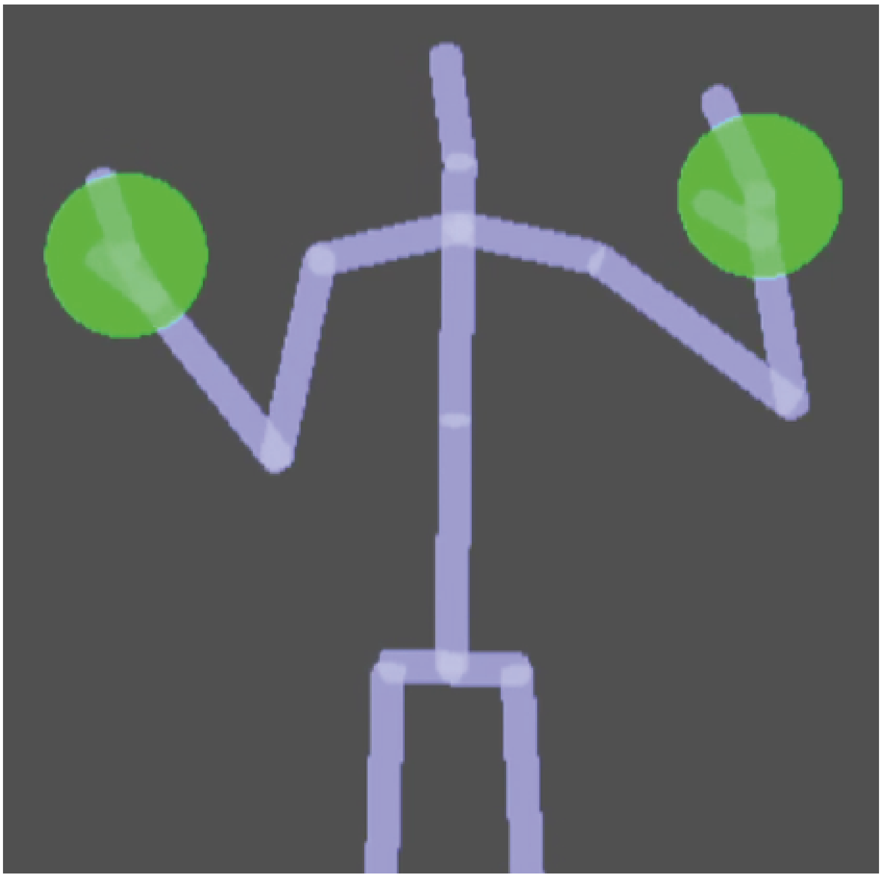

Figure 4. Body tracking using Microsoft Kinect v2.

Figure 4 shows how the Kinect creates a humanoid skeleton by tracking the user using the depth camera. It maps the depth data onto a virtual humanoid skeleton whose motion can then be mapped to animations or real-time avatars.

With the current advancements in computer vision and deep learning methods for identifying trackable landmarks on a person, even the depth camera can be disregarded for markerless tracking. [45] demonstrates a convolutional neural network capable of tracking motion in a regular video. This means that even by using a regular camera, we can perform markerless motion capture. Motion capture from video cameras can be improved by attaching some markers on the person which stand out in the video. A full-body suit similar to marker-based techniques is not necessary. This is demonstrated by Zult et al. in their study [46] where they compare data of 3D gait analysis from a 3D motion capture system and a simple 2D video camera. The results indicate that the video camera produces analysis results similar to a high-cost 3D motion capture system. Instead of markers, different coloured clothes that stand out from the background, along with different coloured hands and feet are also capable of producing accurate tracking [47]. There are even methods for body tracking using Hidden Markov Models [48]. A comprehensive literature review of various methods for body tracking and their popularity is presented in [49].

### *2.4.2. Avatar Embodiment*

Having an avatar, also known as a virtual body, inside a VE can add a new layer of realism to the VR system. Skarbez et al. performed a study [6] to measure the Psi and PI proposed by Mel Slater for evaluating a user s immersion inside a VE [1]. They measured which aspect of a VE had an impact on Psi the most, look and feel of the environment, physical behaviour of objects, or improved body tracking. Results show that for most participants, having a virtual body impacted Psi the most.

Avatar Embodiment can be achieved using various methods discussed in Section 2.4.1, including the simplest three-point tracking using just the HMD and the controllers and letting a good IK algorithm calculate the rest of the body movement. The Microsoft Kinect is also a popular method for low-cost avatar embodiment when the study involves a lot of movement and solving for IK is not accurate enough. There are even solutions available for tracking individual ngers, such as Leap Motion Controllers, or the built-in nger tracking in the Oculus Quest 2. [50] demonstrates a system for creating a body using a combination of Oculus Rift for HMD, Microsoft Kinect for body tracking and Leap Motion for nger tracking.

Having an illusion of virtual body ownership can cause psychological reactions. The most famous example of this is the rubber hand illusion [51] which demonstrates the illusion of ownership of a rubber hand. Slater et al. demonstrate this effect on a virtual arm inside a VE in their study [52] instead of a physical rubber arm. There is even a study to reduce people s fear of death through a virtual Out-of-Body experience [53].

Virtual bodies can greatly enhance the education eld as well. A study by Mina C. Johnson-Glenberg demonstrates the use of the virtual body in education [54]. The study demonstrates that hand gestures facilitate learning. Therefore, having a virtual body can have a similar effect in remote teaching as in physical teaching.

## 2.5. Motivation for Current Study

The section 2.2.1 establishes the eld of telepresence and the Section 2.4.2 demonstrates the bene ts of Avatar Embodiment inside VR. This study combines these two fields and probes the aspect of improving VR Telepresence using Avatar Embodiment. The user wears the HMD displaying a 360-degree live stream from the camera present at a remote location and embodies a virtual avatar that mirrors their hand movements. The term Self-Avatar is used here to describe the virtual body the user is currently embodying.

Previous studies on Avatar Embodiment have focused solely on improving a VE. This study aims to study the effect of having a virtual body in a real location. This opens a future of having teleoperation systems with robotic arms which the user can control in real-time as envisioned by Minsky when he coined the term Telepresence [5].

The telepresence system is achieved by live-streaming from a 360-degree camera as demonstrated in Section 2.3.1. To avoid VR sickness discussed in Section 2.1.4, the camera is stationary and the user is seated on a chair. The body tracking is achieved using a three-point tracking IK system which was demonstrated in Section 2.4.1.

# 3. METHODOLOGY

This chapter explains the tools used for designing the VR environment that enabled the user to experience the telepresence scenario while embodying a self-avatar. Data points collected during the user study are also explained here.

## 3.1. Development Tools

The software used for the telepresence experience is created using the Unity Game Engine version 2020.3.1f1 LTS (Long Term Support), which has a built-in official package for calculating IK called "Animation Rigging" [55]. Since the user's body motions are tracked using a three-point tracking system, IK is needed to calculate the motion of the rest of the body parts. Another popular package available for solving IK in Unity is called "Final IK" [56]. Final IK is available as a paid asset and provides advanced features and, at the time of this writing, is overall better than the in-built animation rigging. However, in this study the subject, and hence the virtual body have restricted movements, therefore the freely available animation rigging package is ample for the job.

For the VR system, Oculus Rift S is used as it boasts a minimalist setup and the least number of accessories for it to work, an HMD, and one controller per hand. Moreover, the controllers also provide capacitive touch functionality for a rough tracking of fingers. The HMD tracks the head movement of the virtual body, while the controllers track the position and orientation of the left and right wrist respectively. This creates a three-point tracking system for the IK algorithm to solve. The wrists solve for the position and orientation of the elbow and shoulder joints, while the head determines the position and orientation of the chest. Figure 5 shows the body being tracked along with simple finger animations controlled by the controllers.

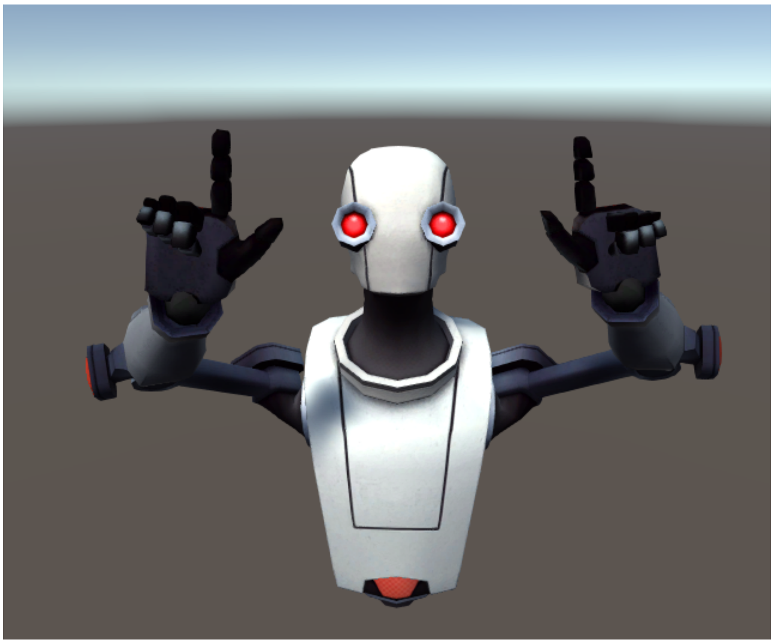

Figure 5. The Self-Avatar that the user embodies.

Lastly, for the telepresence setup, a Ricoh Theta Z1 360 camera is used. Ricoh 360 cameras are the only cameras commercially available that provide a stitched equirectangular (see Section 2.3.1) video feed directly as a system webcam with minimal delay. A caveat for using this camera is that the port for tethering it to the PC is located at the bottom and most tripods have a broad base to support big cameras and thus, it obscures the port making the camera impossible to use. To overcome this challenge, a special monopod is used, namely "Bushman Monopod V2" which has a narrow base to allow room for the wire to be attached to the base of the camera. Figure 6 shows the camera mounted on top of the monopod.

Figure 6. Camera mounted on the stand.

Since it was crucial for the study that the user has a real-time conversation with the confederate, the streaming quality was dropped to FullHD resolution to reduce latency to a level not noticeable by the user. The confederate is part of the research group and is responsible for having a conversation with the user.

### 3.2. Study Setup

This study aims to study the effect of self-avatar in a telepresence environment and not introduce novel methods for latency-free communication between devices in the system. Hence, all connections will be wired to minimize any possible latency. Thus, the camera was chosen to be placed in close proximity to the user s actual location which reduces latency in the video feed as well as provides directional audio to the user without relying on a 3D audio technique.

