An Examination of Presence and Engagement in Video Conferencing Systems and Virtual Environments

Mark B. Manuel

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Science & Applications

Denis Gračanin, Chair

R. Benjamin Knapp, Co-Chair

Thomas L. Martin

December 10, 2021

Blacksburg, Virginia

Keywords: Virtual Environments, Extended Reality, Human Computer Interaction,

Remote Collaboration, Embodied Interaction

An Examination of Presence and Engagement in Video Conferencing Systems and Virtual Environments

Mark B. Manuel

(ABSTRACT)

Extended Reality (XR) is an upcoming field of technology that has garnered interest from researchers in the last few decades. This increased interest is largely due to the development of powerful hardware like the Microsoft HoloLens, Oculus Quest and the Magic Leap. Several companies like Microsoft, Meta Platforms, Apple and Nvidia are touting the rise of a new “metaverse” - the next generation of the internet, that will blur the lines between physical and virtual presence. This thesis explores the use of web-based XR platforms in Computer-Supported Cooperative Work (CSCW) as an alternative to contemporary video conferencing tools.

We conducted a user study with 15 subjects to evaluate web-based XR platforms (Mozilla Hubs) with video conferencing (Zoom) and examined subject attention and success in remote collaborative tasks. We also proposed a new system design to support embodied interactions in XR. This system was tested by measuring the communication latency between two collaborators separated by varying distances. Our system performance evaluation suggests the feasibility of support embodied interactions, with a minimal latency of 120ms across a distance of 4700 miles.

An Examination of Presence and Engagement in Video Conferencing Systems and Virtual Environments

Mark B. Manuel

(GENERAL AUDIENCE ABSTRACT)

Extended Reality (XR) is changing the way we interact with digital content. In 2021, several companies like Microsoft, Meta Platforms, Apple and Nvidia are developing devices that allow users to physically interact with virtual content in 3-dimensional space. These technologies bring with them the promise of better remote communication and collaboration. Users will be able to enter these 3d virtual spaces as avatars and will be able to interact with digital media just like they would with real world objects. This thesis explores the use of web-based XR platforms in supporting remote collaboration as an alternative to contemporary video conferencing tools.

We conducted a user study with 15 subjects to compare differences in web-based XR platforms (Mozilla Hubs) with video conferencing (Zoom). We proposed and evaluated a new system design to support more natural and intuitive interactions in XR. This system was tested by measuring the communication latency between two collaborators separated by varying distances. Our system performance evaluation suggests the feasibility of support embodied interactions, with a minimal latency of 120ms across a distance of 4700 miles.

Acknowledgments

This journey through graduate school has been an extraordinary growth period for me and I have been fortunate to have so many wonderful people to support me along the way.

Firstly, I am extremely grateful to my advisor Dr. Denis Gračanin for your consistent support throughout my master's program. This thesis would not have been possible without your advice and encouragement. I would also like to especially thank my co-advisor, Dr. Ben Knapp for all of your support over the past few years. It was by working with you on the Mirror Worlds project in 2016, that I first began my foray into research. I sincerely appreciate all of the kind, motivating words and advice that you've provided me with as my co-advisor. I would also like to thank Dr. Tom Martin for taking an interest in this work.

I wish to express my heartfelt gratitude to my father, Balfour Manuel, my mother, Metty Manuel, my brother, Alan, cousins, Beri, Dale, Gale, Gabriel, Jean and all of my extended family members for helping me in uncountable ways throughout my time in graduate school. Mohamed Handosa, Matt Tucker and Reza Tasooji, I thank you for helping me with some of my earliest research projects. I learned so much from you! Nikitha Chandrashekar and Juwon Park, I will cherish the time we spent together working on the Trauma Project. Sam Williams, Shaoli Dasgupta and Gunnar Nelson, it was a pleasure to work with you on the Virtual Museum Project.

Finally, this acknowledgement would be incomplete without mentioning my incredible friends, Archi Dasgupta, Poorvesh Dongre and Rifat Sabbir Mansur. I would not be the person I am today without the adventures that we have had together. You guys have helped me power through some of my most challenging days and for this I am truly grateful.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"*In the past, before phones and the Internet, all communication was face-to-face. Now, most of it is digital, via emails and messaging services. If people were to start using virtual reality, it would almost come full circle.*"

— - Palmer Luckey, *Oculus VR*

Remote communication technologies have been evolving for centuries. Letter writing, telegraphs, the telephone and email are all tools that we developed to better communicate with each other across large distances. These tools have connected different parts of the world, helped spread information and have accelerated technology growth. Today, one of the most common means of remote communication is video conferencing. Video conferencing applications are capable of disseminating both aural (audio) and visual information in real-time. Having access to both of these modes of information, communicators are able to feel some sense of presence and co-location with their remote collaborators. However, these applications are still limited in their ability to make users feel truly co-present with one another. Video conferencing tools are unable to convey the full range of human expression, which includes body language and other non-verbal means of communication. As a result, there is still a significant difference between in-person communication and remote communication. Therefore, there is a need for a more immersive communication medium that can enable better collaboration between remote users.