From Figure 5 we can also observe that the body is that of a robot. Having a realistic human appearance may create a bias among people of different gender, race, etc. [57] demonstrates that machine-like and cartoon-like avatars had a stronger ownership illusion when compared to realistic human avatars. We also observe that the avatar s legs are hidden and only the torso along with the head and hands are visible. Pan et al., in their study, [58] concluded that the introduction of foot tracking had no signi cant effect on the overall result unless they analyzed the participants movement behaviour. In this study, the participants will not be moving, hence the introduction of foot tracking is not necessary.

Keskinen et al. (2019) in their study [59] observed that the user s own height with respect to the camera height had no effect on the viewing experience and concluded that a height of 150 cm was the sweet spot for the camera mount irrespective of the user s actual height. However, during the pilot studies, participants felt too high at 150 cm, while sitting. Ultimately, the camera height was set to be equal to the eye height of the confederate when seated, 117 cm. Therefore, in this study, the user would be seated and the camera would be mounted at a height of 117 cm from the ground.

Certain behavioural metrics were recorded during the study. To stimulate the user, two events take place during the sessions. First, an object is thrown towards the camera abruptly. Second, the confederate walks towards the camera which might cause reactions from users. This is demonstrated in [59], where it is shown that the users were only comfortable when other people were at least 1 meter away from the camera.

Having 3D audio from the confederate is also essential to enhance the experience [60]. Since this study does not focus on virtual audio techniques and since the user would be sitting for the entire duration of the study, real-life audio can be substituted for 3D audio techniques. This is achieved by seating the user directly behind the camera and instructing them not to turn their head all the way back for the duration of the study (See Figure 8). This way the user hears the audio directly from the source in a natural fashion. There could be some delay with the lip movement in the video and the audio, but that was resolved with the use of a mask covering the confederate s face (See Figure 7). The area where the user is seated is concealed from the area where the camera is situated by a curtain.

Three crucial facets for achieving a sense of telepresence are presented in [31]:

- The user should receive the same sense of sensory information that you expect them to receive had they been present physically at the remote location.
- The user should be able to move the sensing devices (for eg, the HMD and controllers with sensors for capturing head and hand motion) around in the remote location.
- The user should be able to modify and interact with the remote location.

In this study, the user receives accurate visual and audio information with negligible latency. The user is able to move around their virtual hands in one part of the study and therefore, they are able to move the sensing device around. Finally, the user is able to have a live conversation with the confederate and thus, interact with the remote environment. Therefore, the study takes into account all the points mentioned above.

### 3.3. User Study

Before the user arrives, their ID is determined. The study is performed in a counterbalanced scenario where the script is broken into two parts, one is experienced with a body and one without a visible body, in alternate orders for each user. For example, if the first user got scenario A (without body) as their first scenario, then the next scenario will be scenario B (with the body). Then the second user would get scenario B as their first scenario and scenario A as their second. The ID has the first scenario encoded in name, for example, 01A, 02B, 03A, etc., the last letters A and B determining their first scenarios. Then the counterbalancing sheet, questionnaires, and consent forms are pre-filled with the researcher's details and the user ID.

After the user arrives, they are first asked to sit on a chair that is fixed at a predetermined location. After several trial and errors, the chair location was fixed so as to avoid hitting any obstacles with the hands while moving them around inside the VR. They are then asked for their consent on having their video recorded. After reading out the instructions, explaining the controls, and letting the user put on the VR headset, the first scenario is played. If the user has a body, they are instructed to practice the thumbs up and pointing gestures at least 3 times. When they are ready to start, the confederate is signalled to start the process. After the confederate enters the room, the curtains which separate the user from the location of the camera, are slightly moved without the user's knowledge to allow the confederate to see the user's gestures.

Figure 7 shows the confederate sitting at their assigned location. They give a live tour of the equipment lab to the user and explain various equipment to them. The confederate has a script with them which they use to act out their part in the conversation (See Appendix 6 for the script). The script is designed in such a way that it does not rely on the user's responses to progress. Also in the script, the user is asked to point out various objects in the lab so as to make the conversation interactive and to encourage the user into using their hands more. At the end of the first scenario, the confederate picks up a dummy rat and pretends that the rat bit them. The rat is then thrown towards the camera for stimulating a reaction from the user. Similarly, at the end of the second scenario, the confederate walks towards the camera for achieving the same goal of attempting to break the presence. After the confederate leaves, the curtains are drawn before asking the user to remove the headset. Figure 9 shows the user inside the session and his video being recorded from two angles. The angle from the back also shows the curtains that separate the user from the supposed remote location.

Figure 7. The confederate sitting at their assigned position.

Figure 8. The User.

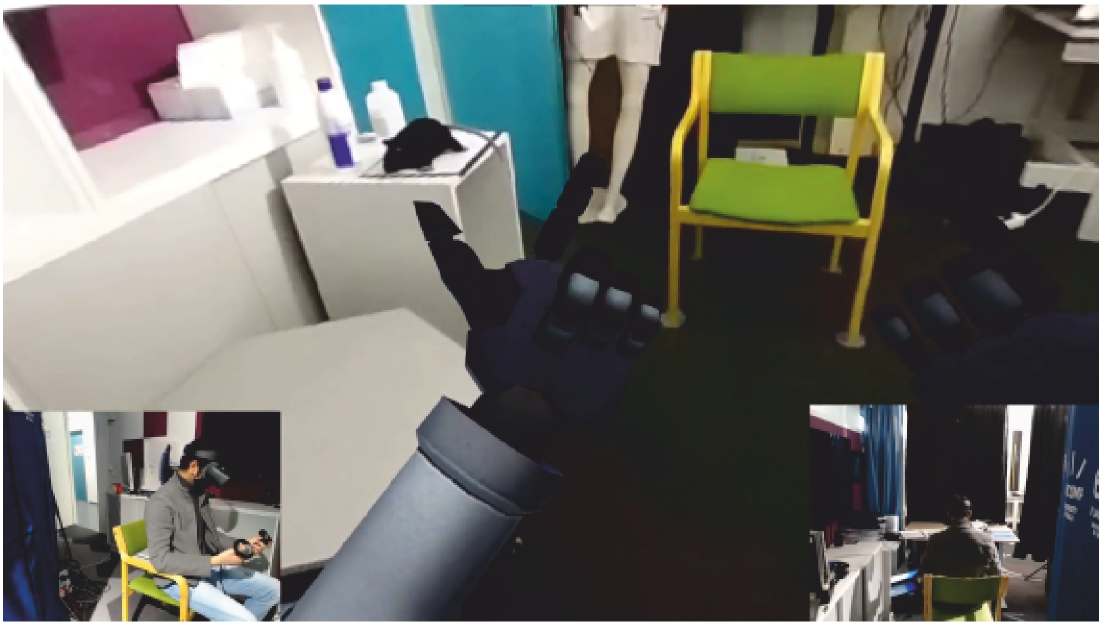

Figure 9. The User inside the session.

Figure 8 shows the user when the scenario is undergoing. They are instructed to hold the controllers regardless of the scenario to avoid creating a bias due to not holding the controllers in one scenario. They are also instructed to not look to their back so as to avoid having them accidentally see themselves when the curtains are open. If they consented to the video recording, they are also recorded from two angles, a side view, and a back view. These recordings are for recording their reactions to events in the script where the confederate tries to stimulate the user s response by throwing stuff at them or walking towards them. After each scenario, they are rst asked to draw and annotate a graph based on their experience (Figure 10). The graph describes the variation in the user s sense of location, whether they felt sitting with the confederate or at their original location (y-axis) with respect to the duration of the session (x-axis). Then they are asked to ll out a questionnaire with appropriate questions regarding their experience. After both scenarios are done, they are presented with a post-experiment questionnaire that asks for their demographics and overall opinions.

Aside from video data, the position coordinates and Euler angles of the HMD and both controllers in Unity s world coordinate system are also recorded for every frame. Figure 11 show the raw position and rotation data of the HMD and the two controllers plotted using a line graph.

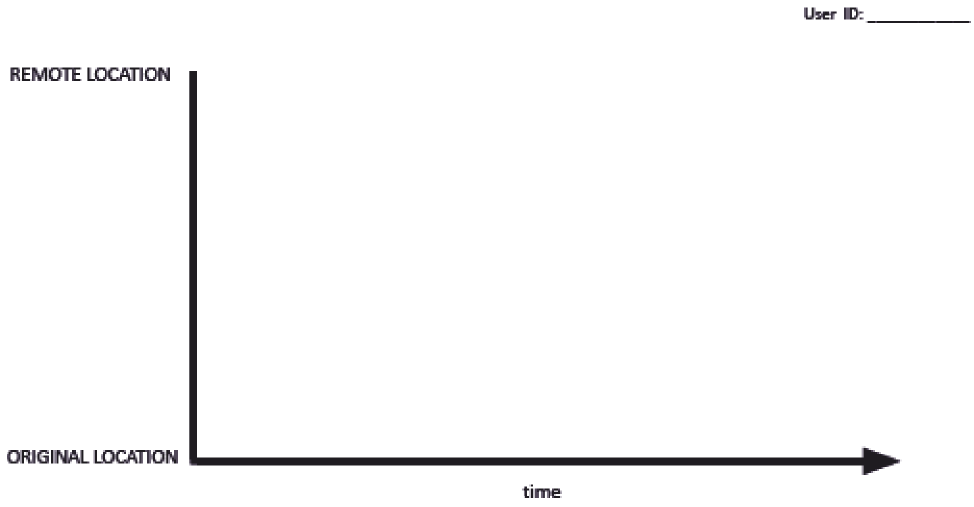


Figure 10. Presence Graph.

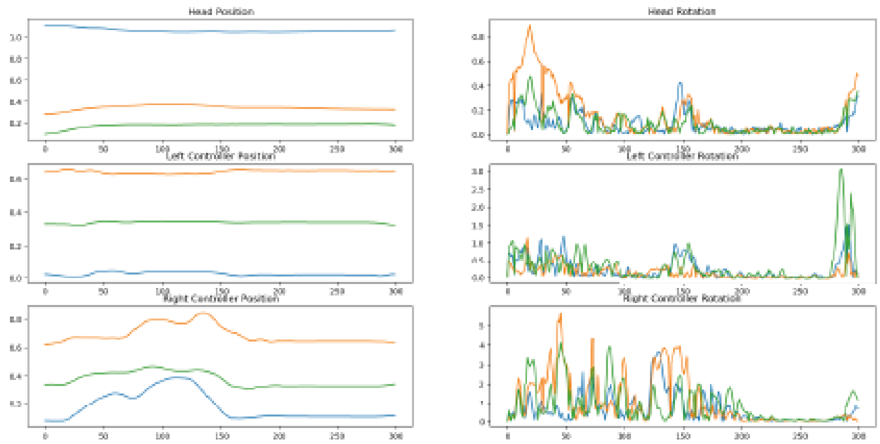


Figure 11. Raw position and rotation data.