1.1 Motivation

Extended Reality (XR) can serve as a suitable alternative to video conferencing. With XR, the spatial barrier presented by video-calling applications can be remedied, allowing for even greater immersion. In XR space, remote participants can view each other as holographic or digitally-rendered avatars co-located in the same physical/virtual space as themselves. Imagine something akin to the round table scene from the movie Kingsman [32], in which dignitaries across the globe are co-located as holograms in a mixed-reality enabled round table conference. In such spaces, participants would be able to personalize their avatars to better represent themselves [3, 71]. Body tracking and facial recognition would enable avatars to a broad range of gestures and facial expressions. Remote participants would be able to share and explore 3D objects and environments, leading to a richer, more immersive collaborative experience [4].

XR teleconferencing, with all of these additional layers of information that can be provided through it, could be considered an improved iteration over video-based communication technologies [107]. However, the feasibility of using XR environments as a replacement for existing video communication technology requires further evaluation. In this thesis we seek to compare user presence and engagement differences between online virtual environments and video conferencing. We chose online to evaluate online virtual environments over fully immersive XR environments to account for accessibility issues that users might have in obtaining XR hardware.

Additionally, we propose a system design to support embodied interactions in XR environments. Our rationale was that XR platforms with embodied interaction would make users feel more comfortable navigating the environment and provide for more immersive collaborative experiences.

1.2 Research Questions

The goal of this thesis is to evaluate the scope of using online XR environments for day-to-day remote collaboration tasks. Since the advent of the COVID-19 pandemic, video conferencing has become the primary way for most of us to work from home. Schools, offices and government agencies across the world are using commonly available video conferencing tools like Zoom or Skype to conduct their day-to-day operations [81]. These tools, however, are limited in their ability to recreate the feeling of being co-located. Frequent use of video conferencing tools can also lead to burnout and Zoom Fatigue [108]. XR environments can help address some of these challenges by providing users with richer meeting spaces to collaborate in. XR environments can be imbued with additional affordances like avatars with expressive facial features and body-tracking [17, 60]. The popularity of web-based XR environments has been growing since the pandemic with platforms like Mozilla Hubs and VirBela being used to support academic conferences attended by several hundred attendees (ISMAR 2020, IEEVR 2020). These platforms have usability affordances (e.g., personalized avatars, game-like gestures and spatialized audio) that make them more immersive and interesting for users. This thesis explores the potential for such platforms to become the standard medium for remote collaboration. We specifically explore the research questions listed below. These are described in more detail in Chapter 3.

1.2.1 Can online virtual environments can serve as a suitable medium for remote collaboration?

The subject of presence in virtual environments is a common area of research. However, the difference in levels of presence and engagement provided by video conferencing systems and virtual environments has not been clearly explored. This is especially relevant now, as remote work has become more prevalent than ever before. Video communication systems like Zoom, WebEx and Skype provide basic services for communication between participants, yet they cannot fully capture and recreate the effect of being "present" at one's workplace. Virtual Environments like Mozilla Hubs, VirBELA and EngageVR can replicate real-world environments and provide users with a keener sense of being present at their workplace. The purpose of this study is to understand the extent to which study subjects experience the feeling of "being there" [67] while performing simple tasks using VCEs and VEs. The findings from this thesis can be utilized to design more user-friendly remote-workplace/collaboration tools.

1.2.2 Can online virtual environments provide higher levels of presence and engagement for multi-user meetings when compared to contemporary video conferencing systems?

Virtual Environments (VR-based) are engaging, immersive experiences that capture the entirety of a user's visual field, transporting them new and unique environments. The nature of the medium makes it highly effective at engaging users with virtual content and making them feel present in the virtual world. However, fully-immersive VR relies on the use of special hardware which may not be accessible to all users. Online virtual environments can help bridge this accessibility gap. The efficacy of online virtual environments in providing users with experiences similar to fully-immersive VR needs to be further explored. One of the goals of this thesis is to evaluate whether current online virtual environments are robust enough for users to frequently use for remote collaboration. The feasibility of these virtual environments as replacements to video-conference-based meetings is explored in this work.

1.2.3 Can 3D virtual environments (VEs) that support embodied interactions be used for remote synchronous meetings?

Finally, we also want to explore new ways to enhance user's sense of presence in XR environments. Embodiment has been shown to play a large role in how present and engaged users feel in a virtual environment. Body-tracking hardware and software need to be developed and integrated with XR platforms appropriately in order to support embodied XR interactions. The design and implementation of one such system is described and evaluated in Section 4.

1.3 Approach

Video conferencing systems like Zoom, Microsoft Teams and Skype are popular collaborative tools that used by millions of people daily [90]. Collaborative virtual environments like Mozilla Hubs and Skype on the other hand are new and only used by a small set of people [88]. While fully immersive virtual environments are predicted to be the successors of modern day remote collaborative tools, there are certain barriers to its widespread adoption. The present reliance on specific hardware makes collaborative XR environments difficult for everyone to access. This trend is expected to persist for at least another few years until the price of the technology becomes commercially viable. Until such a time, intermediary platforms like Mozilla Hubs and VirBELA can serve as stepping stones towards fully immersive XR. The benefit of these intermediary platforms is that they are platform agnostic. Individuals can use these platforms through their web browser or, if they have access to XR hardware, they can even access these platforms through their XR devices. This flexibility can help speed up the adoption of these intermediary platforms.

In addition to being flexible and platform-agnostic, these intermediary XR platforms will also need to provide users, at minimum, with a comparable experience to traditional video conferencing systems. Joining meetings should be easy and intuitive. Users should be able to perform their everyday collaborative tasks, like discussions and presentations, just as easily, if not better than how they perform them using video conferencing tools. Due to the novel nature of these intermediate XR platforms very few studies have explored the usability differences between using videoconferencing and online XR platforms. This thesis seeks to address this gap by comparing the usability of the two platforms. Fifteen users were made to perform 3 similar tasks across both platforms and their usability was measured using the System Usability Scale (SUS) questionnaire [8]. User attention was also measured through eye-tracking data captured during each study session.

This thesis also sought to explore methods to enhance user presence in collaborative virtual environments. One such method that we explored was the incorporation of body-tracking in shared virtual environments. Embodied interactions have been shown to improve user presence in a platform [15, 52, 61]. With our goal of increasing user presence and engagement, we decided to pursue this approach. We designed a system with three components that allow two remote collaborators to share the same space and to interact with shared virtual objects. The complete description of this system can be found in Section 4.

1.4 Contributions

We present the current state of the art in XR and remote collaboration technologies. Additionally, we present our findings from a user study comparing attention and engagement differences between contemporary remote collaboration tools. Finally, we describe the design of a system capable of supporting embodied interactions in XR environments. The contributions of this work include:

- A literature review about the feasibility of XR for Computer-Supported Cooperative Work. This literature review provides background for XR and embodied interactions. It concludes by presenting the current limitation of XR, along with some solutions for addressing these limitations.
- A user study with 15 subjects evaluating attention and engagement differences between video conferencing and online virtual environments. This study makes use of eye-tracking to measure user attention while performing similar tasks.
- A system design and evaluation of an setup designed to support embodied interactions in shared XR environments.

1.5 Thesis Structure

In this chapter (Chapter 1), we addressed the research problem and provided an overview of commonly used remote communication tools.

Chapter 2 discusses literature on the history of Computer-Supported Cooperative Work (CSCW) and the applications of remote collaboration tools. It also provides background for XR and explores it's potential for enhanced remote collaboration.

Chapter 3 defines the hypothesis for this thesis and Chapter 4 discusses the approach taken to test these hypotheses.

In Chapter 5 we discuss the results that we observed from conducting the study.

Finally, Chapter 6 provides the conclusion and talks about future work that can be done to extend the work presented in this thesis.

Chapter 2

Literature Review

This chapter covers several topics that provide the background for the work presented in this thesis. We begin by providing an overview of Extended Reality and Computer-Supported Cooperative Work in the context of enabling remote collaboration. We describe how distributed XR and body tracking sensors can be used to support embodied interactions in real-time. Next, we talk about remote collaborative tools and how eye tracking can be used to measure user attention on collaborative platforms. Finally, we explore several studies where XR applications in order to facilitate remote collaboration between users.

2.1 Introduction

2.1.1 What is XR and why is it important?

Extended Reality(XR) is an umbrella term that refers collectively to Augmented Reality (XR), Virtual Reality (VR) and Mixed Reality (MR). It comprises the entirety of Milgram's Reality-Virtuality continuum [65]. Milgram places real and virtual environments on the extreme ends of this spectrum, with MR representing the entire range of blended reality in between these two ends (Figure 2.1). With news of the metaverse in 2021 and evolving XR hardware, Milgram's virtuality continuum bears special relevance today.

Ivan Sutherland, often regarded as the "father of computer graphics", developed the first XR system (VR) at Harvard in 1966 [94]. In 1990, Tom Caudell and David Mizell coined the term "augmented reality" to refer to overlaying computer-presented material on top of

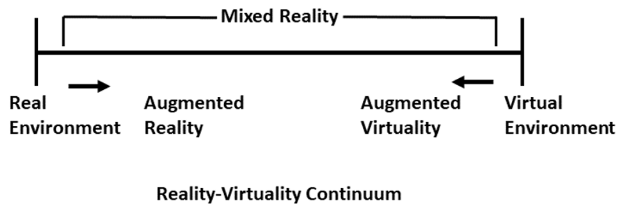


Figure 2.1: Milgram's Reality-Virtuality Continuum [65].

the real world [12]. They created one of the earliest XR applications which was used to train pilots. Later in 1999, the ARToolKit was realeased by the Nara Institute of Science and Technology, allowing developers and researchers to create and deploy their own XR applications [46].