## 3.4. Questionnaires

There were three sets of questionnaires presented to the user. Two after finishing each of the scenarios, and one post-experiment questionnaire after both scenarios are over and their respective questionnaires have been submitted. Google forms service was used to design and present the questionnaires to the users. The quantitative questions were designed using a Likert scale [61] with a range of 1 to 7 as it is the most popular scale used in other studies as well, where 1 represents strong disagreement and 7 represents strong agreement with the statement provided in the question.

The questionnaires after the scenarios were usually divided into three or four sections, depending on the scenario. The first section determines the user's sense of presence (See Appendix 1 for the full questionnaire). This questionnaire is based on the work of Slater and his colleagues and is commonly known as Slater-Usoh-Steed (SUS) questionnaire as described in [62]. There are 6 questions revolving around 3 main themes:

- The sense of being inside the VE.
- The degree at which the VE becomes the dominant reality.
- The degree at which the VE is remembered as a place.

The questionnaire used has been modified slightly for measuring the sense of presence in a remote real environment rather than a VE. The SUS questionnaire has been a standard in a number of studies that measure PI and Psi (For e.g. [6]). As described already, the SUS questionnaire also uses the Likert scale in the range of 1 to 7.

The second section of the questionnaire was used for measuring the Co-Presence with the confederate (See Appendix 2 for the full questionnaire). The questionnaire is a modified version of the Co-Presence questionnaire presented by Casanueva et. al. in their study [12]. They conduct a study in which participants interact with each other, however that is not the case in this study, where the participants interact with the confederate. Hence, the questions are modified to fit this study.

The third section of the questionnaire collected qualitative data on any sort of break in the user's presence (See Appendix 3 for the full questionnaire). According to the script, the confederate is supposed to throw an object towards the camera and later approach the camera to pick it up. This is done intentionally to break the user's sense of presence and the questionnaire exists to get their experience during the stimulus and any sort of challenge they faced while recovering from the stimulus.

The final section is exclusive to the scenario where the user has a visible body. This questionnaire measures the sense of embodiment of the virtual avatar (See Appendix 4 for the full questionnaire). The questionnaire is presented by Gonzalez-Franco M. and Peck T. C. in their study [63]. They presented a standard questionnaire for most of the embodiment related studies which were conducted or can be conducted in the future. Out of all the questions present, a few were selected that matched with the current study and an extra question was added which asked the user whether they actually felt like they had embodied a robot (since the avatar used is a robot's body).

After all these questionnaires have been filled in both scenarios, the user is presented with a post-experiment questionnaire (See Appendix 5 for the full questionnaire). This questionnaire surveyed the user's preference out of the two scenarios (with the body or without one) and their demographic information such as gender, age, video gaming history, history with VR, etc.

# 4. RESULTS AND DISCUSSION

This study involved 20 participants and the following data is collected from them:

- Video and Audio Data including screen recording of the experience
- Position and Rotation of HMD and two controllers in the 3D space
- Graphs depicting the variations in their sense of presence throughout the experience (See Figure 10)
- Various Questionnaires

The questionnaire results are analyzed using the charts generated automatically by google forms and using the SPSS software for advanced analysis. The coordinates data is analyzed using the Pandas library of the Python programming language.

## 4.1. Questionnaires

### *4.1.1. Post Experiment Questionnaire*

This section surveyed the user s preference out of the two scenarios which they experienced and their demographics. The rst question asked which of the two scenarios they liked the best. 18 out of 20 participants preferred the experience with the body. The open-ended question that followed this question asked them why they preferred one scenario over the other. Some popular responses on their preference of having a virtual body are:

- They were able to point at objects when the confederate asked them to.
- They felt ownership of the robot s hands.
- They felt weird when they didn t have hands and were told to point.
- They were able to use gestures when talking with the confederate.
- Thought it was cool.

The users who preferred the experience without the body thought that the robot hands did not look like their hands or they were in the way.

Next, they were asked if adding the virtual body had any improvement in their experience. 16 out of 20 participants felt that the addition of a virtual body improved their experience of talking with the confederate. The reason for improvements, as answered by the users, was mainly due to the ability to use body language in addition to the speech. This effect is noted by [54] where the addition of hand gestures had a positive effect on learning.

Further questions were on user s demographics. The results are summarized below:

- 65% (13) of them were male and 35% (7) were female.
- Majority of the participants were in the age group of 24-29.
- Majority of the participants (55%) have never played video games or play once or twice in a year.
- Majority of the participants (75%) have used VR at least once in their lifetime.

### *4.1.2. Presence Questionnaire*

The SUS questionnaire (refer to section 3.4 and Appendix 1) was used for measuring the PI experienced by the user in both scenarios and then compared which of the two scenarios yielded more sense of presence. All questions in this section were quantitative in nature with responses in the form of a Likert scale in the range of 1 to 7. As instructed in [62], a score of 6 or 7 is considered to be 1 and anything less than 6 is considered to be a 0. In the subsequent diagrams and tables, "A" refers to the scenario without a body and "B" refers to the scenario with a visible body.

Table 1. SUS Data Description

| | SUS A | SUS B |
|---|---|---|
| Mean | 3.70 | 3.15 |
| Median | 4.00 | 3.50 |
| Std. Deviation | 1.838 | 2.254 |
| Skewness | -0.863 | -0.084 |
| Kurtosis | -0.242 | -1.397 |

As seen from Table 1, both the scenarios are negatively skewed. The skewness of scenario B is closer to 0 than the skewness of scenario A, which indicate a normal distribution in B. Having a normal distribution implies that all values are close to the mean value. However, looking at the Kurtosis value, we nd that B is more negative than A which indicate extreme values in data and possible outliers. This effect can be seen when we plot the box plots of individual questions in the questionnaire as seen in Figure 12. This plot shows a higher mean value in the B plot, however, accompanied by a great number of outliers.

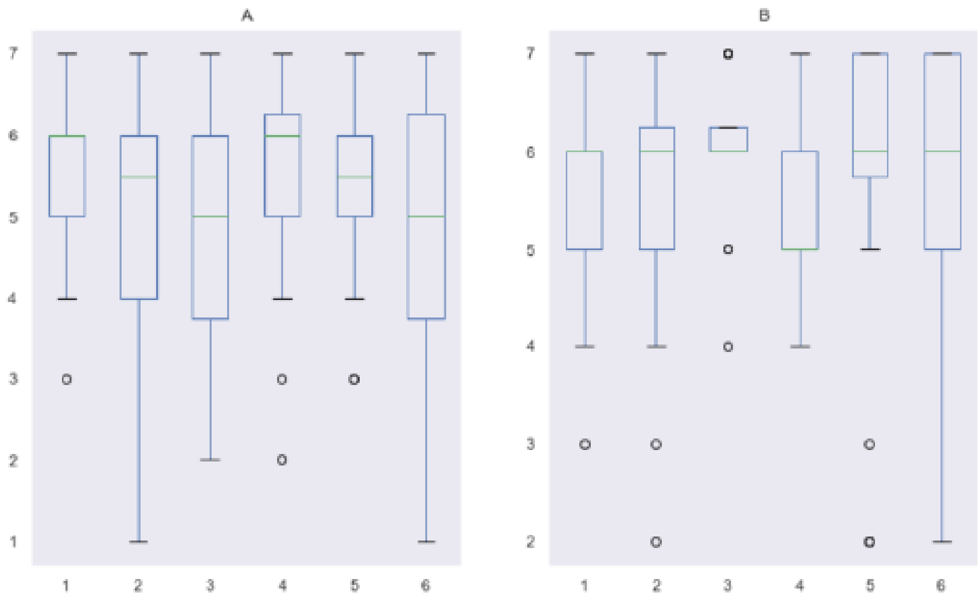


Figure 12. SUS Box Plots (x-axis: question number, y-axis: user s score).

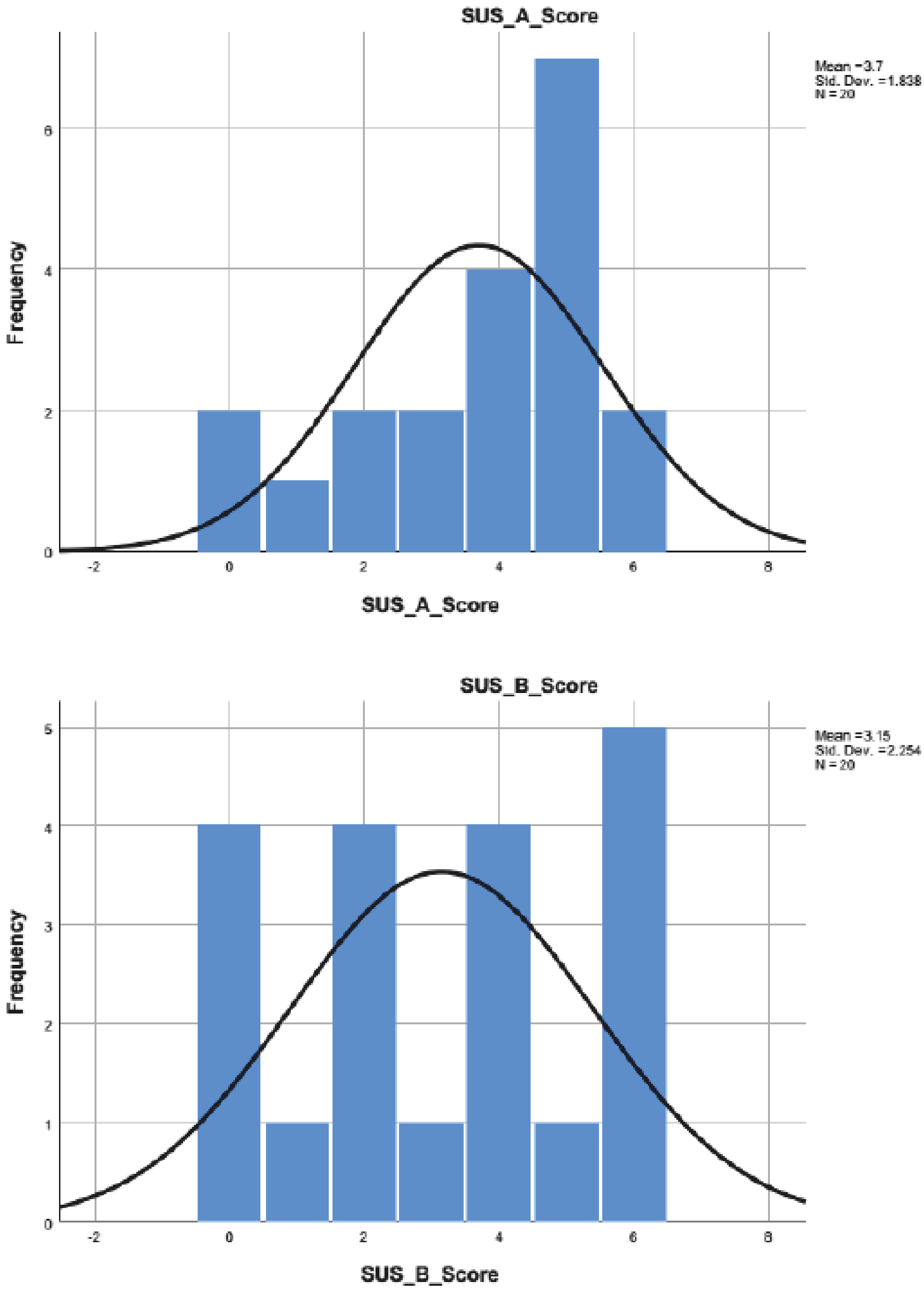


Figure 13. SUS Data Histogram.

The histogram in Figure 13 also shows the extreme values in the B condition. These data points conclude that there was an extremely varied sense of presence in the scenario with the body. While a higher number of people experienced a boost in PI with the addition of a virtual body, a signi cant number of people experienced a decline in PI compared to the scenario without a body. To explore this further, analysis is done on whether the scenario order had any effect on these results. The data is divided into four groups, scenario A data of users that began with scenario A, scenario A data of users that began with scenario B, scenario B data of users that began with scenario A, and scenario B data of users that began with scenario B. For this, the samples are considered independent of each other and Mann-Whitney U Test is performed on them.

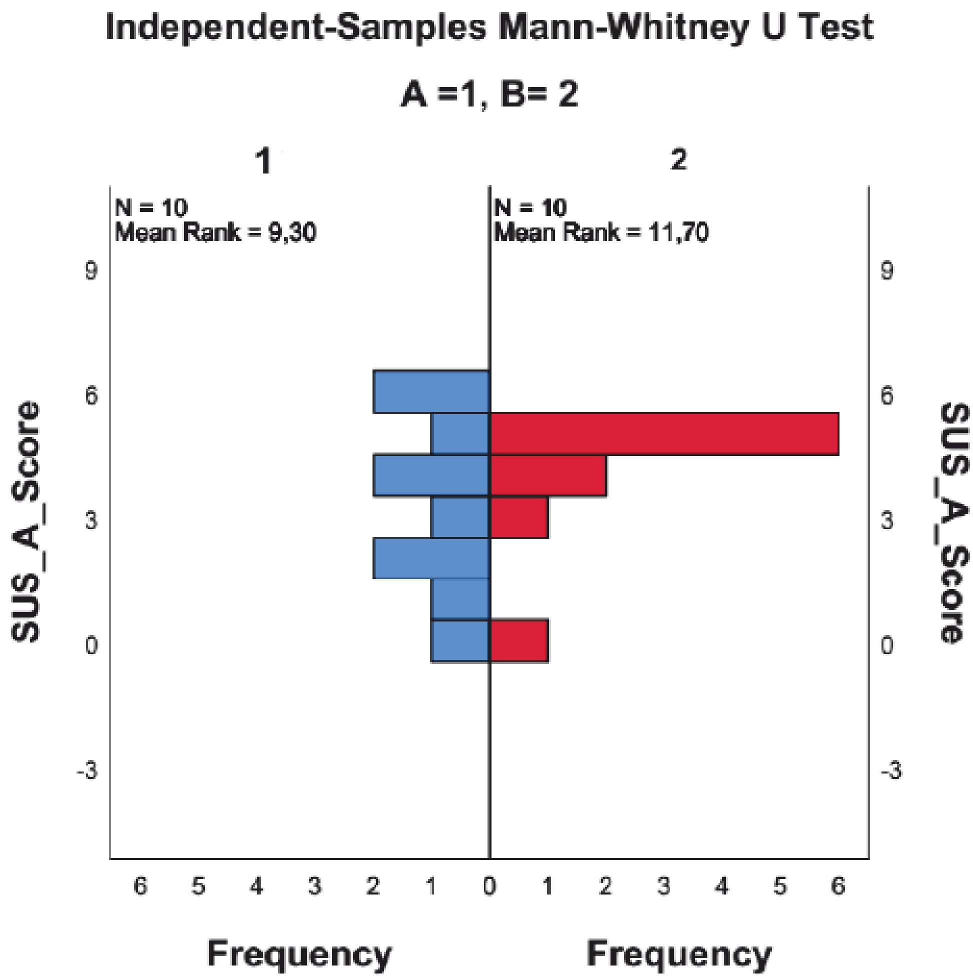


Figure 14. SUS A with A first (left) and B first (right).

In Figure 14, the SUS score for scenario A is shown for users with scenario A as their first scenario on the left, and users with scenario B as their first scenario on the right. Similarly, Figure 15 shows the SUS score for scenario B for users with scenario A as their first scenario on the left and scenario B as their first scenario on the right. These plots show that the users had a higher PI in their second scenario regardless of whether they started with a virtual body or without one. Half of the users that started with a virtual body did not get enough time to feel present in the remote environment, hence the outliers in the scenario B data.

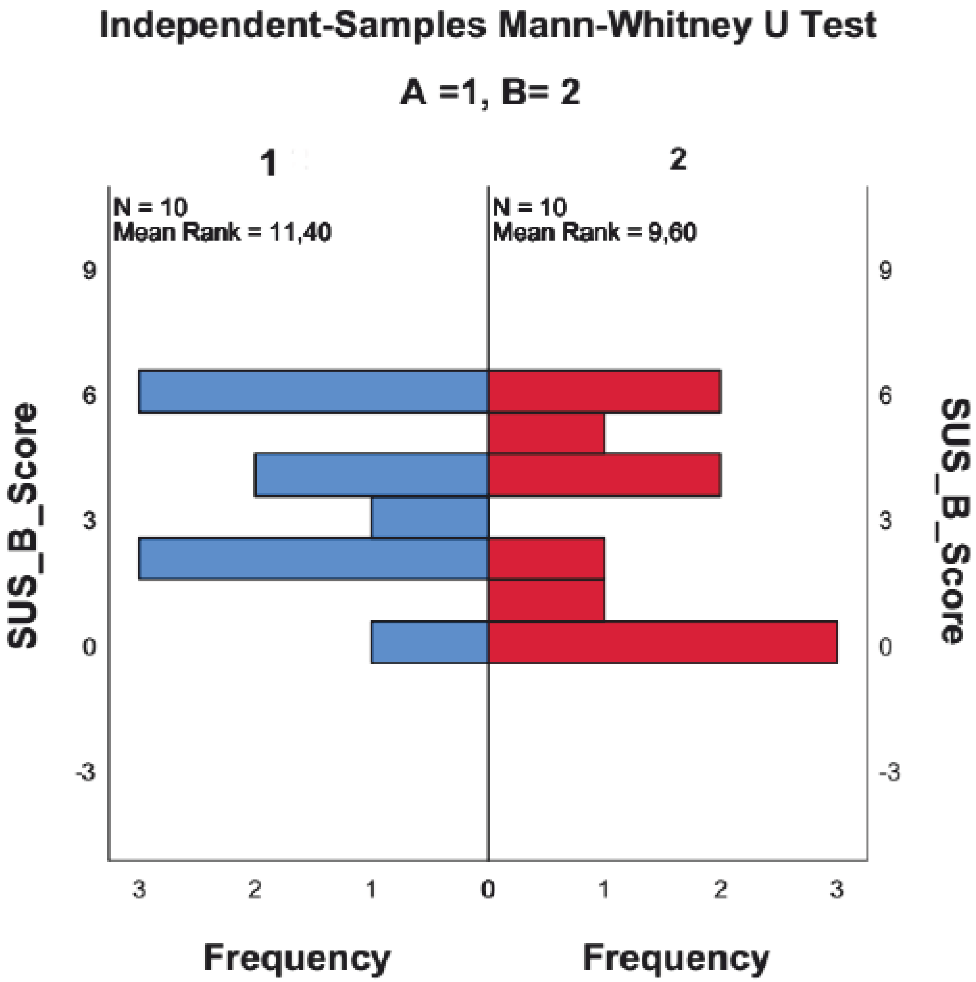


Figure 15. SUS B with A rst (left) and B rst (right).

### *4.1.3. Co-Presence Questionnaire*

This questionnaire determined the sense of social presence the user felt in the remote environment when they were interacting with the confederate. The questions in this section were also quantitative in nature with response in the form of a Likert scale from 1 to 7. Figure 16 and Figure 17 shows the histogram of responses of scenario A and B respectively.

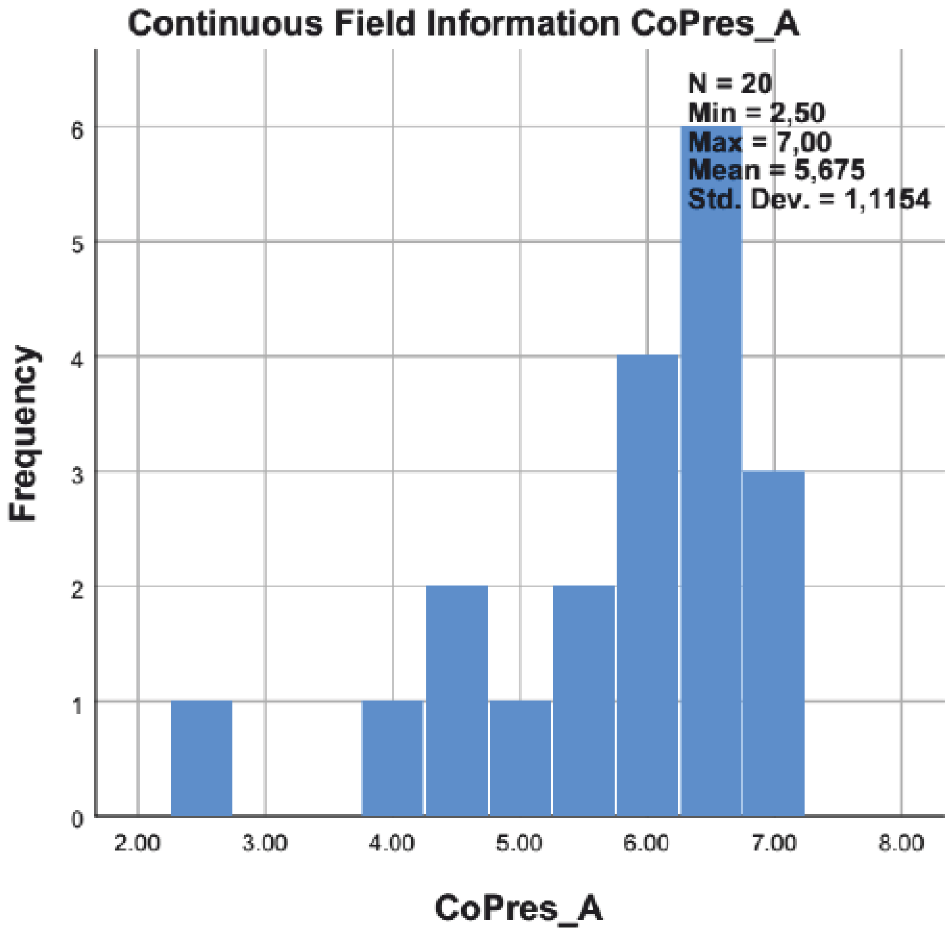


Figure 16. Co-Presence histogram of scenario A.