Ladwig and Geiger [50] consider how the advancement of technology in the near future will allow us to create the "ultimate device" that will be capable of making the real world indistinguishable from the virtual world. Such a system would provide "realistic and complete embodied experiences" by utilizing multiple human senses including haptic, sound or even smell and taste. This information if sent over a network and recreated could also facilitate remote collaboration [50].

XR has many advantages when used in education [14], training [18, 41] and in collaborative tasks [5]. It presents the unique ability to create hybrid environments consisting of real and virtual objects. This can be used to overlay additional information, test problem solving skills and to simulate real-world scenarios in a cost-effective manner. It provides an opportunity for experiential learning which is of special importance to disciplines that benefit greatly from hands-on experiences over traditional education; such as those in the medical field. Additionally, decreasing costs and increasing availability of XR devices have made XR applications more popular and attractive for vocational training [36].

The key benefit of XR is the ability to integrate virtual environments and objects into a physical space. Adding virtual constructs to an empty room allows users to engage in interactions that would not be available otherwise. This ability can be used to provide a user with unique experiences without the need to purchase expensive equipment or travel to distant locations.

XR-enhanced environments provide additional features and details to end-users which can enhance a user's experience. Physical environments can be enhanced to be made more intuitive and user-friendly using such technologies. The ability to easily augment physical spaces have enabled the possibility of creating immersive and scalable environments at a low cost. These immersive environments have various applications which include education and skill-training. If the learning is to be done in situations where it could pose a threat to the user or the environment is difficult or expensive to create, an XR environment can mitigate these challenges while allowing the user to achieve the learning outcomes.

Furthermore, XR allows for the simulation of specific scenarios which would be difficult or expensive to simulate in a physical environment. For example, creating an instance of a system failure at a power plant would be difficult to create using real actors and a physical space. XR not only enables the recreation of existing training scenarios, it offers the ability to enhance them and extends the range of topics that can be taught.

2.1.2 Networked and Distributed XR

XR environments can be tailored to support multiple users at the same time. Singhal and Zyda coined the term Networked Virtual Environment (NVE) to define such a space that allows multiple users to remotely interact with each other and the shared XR environment [85].

Robust network support with minimal latency is imperative to sustaining an adequate level of user comfort in such environments. Delaney et al. describe the problem of maintaining a consistent world view for all remote participants [20, 21]. They explained the link between consistency and several properties of NVEs including temporal and spatial synchronization, environment responsiveness and fidelity. Hamza-Lup and Rolland explored similar challenges to maintaining NVE consistency and proposed a novel criterion for the categorization of XR environmnets [35]. They demonstrated that despite high latency, NVE states can be kept consistent for multiple users by keeping the frequency of updates low.

Gracanin et al. reviewed several architectures and approaches to improve Quality of Service (QoS) levels and identified contemporary challenges [33]. They described an approach to reduce network requirements by optimizing application level performance, based on user QoS specifications.

These studies suggest that with a sufficiently low latency (below 100 ms), XR applications can be developed to support multiple users [25, 51]. This opens up the potential for XR to be used as a tool for collaboration between several remotely located users. In Section 4.2 we describe one such approach to supporting real-time embodied interactions between remote collaborators in XR.

2.1.3 Computer-Supported Cooperative Work (CSCW)

The term CSCW originated in the 1980s as technologists started to seek diverse perspectives from economists, sociologists, organizational psychologists and educators on how to best promote group activity at workplaces [34]. According to Grudin, early attempts at creating collaborative tools were unsuccessful largely because developers lacked a proper understanding of how groups of people worked together and the precise role of technology in assisting in this kind of of work. CSCW attempted to address this gap with a broader focus on user requirements extraction.

The term *groupware* is used to collectively refer to any technology that promotes CSCW (email, collaborative writing, video conferencing). Groupware is designed to promote work practice by providing groups of users with better communication mechanisms and tools for problem-solving [70]. Other benefits provided by groupware technologies include: telecommuting/remote work, faster communication and better group and workflow coordination [100].

CSCW was novel in the 1980s because it enabled remote, asynchronous communication between people [28]. However, groupware has expanded to also include tools that allow for synchronous interactions between remote collaborators. Dourish and Bly further explored the use of groupware and media space technologies to support general awareness at workplaces [23]. These tools would provide users with several views of their workplace, allowing them to observe other social interactions and work that was happening around them. Dourish envisioned that these tools would allow for better collaboration and team building across distributed teams. XR can be used to support such tools that allow users to spontaneously connect with remote collaborators. Billinghurst suggests that can be used to enhance a physical space and create “intuitive 3D interfaces” that can support CSCW [5]. This is especially useful as it would allow for better remote synchronous in addition to asynchronous collaboration.

The Johansen Matrix (Figure 2.2) provides an intuitive breakdown of several common CSCW tasks [44]. These tasks can be classified as:

- Same-time/same-place: These are synchronous tasks that require all collaborators to be present. Examples of these tasks include: in-person meetings, round-table brainstorming sessions and white-boarding.
- Same-time/remote: These are synchronous tasks that do not require collaborators to be physically co-present. These tasks are usually performed through the use of a shared collaborative medium. Examples of these tasks include: messaging, video-conferencing and shared text-editing.