To analyze this section, we perform a Wilcoxon Signed Ranks Test to determine whether there is any difference in average values between the scenarios. Table 2 shows that the asymptotic signi cance is larger than 0.05, thus strongly suggesting retention of the null hypothesis. This shows that there was either little difference in the Co-Presence experience of the users, or the sample size was too small to make any signi cant conclusions.

Table 2. Co-Presence Wilcoxon Signed Ranks Test Analysis

| Null Hypothesis | Test | Asymtotic Sig. (2-sided test) | Decision |
|---|---|---|---|
| The median of differences between Co-Presence A and Co-Presence B equals 0. | Related-Samples Wilcoxon Signed Rank Test | 0.336 | Retain the null hypothesis |

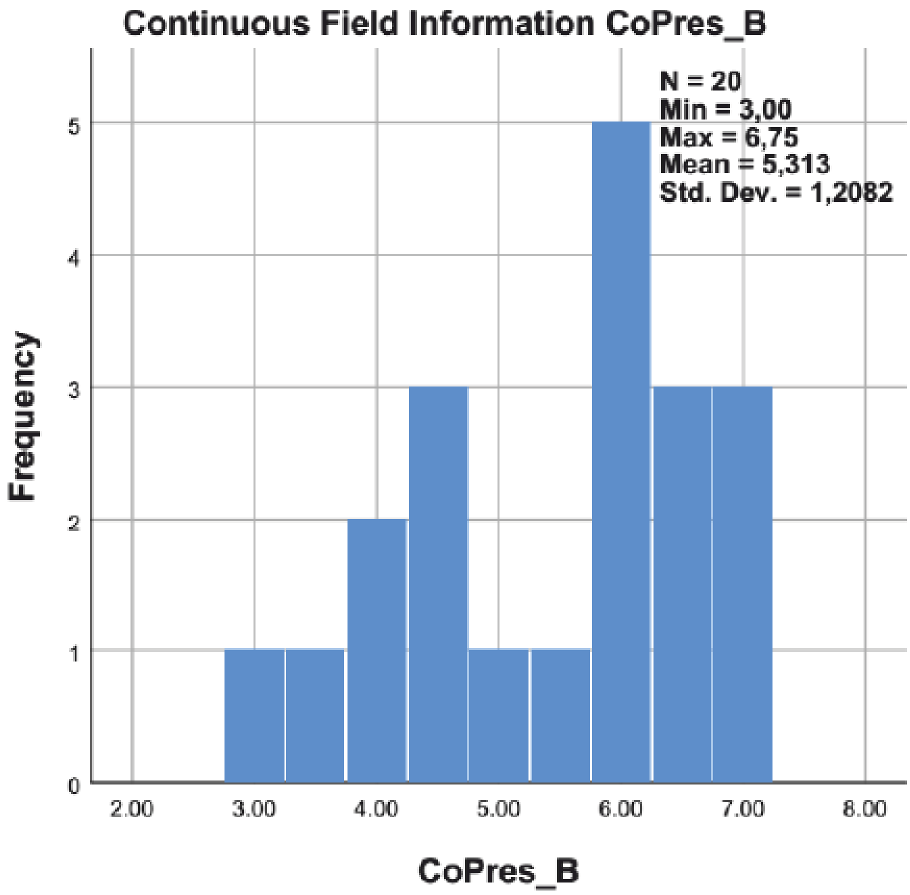


Figure 17. Co-Presence histogram of scenario B.

The results from Mann-Whitney U Test in Section 4.1.2 and the Wilcoxon Signed Ranked Test in this section indicate that the sample size of 20 participants was too small to draw any meaningful conclusions. Therefore, this work will serve as a pilot for a more comprehensive study with bigger sample size.

#### *4.1.4. Virtual Body Questionnaire*

This questionnaire was exclusive to scenario B (with a virtual body) and measured the sense of embodiment the user felt towards an avatar. [63] proposed that the total embodiment is the function of 6 variables pertaining to the Avatar Ownership, Agency, Tactile sensations, Location, Appearance, and Response. The total embodiment is calculated as follows:

Total Embodiment = ((Ownership/5)*2 + (Agency/4)*2 + (Tactile Sensations/4) + (Location/3)*2 + (Appearance/4) + (Response/5)) / 9

However, in this study only a few questions were selected out of the many questions provided in the paper as most of them did not t in a telepresence scenario (for example, Tactile Sensations). Since there were only a few questions, we can pick the last question, "I actually felt like I embodied the robot" as the measure among participants.

Table 3. Robot Avatar Embodiment

| | I actually felt like I embodied the robot |
|---|---|
| count | 20 |
| mean | 5.35 |
| std | 1.38 |
| min | 2.00 |
| 25% | 4.75 |
| 50% | 6.00 |
| 75% | 6.00 |
| max | 7.00 |

The results shown in Table 3 show that more than 50% of the participants gave a score of at least 6 to the presented question. This score is also evident from the results of the rst two questions in the post-experiment questionnaire (See Section 4.1.1). Therefore, we can safely conclude that majority of users felt ownership of the virtual avatar.

#### *4.1.5. Break in Presence Questionnaire*

This section contained qualitative questions that asked for any moment in the scene which caused a break in the user s sense of presence from the remote environment. These questions were added in response to the intentional attempts by the confederate to break the user s presence by throwing an object towards the camera and by physically approaching the camera. This was done to stimulate a response from the user which is recorded by the video and coordinates data.

According to the Break in Presence Questionnaire responses and the variation in presence graphs, the majority of users felt a break in presence when the object was thrown towards them, however, fewer users felt a break in presence when the confederate approached the camera. The approach towards the camera might be due to cultural differences from my personal observations and could be a suitable addition to the demographics of future study.

### 4.2. Variations in Presence Graph

The users were presented with a graph (shown in Figure 10) after each scenario to measure their variations in presence with respect to the duration of the session. Based on the responses, the users can be categorized into 4 main categories:

- Overall Increasing Presence: These users had a low sense of presence at the beginning which increased by the end.

- Overall Decreasing Presence: These users had a high sense of presence at the beginning which decreased by the end.
- Stable Presence: These users had an almost similar presence throughout the experience.
- Irregular Presence: These users had a somewhat irregular presence throughout the experience.

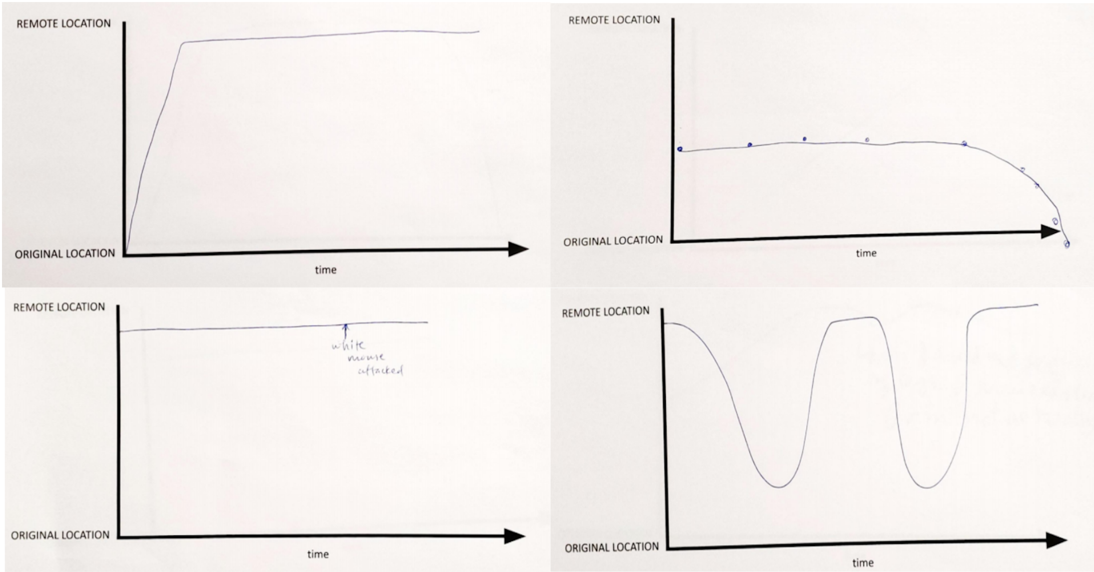


Figure 18. Graph examples of different categories (Top Left: Overall Increasing Presence, Top Right: Overall Decreasing Presence, Bottom Left: Stable Presence, Bottom Right: Irregular Presence).

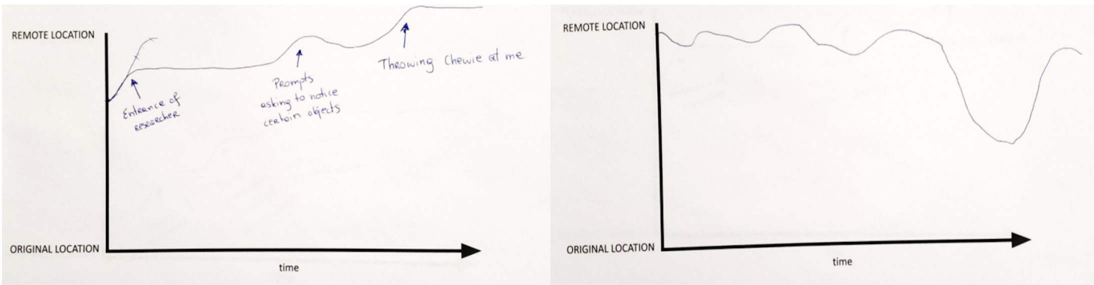


Figure 19. Different variations in response to object being thrown towards them.