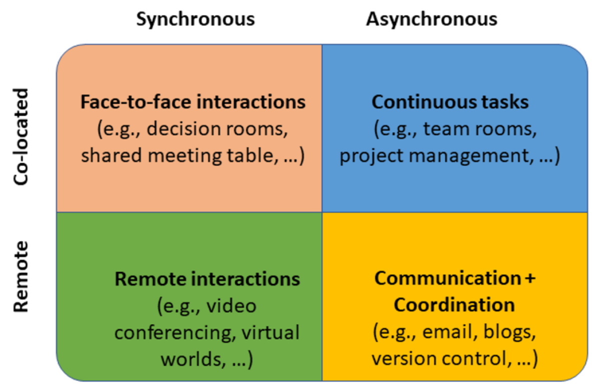


Figure 2.2: Johansen's CSCW Matrix [44].

- Asynchronous/same-place: These are larger on-going tasks where collaborators are not required to work at the same time. These tasks are location-dependent. Examples include: large public displays and project management.

- Asynchronous/remote: These tasks are primarily means of notifying collaborators or keeping track of events. Examples of these tasks include: email, calendars and version control.

A large part of CSCW involves envisioning trends in group work and workplace team dynamics. Workplace practices have been evolving over the last few decades, especially after the advent of computers. Workplace architecture can play an important role in determining local work practices. Strietz et al. [93] present an early conceptual framework into how data visualization can be better facilitated by workplace architecture, making the workplace environment a multi-modal interface to access relevant information. They leverage the ideas of ubiquitous computing [105] and invisible technology, stating that they will play an important role in making these adaptable architectural spaces a reality. Strietz et al. argue that despite information and communication technologies having reshaped work processes, the design of work spaces have largely remained the same. To remedy this they propose the idea of *cooperative buildings*: "flexible and dynamic environments" capable of "supporting and augmenting human communication and collaboration". These buildings operate around three dimensions: 1) the individual and the group, 2) real and virtual content, and 3) the local and global environment [92]. Finally, *cooperative buildings* would comprise of *roomware*: an amalgamation of everyday office room elements and computer-based information devices.

In additional to room-scale collaborative tools, individual work items can also be imbued with groupware-like properties. Fjeld et al. leverage Activity Theory - "the concept of tools mediating between subjects and objects" to develop tangible groupware [28]. This kind of tangible groupware is very close to the idea of collaborative XR environments. XR is unique and well-suited in this regard because it imbues everyday objects with additional digital properties, otherwise physically impossible to create. The added layer of tangibility makes interacting with virtual content more manageable. Further research into haptics for XR should make these interactions more seamless and help blend them better with regular interactions with real-world objects [55].

2.1.4 XR as a successor to modern day video conferencing applications

Video conferencing applications like Skype and Facetime are highly effective in providing visual and aural communication between users. These applications, however, are limited to two-dimensional representations of information in our three-dimensional world. While two-dimensional video conferencing applications are effective at providing visual and aural modes of remote communication, what they lack in is immersion. Users engaged in real-time video conversations do hear and see each other but they are also reminded of a distinct spatial boundary that is separating them, i.e. they are not having an immersive shared experience, as they would have had if they were physically co-located. Many non-verbal cues, including proper eye-contact, are lost over video-based communication systems [5]. Furthermore, users have no control or influence over the physical environment of their remote collaborators [31]. This tangible feeling of spatial separation can be a significant barrier in terms of enabling collaborative activities between remote users. Therefore, there is a need to expand on this mode of communication to facilitate remote interactions between participants in a shared three-dimensional space that is closer to natural human communication.

XR is a great medium to bring about this change in human remote communication. Over the last decade we've seen an increase in availability and affordability of XR devices like the Microsoft HoloLens [63], Magic Leap [54] and the Oculus Quest [78]. These devices are being used to develop a variety of applications in the domains of education [19, 89], training [36] and remote assistance [80]. XR provides uses with several real-world affordances like gestures, eye gaze tracking and shared spatial anchors that make communication easier. These affordances are familiar, intuitive and closer to in-person communication, making XR well-suited to supporting remote collaboration.

2.2 Interactions in XR

A large body of XR research focuses on interactions with real and virtual objects within virtual environments. Because of the complex nature of interactions within such spaces, XR devices must support multi-modal forms of interaction, i.e. they must provide for more than one way of interacting with the XR environment. Failure to provide multiple ways of interacting with XR content can lead to poor task performance and dissatisfaction [37]. XR interactions need to be intuitive and easy to learn. To address this need, in Section 4.2, we propose an approach to incorporating embodied interaction in XR environments.

2.2.1 Multi-modal interactions

Currently, XR devices provide limited modes of interaction with the XR environment, largely because of the relative newness of XR technology. These interactions are presently restricted to gaze-based interactions, gesture-based interactions and voice command-based interactions. However, these kinds of interactions might soon prove to be limited in their ability to handle 3D content [6]. As XR devices begin to see utility in more complex environments there will be a growing need for more sophisticated, multi-modal interactions with XR environments.

2.2.2 Embodied interactions

In an extended reality environment, embodied interactions refer to the ability of a user to interact with virtual objects by using their physical body as an input modality. This is usually achieved via the use of an external tracking device that records the user's pose in real-time and conveys this information to the virtual environment. The virtual environment in turn uses this information to update the user's state inside the environment. Embodied interactions allow the user to take advantage of affordances and embodied cognition in an XR environment, allowing them to interact with the environment in a more natural, intuitive manner [47].

Studies on embodied interactions in XR suggest improved usability and better task performance in collaborative activities. Minatani et al. developed a remote augmented reality system that allows people in far off locations to share a common augmented reality space [66]. The authors developed a method to represent user's upper body and hands on the table to allow to remote users to play a game of Othello, as if sitting across from each other. Peters et al. [74] describe their approach the challenge of connecting different hardware and software components in XR/VR. They made use of Unity 3D Cluster Rendering, distributed meetings and team collaboration. They describe the development of the AMELIO project, a serious, team collaboration game in XR with naturalistic interaction among users. They used WorldViz equipment for user input and Volfoni equipment for stereoscopic vision and user feedback.

2.2.3 Relating multi-modal and embodied interactions

The novel nature of extended reality makes interactions within XR environments unlike most other human-computer interaction modes that we are familiar with today. These interactions must be both realistic enough to be intuitive and different enough from physical interactions to facilitate interactions with objects that would not otherwise be possible in the physical world. *Blended Interactions* [43] advocate "blending the virtues of familiar concepts from the physical or social realm with the expressive power of the digital world in a considered manner" in order to retain the best properties of each. Because of the added complexity of extended reality, both embodied interactions and multi-modal interactions are necessary for an immersive experience in an XR environment. Embodied interactions are intuitive to the user, while multi-modal interactions provide users with more than one way of interacting with the XR environment. Together, these two modalities can provide for a familiar, yet unique experience for the user of an XR device.

Multi-modal interaction forms are important to maintain the usability of an XR application. As users are still unfamiliar with the different modes of interacting with XR environments, having the redundancy of multiple means to perform the same task allows the user more flexibility and control over the XR environment. In the future, we are likely to see more and more biological modalities (olfactory, haptic and auditory feedback) become integrated with XR applications [6], till eventually the lines between interactions within physical realities and virtual realities become indistinguishable, blending together in the truest sense of extended reality.

2.3 Remote collaboration and user presence

Remote communication technology has been evolving at a rapid pace in the 21st century. Zoom, Skype and WebEx are all widely used video conferencing platforms. Since the advent of COVID-19 in early 2020, remote work has become the norm. This trend will continue to persist into the future [24]. However, with a surge in video conferencing and work-from-home meetings, the phenomenon of "Zoom fatigue" has also become quite common. This has been attributed to the added cognitive load of having to focus more intently during video-meetings, while struggling with the dissonance of performing work-related tasks in a contrary environment [108].

Several factors make video conferencing strenuous include increased self-awareness, frequency of meetings and the lack of body language cues. As a result, people using video conferencing tools find it difficult to stay focused and work productivity suffers [103]. Therefore, there is a need to develop more engaging remote work tools that can capture and relay a shared sense of presence among colleagues. XR can help address this challenge by creating immersive meeting spaces for users to remotely interact within [4].

2.3.1 Eye gaze, gestures and presence

Social Visual Gaze plays an important role in human communication.Visual gaze and gestures help us feel more present during social interactions [48]. Vrzakova [104] conducted a mixed-media experiment where a buyer and seller negotiated the price of a car exchange. The effect of eye gaze in negotiations was examined. They observed that video conferencing degrades visual (eye gaze, hand and body gestures) and discourse cues (turn-taking), which limit coordination and trust building.

Video conferencing disrupts our natural ability to use gaze and gestures and detracts from a user's sense of presence. Sirkin et al. [86] developed a swiveling display screen as a mechanical proxy to indicate which direction a subject was looking, while in a video-conferencing meeting. They noted that subjects reported greater engagement in conversation and more accurate responses with the mechanical proxy than without. The lack of a sense of where a remote collaborator is looking makes it difficult to relay spatial information over video-conferencing platforms. XR environments address this limitation by providing users with a shared 3d environment to communicate with each other [26].

Video conferencing primarily involves information transfer over two channels of information: visual and auditory. Extensive research has been conducted into the use of these two mediums for remote collaboration. Doherty et al. [22] examined gaze aversion and shared-gaze in video-communication tools similar to Skype and Zoom while Alam et al. have explored how people communicate via speech-only interfaces [1]. It has also been found that attention varies with different tasks, with the average human attention span being around 8 seconds while browsing online content [7].