Figure 18 shows an example of the aforementioned categories. The graphs were used as an auxiliary to the Break in Presence Questionnaire (Section 4.1.5). Analysis of these graphs presented a fascinating data point. While some of the users had a dip in the graph when the object was thrown towards them, many of the users actually had an upward trend in the graph when the object was thrown towards them (See Figure 19). This illustrates that some users felt an increase in PI when the object was thrown towards them.

### 4.3. Video and Coordinates Data

While the session was underway, the user s video was being recorded along with the 3D coordinates of the HMD and two controllers (See Figure 11). After trying out different plotting techniques of the coordinates, it was decided to lter out the coordinates data recorded during the two stimulation events, when the object is thrown towards the camera, and when the confederate approaches the camera. Since the coordinates and the video were recorded independently from each other, they were synchronized carefully and then the stimulus portion was cropped out. For the object throwing part, 200 frames were selected and for the confederate approaching part, 300 frames were selected. Since, the data was recorded at 60 frames per second,
$200$ frames $= 300/60 = 3{,}33$ seconds, and
$300$ frames $= 300/60 = 5$ seconds.

For plotting the data, average vector distances from a baseline were calculated. For the baseline, we calculate the average of 10 frames prior to the rst frame of the stimulus data. Since we are dealing with vectors, we disregard the rotation data and focus solely on the positions. Length of a vector from the three position coordinates is calculated as $|\vec{V}| = \sqrt{V_x^2 + V_y^2 + V_z^2}$, where $V_x$, $V_y$ and $V_z$ are the three positional coordinates. The difference between these vectors and the baseline vector is calculated and then all the resulting vector differences are averaged out. In the graphs below a histogram is plotted for data with a virtual body (orange) and without a body (blue). h_p, l_p, r_p are the vector lengths of HMD position, left controller position and the right controller position respectively.

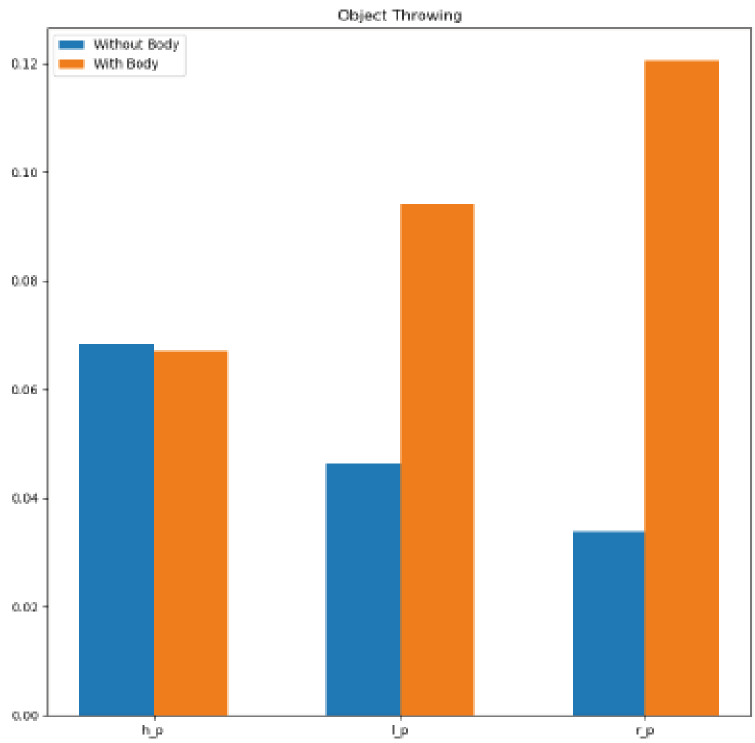


Figure 20. Coordinates Data when the object is thrown towards the camera. (y-axis in metres)

Figure 20 and Figure 21 show signi cantly higher movements with a body. This provides a strong evidence that the users had a stronger response to a stimulus when they had a virtual body. Having a stronger response when having a virtual body could be interpreted as the user having a higher sense of presence with a virtual body. Although other factors such as user s own personality could affect the results. Therefore, a higher sample size is needed to draw any meaningful conclusion.

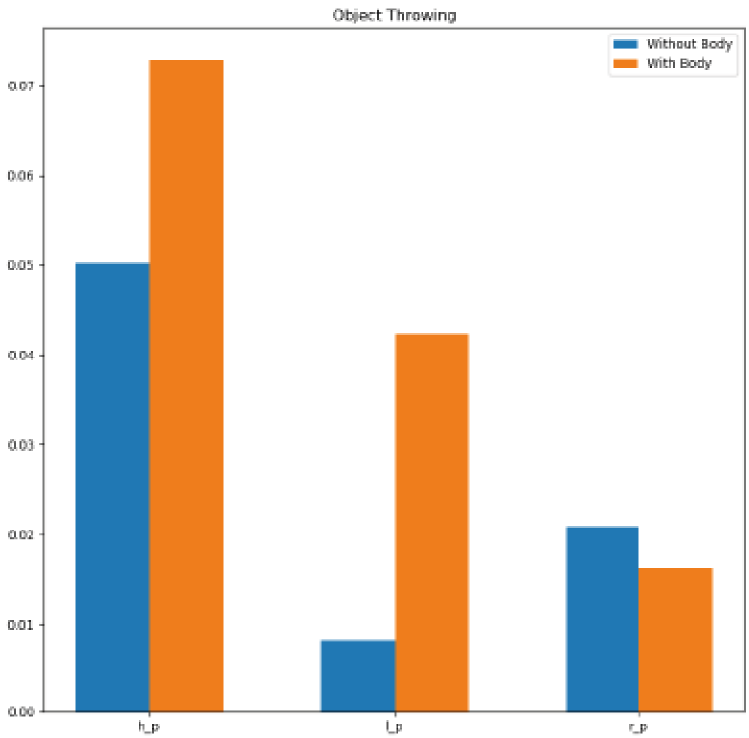


Figure 21. Coordinates Data when the confederate approaches the camera. (y-axis in metres)

# 5. CONCLUSION

In this thesis, the aspect of augmenting the VR Telepresence experience using a self-avatar is explored. The study is conducted on 20 participants, 13 male and 7 female, where they experience a telepresence scenario using live footage from a 360-degree camera. The scenario is broken into two parts, one while embodying a self-avatar and one without any visible body. Video data, coordinates data, questionnaires and variation in presence graphs are the different data points collected from the users and comparison is made between data collected while the user had a virtual body, and when the user did not have any visible body.

Compiling the results from all these data points show that the majority of users preferred the experience with a virtual body, which was an expected result. Unexpectedly, however, that did not necessarily improve their PI score. According to some open-ended answers, the length of the session was too short and the users did not get enough time to fully immerse themselves in the session. Hence we can see outliers in the results of the scenario with the virtual body as there was a conflict of interest between improvement of experience with a virtual body and not getting enough time to feel present in the remote environment. This is proven by the fact that the users tend to give a higher PI score to the second scenario irrespective of them having a virtual body or not. Results from other questionnaires indicate that the sample size was too small to draw any meaningful conclusions.

The users who preferred the experience without any body felt like the hands were not realistic enough or they got in the way. Video data show them being unable to form gestures with the controller buttons as they were not used to doing so. This drawback is possible to resolve by tracking hands without any controllers. The new Oculus Quest 2 VR system has the capability to track hands motion and finger movements without needing any controller or markers. Thus future studies might not have to deal with this issue.

Although the hypothesis presented by the thesis did not reach a conclusion, feedback from the users favours towards having a virtual body. The results gained from this study can be used to improve the future study with a larger group of participants that will conceivably provide meaningful results to answer the hypothesis.

## 5.1. Future Work

The original vision of telepresence that Marvin Minsky proposed in 1980 was a system where a user is able to operate robotic arms in a remote location that mirrors the motion of the user's hands in real-time. Having a virtual body is certainly a step in the right direction for achieving that vision. A self-avatar can also have an impact on social interactions as is noted by the results of the post-experiment questionnaire. The users enjoyed being able to use body language in communication along with the speech and that adds an extra layer of functionality on top of regular video conferencing.

Since this study simulates a remote environment, future studies can be performed on introducing novel techniques for a latency-free and 3D audio system from a remote location. With the advancements in the 5G mobile networks, creating a system capable of this is certainly plausible. Following Minsky's vision and many science fiction author s dreams, future work in the robotics eld that enables a robot to mirror a human s actions is also an appealing topic for future studies.

# 6. REFERENCES


[1] Slater M. (2009) Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philosophical Transactions of the Royal Society B: Biological Sciences 364, pp. 3549–3557. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2009.0138.

[2] Stavar A., Dascalu L.M. & Talaba D. (2011) Design, test and experimental validation of a vr treadmill walking compensation device. In: L.M. Camarinha-Matos (ed.) Technological Innovation for Sustainability, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 402–409.

[3] Kim M., Jeon C. & Kim J. (2017) A study on immersion and presence of a portable hand haptic system for immersive virtual reality. Sensors 17. URL: https://www.mdpi.com/1424-8220/17/5/1141.

[4] Caserman P., Garcia-Agundez A., Konrad R., Göbel S. & Steinmetz R. (2019) Real-time body tracking in virtual reality using a vive tracker. Virtual Reality 23, pp. 155–168.

[5] Minsky M. (1980) Telepresence. OMNI magazine URL: https://web.media.mit.edu/~minsky/papers/Telepresence.html.

[6] Skarbez R., Neyret S., Brooks F.P., Slater M. & Whitton M.C. (2017) A psychophysical experiment regarding components of the plausibility illusion. IEEE Transactions on Visualization and Computer Graphics 23, pp. 1369–1378.

[7] Popovic M.B. & Bowers M.P. (2019) 2 - kinematics and dynamics. In: M.B. Popovic (ed.) Biomechatronics, Academic Press, pp. 11–43. URL: https://www.sciencedirect.com/science/article/pii/B9780128129395000021.

[8] LaValle S.M. (2019) Virtual Reality. Cambridge University Press. URL: http://vr.cs.uiuc.edu/vrbook.pdf.

[9] Shen X. & Shirmohammadi S. (2008) Virtual and Augmented Reality, Springer US, Boston, MA, chap. Virtual and Augmented Reality. pp. 962–967. URL: https://doi.org/10.1007/978-0-387-78414-4_253.

[10] Ellis S. (1994) What are virtual environments? IEEE Computer Graphics and Applications 14, pp. 17–22.

[11] Dodgson N. (2004) Variation and extrema of human interpupillary distance. In: Proceedings of SPIE - The International Society for Optical Engineering, vol. 5291, vol. 5291, pp. 36–46.

[12] Casanueva J. & Blake E. (2001) The effects of avatars on co-presence in a collaborative virtual environment. In: Proc Ann Conf SA Inst of Computer Scientists and Information Technologists, pp. 1–N.