In this thesis, we explore the difference in user attention across video conferencing and online virtual environments. Users were given three different tasks to perform on Zoom and Mozilla Hubs. We measured and compared the differences in user eye-movement while completing these tasks. These differences are explored in more detail in Section 5.1.1.

2.3.2 Eye tracking analysis

Eye tracking is commonly used as a technique to study what people pay attention to in physical and digital contexts. Variations in eye movements, duration of fixations and patterns of visual searching provide important information about how a person is responding to visual stimulus [75]. Eye-trackers can used to create heatmaps for the most commonly looked at items on a screen. This information is useful for UX designers to engineer websites with ideal layouts that direct customer attention towards desired items.

Commonly measured variables are: fixations, saccades and gaze path [39]. Fixations are periods when the eyes are stationary and are taking in information. This can range from 60 to several hundred milliseconds. Saccades are rapid eye movements that take place between fixations. Gaze paths represent the combination of fixations and saccades over time.

Eye tracking analysis is based on an important assumption that there is a relationship between our fixations (what we are looking at), our gaze and what we are thinking about [77]. However, eye-tracking does have the following limitations:

- Eye-tracking can tell you what someone is looking at, but not what they perceive.
- It cannot tell you whether the visual attention is accompanied by positive or negative valence.
- It does not account for peripheral vision/perception.

Mansour and Kuhn [57] evaluated the relationship between gaze behaviours, social activity and clinical traits (such as anxiety and social interaction problems). They noted that participants engaging in a Skype conversation spent less time looking at a collaborator's eyes. Higher levels of gaze anxiety reduced the amount of time spent looking at face overall. They also noted that the duration of time spent talking by the collaborator was negatively correlated with gaze time on the face and eyes. Areas of interest were manually coded frame-by-frame (face, eyes). Mansour and Kuhn calculated the proportion dwell time for each interest area, which represents the proportion of time participants spent fixating each of the areas of interest (excluding saccades and blinks), relative to the duration of the entire trial segment. They used ANOVA analysis was used to compare differences between participant gaze behaviors.

Hessels et al. also examined the relationship between gaze behavior and Autism and Social Anxiety Traits [39]. They hypothesized that Autism Spectrum Disorder (ASD) and Social

Anxiety Disorder (SAD) traits are negatively correlated to fixation duration on the eyes, positively correlated to fixation duration on other parts of the face. [16]. Areas of interest were manually coded to include: eyes, nose and mouth. Dwell times were computed in order to determine how long and how often observers looked at the areas of interest.

Tsai et al. explored the role of visual attention in problem solving by examining student performance on a multiple-choice science problem using an eye-tracker [98]. They looked at the amount of time spent by students looking at the answer choices and the factors that would lead them to the correct result. They did this by creating specific focus areas (LookZones) and keeping track of student fixation durations. Two paired t-tests were used to compare fixation durations between selected/rejected options and relevant/irrelevant factors.

All of the above studies use eye-tracking to evaluate what users pay attention to. In Section 5.1 we describe the results of a user study designed to compare differences in user engagement across video conferencing and online virtual environments. Eye-tracking data was collected while participants performed three tasks. Like the studies described above, we defined focus zones/areas of interest and measured the amount of time participants spent focused on these areas. We performed ANOVA analysis to examine if there were significant differences in attention time between the two remote collaboration platforms.

2.4 The role of XR in CSCW

Current CSCW interfaces often introduce seams and discontinuities into the collaborative workspace [5]. XR, however, is well-suited for CSCW (Computer-Supported Cooperative Work) because it facilitates seamlessness and enhances reality. In an experiment conducted by Billinghurst where participants were asked to complete logic puzzles using 1) face-to-face collaboration with real objects, 2) co-located XR collaboration with virtual objects and 3) co-located projection based collaboration with virtual objects, Billighurst observed that

XR provided results that were very close in those observed in the face-to-face collaboration method.

In order to support remote collaboration in XR, virtual objects need to given both "objective" and "subjective" properties [40]. Objective properties are those that remain the same for all users, while subjective properties of virtual objects vary based on the task that a user is engaged in.

A popular technique for facilitating remote collaboration among users is via the use of virtual annotations or shared virtual landmarks [69]. Annotation typically involves layering additional virtual content over other virtual or physical objects [95]. Wither et al. [109] describe annotations in XR as "virtual information that describes in some way, and is registered to, an existing object". They classify annotations as having 1) a spatially dependent component, such as location and a 2) a spatially independent component, such as as color or pattern that makes them unique and distinguishable.

Early prototype applications that make use of annotations include: "VideoDraw" [96], "DigitalDesk" [106] and "SEMarbeta" [13]. Gauglitz et al. describe the creation and an experiment involving a system that provides remote-controlled XR annotations for virtual scene navigation. They make use of "monocular vision-based simultaneous localization and mapping" (SLAM) and surface-modeling to create spatial annotations that are viewable on a local users XR display. Doing so allows a remote collaborator to highlight objects of interest to the local user. Such annotations can be incredibly useful for remote assistance or hazardous tasks where an onsite agent and a subject expert/engineer are separated by large distances, yet need to collaborate to accomplish a task. Consider an extreme situation like an explosive-diffusal scene from a movie, where an onsite-agent is getting instructions from a remote-agent over a phone or a radio. In such a perilous situation, having the aid of visual annotations in addition to audio-based information can greatly benefit the on-site agent. An experimental study conducted by [31] involving a "car repair" task highlighted that the majority of the subjects reported better performance when they used the experimental prototype with annotations.