[13] Wheatstone C. (1838) Contributions to the physiology of vision. —part the rst. on some remarkable, and hitherto unobserved, phenomena of binocular vision. In: Philosophical Transactions, vol. 128, Royal Society, vol. 128, p. 371–394.

[14] Sutherland I.E. (1968) A head-mounted three dimensional display. In: Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I, AFIPS 68 (Fall, part I), Association for Computing Machinery, New York, NY, USA, p. 757–764. URL: https://doi.org/10.1145/1476589.1476686.

[15] Artaud A. (1958) The theatre and its double, trans. Mary Caroline Richards (New York: Grove, 1958) 124.

[16] Ware J. (2018), Virtual reality: a (very) brief history, part 1. URL: https://finfeed.com/opinion/ctrl-alt-del/virtual-reality-very-brief-history-part-1/.

[17] Brooks F. (1999) What s real about virtual reality? IEEE Computer Graphics and Applications 19, pp. 16–27.

[18] Mazuryk T. & Gervautz M. (1996), Virtual reality - history, applications, technology and future.

[19] Chang E., Hwang I., Jeon H., Chun Y., Kim H.T. & Park C. (2013) Effects of rest frames on cybersickness and oscillatory brain activity. In: 2013 International Winter Workshop on Brain-Computer Interface (BCI), pp. 62–64.

[20] Cruz-Neira C., Sandin D.J., DeFanti T.A., Kenyon R.V. & Hart J.C. (1992) The cave: Audio visual experience automatic virtual environment. Commun. ACM 35, p. 64–72. URL: https://doi.org/10.1145/129888.129892.

[21] Lin Edwards P. (2009), Active shutter 3d technology for hdtv. URL: https://phys.org/news/2009-09-shutter-3d-technology-hdtv.html.

[22] Iskenderova A., Weidner F. & Broll W. (2017) Drunk virtual reality gaming: Exploring the in uence of alcohol on cybersickness. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play, CHI PLAY 17, Association for Computing Machinery, New York, NY, USA, p. 561–572. URL: https://doi.org/10.1145/3116595.3116618.

[23] Rovira A., Swapp D., Spanlang B. & Slater M. (2009) The use of virtual reality in the study of people s responses to violent incidents. Frontiers in Behavioral Neuroscience 3, p. 59. URL: https://www.frontiersin.org/article/10.3389/neuro.08.059.2009.

[24] Rizzo A.S. & Kim G.J. (2005) A SWOT Analysis of the Field of Virtual Reality Rehabilitation and Therapy. Presence: Teleoperators and Virtual Environments 14, pp. 119–146. URL: https://doi.org/10.1162/1054746053967094.

[25] Browning M.H.E.M., Mimnaugh K.J., van Riper C.J., Laurent H.K. & LaValle S.M. (2020) Can simulated nature support mental health? comparing short, single-doses of 360-degree nature videos in virtual reality with the outdoors. Frontiers in Psychology 10, p. 2667. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2019.02667.

[26] Das D., Grimmer K., Sparnon A., McRae S. & Thomas B. (2005) The ef cacy of playing a virtual reality game in modulating pain for children with acute burn injuries: A randomized controlled trial [isrctn87413556]. BMC pediatrics 5, p. 1.

[27] Harvey C.D., Collman F., Dombeck D.A. & Tank D.W. (2009) Intracellular dynamics of hippocampal place cells during virtual navigation. Nature 461, pp. 941–946. URL: https://doi.org/10.1038/nature08499.

[28] Schell B.R. (2020) Employing virtual reality to reveal individual locusts decision-making. Master s thesis, Universität Konstanz, Konstanz.

[29] Lang F. (2021), Cows are being tted with vr goggles to increase milk production? URL: https://interestingengineering.com/cows-on-russian-farm-get-fitted-with-vr-goggles-to-increase-milk-production.

[30] Suomalainen M., Nilles A.Q. & LaValle S.M. (2020) Virtual reality for robots. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 11458–11465.

[31] Shen X. & Shirmohammadi S. (2008) Telepresence, Springer US, Boston, MA, chap. Telepresence. pp. 849–852. URL: https://doi.org/10.1007/978-0-387-78414-4_229.

[32] Schloerb D.W. (1995) A Quantitative Measure of Telepresence. Presence: Teleoperators and Virtual Environments 4, pp. 64–80. URL: https://doi.org/10.1162/pres.1995.4.1.64.

[33] TelepresenceRobots.com (2021), What is a telepresence robot and what can they do? URL: https://telepresencerobots.com/what-telepresence-robot-and-what-can-they-do.

[34] Baldis J.J. (2001) Effects of spatial audio on memory, comprehension, and preference during desktop conferences. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 01, Association for Computing Machinery, New York, NY, USA, p. 166–173. URL: https://doi.org/10.1145/365024.365092.

[35] Kachach R., Perez P., Villegas A. & Gonzalez-Sosa E. (2020) Virtual tour: An immersive low cost telepresence system. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 504–506.

[36] Schlag P.M., Moesta K.T., Rakovsky S. & Graschew G. (1999) Telemedicine: The New Must for Surgery. Archives of Surgery 134, pp. 1216–1221. URL: https://doi.org/10.1001/archsurg.134.11.1216.

[37] Schultz R., Nakajima R. & Nomura J. (1991) Telepresence mobile robot for security applications. In: Proceedings IECON 91: 1991 International Conference on Industrial Electronics, Control and Instrumentation, pp. 1063–1066 vol.2.

[38] LYU W., ZHOU Z., CHEN L. & ZHOU Y. (2019) A survey on image and video stitching. Virtual Reality & Intelligent Hardware 1, pp. 55–83. URL: https://www.sciencedirect.com/science/article/pii/S2096579619300063.

[39] Li Y. & Ma L. (2006) A fast and robust image stitching algorithm. In: 2006 6th World Congress on Intelligent Control and Automation, vol. 2, vol. 2, pp. 9604–9608.

[40] Insta360 (2020), First look: Inside verizon s 5g vr live-stream of the pro bowl. URL: https://blog.insta360.com/verizon-pro-bowl-vr-sports-broadcasting/.

[41] Fleischer M. (1917), Method of producing moving-picture cartoons.

[42] Calvert T.W., Chapman J. & Patla A. (1982) Aspects of the kinematic simulation of human movement. IEEE Comput. Graph. Appl. 2, p. 41–50. URL: https://doi.org/10.1109/MCG.1982.1674484.

[43] Ginsberg C.M. & Maxwell D. (1983) Graphical marionette. In: Proc. ACM SIGGRAPH/SIGART Workshop on Motion, ACM Press, New York, p. 172–179.

[44] Caserman P., Garcia-Agundez A., Konrad R., Göbel S. & Steinmetz R. (2018) Real-time body tracking in virtual reality using a vive tracker. Virtual Reality 23, pp. 155–168. URL: https://doi.org/10.1007/s10055-018-0374-z.

[45] Jain A., Tompson J., LeCun Y. & Bregler C. (2014) Modeep: A deep learning framework using motion features for human pose estimation. In: Asian conference on computer vision, Springer, pp. 302–315.

[46] Zult T., Allsop J., Tabernero J. & Pardhan S. (2019) A low-cost 2-d video system can accurately and reliably assess adaptive gait kinematics in healthy and low vision subjects. Scienti c Reports 9. URL: https://doi.org/10.1038/s41598-019-54913-5.

[47] Colombo C., Del Bimbo A. & Valli A. (2001) Real-time tracking and reproduction of 3d human body motion. In: Proceedings 11th International Conference on Image Analysis and Processing, pp. 108–112.

[48] Caserman P., Tregel T., Fendrich M., Kolvenbach M., Stabel M. & Göbel S. (2018) Recognition of full-body movements in vr-based exergames using hidden markov models. In: S. Göbel, A. Garcia-Agundez, T. Tregel, M. Ma, J. Baalsrud Hauge, M. Oliveira, T. Marsh & P. Caserman (eds.) Serious Games, Springer International Publishing, Cham, pp. 191–203.

[49] Caserman P., Garcia-Agundez A. & Göbel S. (2020) A survey of full-body motion reconstruction in immersive virtual reality applications. IEEE Transactions on Visualization and Computer Graphics 26, pp. 3089–3108.

[50] Chessa. M., Caroggio. L., Huang. H. & Solari. F. (2016) Insert your own body in the oculus rift to improve proprioception. In: Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISION4HCI, (VISIGRAPP 2016), INSTICC, SciTePress, pp. 755–762.

[51] Botvinick M. & Cohen J. (1998) Rubber hands feel touch that eyes see. Nature 391, pp. 756–756. URL: https://doi.org/10.1038/35784.

[52] Slater M. (2009) Inducing illusory ownership of a virtual body. Frontiers in Neuroscience 3, pp. 214–220. URL: https://doi.org/10.3389/neuro.01.029.2009.

[53] Bourdin P., Barberia I., Oliva R. & Slater M. (2017) A virtual out-of-body experience reduces fear of death. PLOS ONE 12, p. e0169343. URL: https://doi.org/10.1371/journal.pone.0169343.

[54] Johnson-Glenberg M.C. (2018) Immersive VR and education: Embodied design principles that include gesture and hand controls. Frontiers in Robotics and AI 5. URL: https://doi.org/10.3389/frobt.2018.00081.

[55] Unity (2020), Animation rigging v1.0.3. URL: https://docs.unity3d.com/Packages/com.unity.animation.rigging@1.0/manual/index.html.

[56] RootMotion (2014), Final ik. URL: https://assetstore.unity.com/packages/tools/animation/final-ik-14290.

[57] Lugrin J.L., Latt J. & Latoschik M.E. (2015) Avatar anthropomorphism and illusion of body ownership in vr. In: 2015 IEEE Virtual Reality (VR), pp. 229–230.

[58] Pan Y. & Steed A. (2019) How foot tracking matters: The impact of an animated self-avatar on interaction, embodiment and presence in shared virtual environments. Frontiers in Robotics and AI 6, p. 104. URL: https://www.frontiersin.org/article/10.3389/frobt.2019.00104.

[59] Keskinen T., Mäkelä V., Kallioniemi P., Hakulinen J., Karhu J., Ronkainen K., Mäkelä J. & Turunen M. (2019) The effect of camera height, actor behavior, and viewer position on the user experience of 360° videos. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 423–430.

[60] Brinkman W.P., Hoekstra A.R. & van EGMOND R. (2015) The effect of 3d audio and other audio techniques on virtual reality experience. ANNUAL REVIEW OF CYBERTHERAPY AND TELEMEDICINE 2015 , p. 44.

[61] McLeod S.A. (2019), Likert scale. simply psychology. URL: https://www.simplypsychology.org/likert-scale.html.

[62] Usoh M., Catena E., Arman S. & Slater M. (2000) Using Presence Questionnaires in Reality. Presence: Teleoperators and Virtual Environments 9, pp. 497–503. URL: https://doi.org/10.1162/105474600566989.

[63] Gonzalez-Franco M. & Peck T.C. (2018) Avatar embodiment. towards a standardized questionnaire. Frontiers in Robotics and AI 5, p. 74. URL: https://www.frontiersin.org/article/10.3389/frobt.2018.00074.

# 7. APPENDICES

| | |
|---|---|
| Appendix 1 | Presence Questionnaire |
| Appendix 2 | Co-Presence Questionnaire |
| Appendix 3 | Break in Presence Questionnaire |
| Appendix 4 | Virtual Body Questionnaire |
| Appendix 5 | Post Experiment Questionnaire |
| Appendix 6 | Confederate s Script |
| Appendix 7 | Running Checklist |

Please rate your sense of being in the remote environment, on a scale of 1 to 7, where 7 represents your normal experience of being in a place. *

I had a sense of "being there" in the remote environment

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Not at all | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Very much |

To what extent were there times during the experience when the remote environment was the reality for you? *

There were times during the experience when the remote environment was the reality for me

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| At no time | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Almost all the time |

When you think back to the experience, do you think of the remote environment more as watching a movie or more as someplace that you visited? *

The remote environment seemed to me to be more like

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Watching a movie | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Someplace I visited |

During the time of the experience, which was strongest on the whole, your sense of being with the host, or of sitting at your original position? *

I had a stronger sense of

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Sitting at original position | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Being with the host |

Consider your memory of being in the remote environment. How similar in terms of the structure of the memory is this to the structure of the memory of other places you have been today? By 'structure of the memory' consider things like the extent to which you have a visual memory of the remote environment, whether that memory is in colour, the extent to which the memory seems vivid or realistic, its size, location in your imagination, the extent to which it is panoramic in your imagination, and other such structural elements. *

I think of the remote environment as a place in a way similar to other places that I've been today

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Not at all | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Very much so |

During the experience, did you often think to yourself that you were actually in the same location as the host? *

During the experience I often thought that I was at the same location as the host

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Never | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Most of the time |

I felt as if I was right next to the host *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Strongly Agree |

I think that the host felt as if she was next to me during the course of the experiment *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Strongly Agree |

I think the host often felt like talking to a person rather than talking to a camera *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Strongly Agree |

I felt more like talking to a person directly rather than talking through a video conferencing software *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Strongly Agree |

Were there any moments when you became suddenly aware of where you were physically sitting? *

*Mark only one oval.*

- ⬭ Yes
- ⬭ No

If yes, at what point did that happen?

What triggered you suddenly becoming aware of where you were physically sitting?

How did becoming aware of that make you feel?

How easy or quick was your recovery (recovering your sense of being in the remote environment)?

I felt as if the robot body was my own body *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | Strongly Agree |

I felt as if the robot body I saw was someone else *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | ⬭ | Strongly Agree |

It felt like I could control the robot body as if it was my own body *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Strongly Agree |

The movements of the robot body were caused by my movements *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Strongly Agree |

I felt as if the movements of the robot body were influencing my own movements *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Strongly Agree |

I felt as if the robot body was moving by itself *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Strongly Agree |

I felt as if my body was located where I saw the robot body *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Strongly Agree |

I actually felt like I embodied the robot *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | ◯ | Strongly Agree |

Thinking back to both of the experiences, which one gave a better sense of being in the same place as the host *

*Mark only one oval.*

- Without Body
- With Body

Why?

Did you think that seeing the robot body improved the overall experience? *

*Mark only one oval.*

- ( ) Yes
- ( ) No

Why do you think so?

Any other comments regarding the experiment?

Gender *

*Mark only one oval.*

- Female
- Male
- Prefer to self-describe
- Prefer not to say

What is your age? *

What is your current occupation?

How often do you play video games? *

*Mark only one oval.*

- Never
- Once or a just a couple times ever
- Once or twice a year
- Once or twice a month
- Once or twice a week
- Several times a week
- Every day

What kind of games do you play?

For example, first-person shooters, adventure games, etc.

How often do use Virtual Reality? *

*Mark only one oval.*

- Never
- Once or a just a couple times ever
- Once or twice a year
- Once or twice a month
- Once or twice a week
- Several times a week
- Every day

What kind of Virtual Reality experiences do you use?

For example, first-person shooters, 360 videos, adventure games, etc.

If you are experienced with Virtual Reality, how often do you feel uncomfortable using it and why?

R: Researcher, S: Subject

(Researcher enters the room)
R: (Waves hand) Hey, how are you?
PAUSE FOR RESPONSE
R: Welcome to our lab. My name is Alexis and I'll be your host for today. At the Perception Engineering research group, in the Center for Ubiquitous Computing, we focus on researching Virtual Reality, Augmented Reality, Robotics and similar fields. I'm going to tell you about various equipment we use in the lab.
R: Our research group evaluates virtual reality headsets. To do this, we need to perform calibrations of the visual and audio system equipment. For this, we use a dummy human head with cameras in the eyes to test that. Can you guess which device is used for those tests?
PAUSE FOR RESPONSE
R: It's this one here. (Points towards OptoFidelity). This is called the OptoFidelity Buddy. So the shape resembles a human head and we strap the VR headset around the eyes which measures any visual latency and it can also rotate around to measure the refresh rate of the screen. Kinda cool, right?
PAUSE
R: If you look at the table to your right, (Point towards cleanbox) one of the first things you might notice is that box with a white cylinder in the middle. That is called a cleanbox and it's used to sanitize our Virtual Reality headsets using UV-C light. It can kill 99.999% of bacteria and viruses. With the current pandemic situation, this is how we can safely run studies with several people using the same device without the risk of spreading the virus.
R: (Picks up the black rat). I bet you're wondering about these rats behind me, too. This guy's name is Remy and he is our lab rat for testing all sorts of experiments before we try them on humans. Just kidding! Although, he is usually with a white rat named Chewy. Do you see a white rat around here? (Look left and right but not behind and don't block the white rat from the subject).
PAUSE FOR RESPONSE
R: (If the subject can't find it within 10 seconds, just find it yourself). Aah, there she is. (Pick it up and pretend it is biting you). Ah! She bit my finger! (Throw towards the camera) Here, catch!
PAUSE FOR RESPONSE
R: Wow. Well, that was weird. Let's take a break from the tour here so I can look for a bandage and continue in a bit.
PAUSE FOR RESPONSE
(Researcher leaves the lab and subject fills the questionnaire)

(After the second session begins)
R: (After entering, gets scared of the mannequin placed at the entrance) Oh this guy scared me. I really think it could be in a better spot than this. Where would you suggest where we put it in the future?
PAUSE FOR ANSWER
R: Got it / Hmm, I'll figure it out. Ready to continue the tour?
PAUSE FOR ANSWER
R: Great, now I can briefly explain some of the projects we do in this lab. So, it's mainly used for testing out Virtual environments inside Virtual Reality. As you can see this room is pretty spacious, albeit a bit cluttered, but with VR, you can't see the real world. You need a spacious room to allow you some degree of movement. We mainly use this lab for conducting user studies where real world users are invited and they try out our projects and give us their feedback. Have you ever been here before?
PAUSE FOR ANSWER
R: So let's continue on to some more of our research equipment. (Turn towards the mannequin) This mannequin is modeling some of our cool body sensors. Can you see which parts of the body they are attached to?
PAUSE FOR ANSWER
R: Right, those are biosensors that we use in some of our studies. Biosensors measure body signals, like heart rate, blood pressure, and muscle movements. We have special sensors to measure a couple of different things for our research. The sensors that we put on people's fingers measure EDA, or electrodermal activity. That's a measure of the conductivity of the skin on your fingers. When your body becomes physiologically stimulated, your sympathetic nervous system responds with the fight-or-flight response. Have you ever heard the term fight-or-flight?
PAUSE FOR RESPONSE
R: When the fight-or-flight response happens, adrenaline stimulates an increase in heart rate, and you begin to sweat since your body anticipates that you might need to react to some danger. That is why you get sweaty palms when you are nervous about something. The increase in sweat on your fingers increases the conductivity across your skin, which is measured by the EDA sensors. In this way, we can measure if you are responding to something we show you. This is the same technology that's used in lie detectors. Interesting, huh?
PAUSE, LOOK AT CAMERA
R: Ah, I just noticed Chewy is still hanging out over there! Let's put Chewy back with her friend Remy so she can calm down after that last incident.
(Researcher gets close to the camera and picks up the white rat and place it next to Remy)
R: That's better. Well, that concludes the tour. Thank you for joining me and have a lovely day. (Waves) Bye!
(Researcher leaves the lab and the subject fills the rest of the questionnaire).

**RUN SHEET** Subject ID:__________ Date, Study Time:_______________

- ☐ Your phone on silent, Plug in Laptops
- ☐ Check and clean Oculus, Wipe off lenses, Check controller batteries
- ☐ Check that chair and camera are in correct position
- ☐ Check that the curtains are drawn
- ☐ Place the items in their original positions: Black rat near the confederate, White rat at an elevated position on the table at the back.
- ☐ Open all questionnaires, Determine User ID, Pre-sign Consent Form, pre-fill Subject ID, Date, Time and Counterbalancing scenario on laptop
- ☐ Locate the auto.py file located on the desktop
- ☐ Check that the 360 camera is set on live streaming and start the camera app to prevent the camera from going to sleep
- ☐ Make the confederate stand outside the research room

************** Subject comes in **************

- ☐ Subject Phone on silent
- ☐ Ask the user to sit on the chair and ask them to not move the chair
- ☐ Subject Information Sheet and Consent Form
- ☐ Check if subject wrote yes or no about the video consent
- ☐ Ask the subject, Have you been in this room before?: Yes_______ or No________
- ☐ Read Instructions to Subjects
- ☐ Show them how to Put on Headset and the controller mappings
- ☐ Ask if Image is Clear and determine if they are able to see the room and there is no Oculus Menu open
- ☐ Start A/B Scenario: __________ by opening the auto.py file and enter ID and scenario
- ☐ Signal the confederate to enter the room and open the curtains. Close afterwards
- ☐ Make the subject draw the graph, SSQ after 1st Scenario
- ☐ Start A/B Scenario: ____________ by opening the auto.py file and enter ID and scenario
- ☐ Signal the confederate to enter the room and open the curtains. Close afterwards
- ☐ Make the subject draw the graph, SSQ after 2nd Scenario
- ☐ Post experiment SSQ
- ☐ Give Debriefing, Compensation, Consent and Info Page and ask if they have Any Comments?