Szalavari et al. propose a client-server approach for collaborative XR applications with “Studierstube” where an “Environment Server” holds a database containing registration information and mobile object and display data [95]. This database is shared with each connected client device. Client devices have identical data, except for special customizations to the client’s local environment that are made by individual clients. Such a system is capable of providing a balance between shared XR content and localized individual content. “Studierstube” was further extended by Reitmayr and Schmalstieg [79] to create a collaborative game of XR chess. Our proposed framework for embodied interaction uses a similar client-server model to transfer user avatar positions to each listening XR client.

Bauer et al. [2] describe the use of an XR telepointer to direct the attention of remote collaborators. In their experiment, they make use of an HMD and a telepointer to facilitate collaborative tasks between participants. The results from their study demonstrate that pointer-based directions were far more successful in accomplishing navigation tasks than those involving verbal communication.

Poelman et al. [76] describe an application that allows crime scene investigators to collaborate with remotely-located subject matter experts. Their experiment involving a collaborative spatial analysis conducted in a staged crime scene showcase the merits of using XR for remote guidance.

Oyekoya et al. [72] describe their project (BEAMING) that “aims to transport people (visitors) from one physical place in the world to a destination to enable interaction with local people and objects there in real-time.” They use a motion capture suit, hand-tracker and EEG to capture visitor data. This data is then transmitted over the internet to the destination, which is simultaneously streamed to the visitor’s location as a reconstructed virtual environment in real time.

Brown et al. describe a multi-user collaborative experience that “allows web, virtual reality and physical visitors to share a museum visit together in real-time”. Their system involved the use of a PDA, an ultrasonic tracking system, a 3D representation of the gallery, 2D maps and hybrid XR exhibits. The system was built with the intent of providing the three main resources for awareness and interaction: a shared audio channel, awareness of each other’s location and orientation and a common information space. Results from their study indicate a favorable response from their audience towards this collaborative experience [9].

All of the articles above demonstrate work that has been done in the area of supporting remote collaborative activities using XR technology. These studies explored the use of XR in a variety of domains and illustrate the feasibility of XR for remote collaboration. However, despite its many advantages, certain challenges still prevent the large-scale adoption of XR. These challenges are described in the next section.

2.5 Challenges

XR technology has tremendous potential to revolutionize information access and human communication. The familiar and increased affordances provided by XR environments open up large application areas in collaboration, data visualization, and training. However, currently there exist a few challenges that are limiting the growth of XR. The five most critical challenges that are currently affecting XR are listed below:

1. High price point and limited applications: With most augmented reality devices costing around \$2000- \$3000 [54, 63], there is very little incentive for average consumers to purchase these devices. Furthermore, most of the applications that are being developed for XR devices are tailored for retail, visualization and training purposed. The combination of these two factors greatly reduce the appeal of XR technology for the regular consumers. In order to combat this, Microsoft, a leader in XR devices, deployed the Windows Mixed Reality

platform in 2017 and partnered with manufacturers like HP, Dell and Acer to come out with more consumer-friendly devices. These devices currently only support VR experiences but will eventually be able to support XR experiences as well. Virtual reality headsets like the Oculus Quest [78] are much more affordable, however, they are still used primarily for entertainment.

2. Imprecise or limited tracking outdoors: Current research into the tracking capabilities of augmented reality devices have shown them to be imprecise for navigational tasks outdoors [49] [38]. Developing XR hardware with minimal light sensitivity is an ongoing area of research [45, 73].

3. Limited field of view: A significant challenge to XR design is the limited-FOV of most modern HMDs [31, 49]. Milgram and Kishino describe this as the "keyhole effect" [65] where an excessively narrow field-of-view of XR HMD/HUDs cause the user to feel like they are peeping into the XR environment through a keyhole. Large FOVs are difficult to create as they lead to lower angular resolutions [113]. While the FOV of XR headsets today have greatly improved, they still haven't reached the point where XR can be seamlessly integrated into everyday-use. A potential solution to this problem is foveated rendering [10], which refers to the process of selectively rendering high resolution content only for the areas of the human eye that have high visual acuity.

4. Fatigue and eye strain: Head-mounted devices like the Microsoft HoloLens when worn for an extended period of time can cause discomfort in the form of eye strain and headaches [102]. This discomfort can often prevent a user from returning to a useful XR tool or training simulation. While this discomfort can be overcome through repeated exposure, it is initially difficult for a user to get used to such head-mounted devices.

5. Occlusion: Occlusion refers to the visual blocking of objects due to real or virtual objects placed in front of other objects. This is a common problem in XR and leads to a difficulty in separating foreground and background objects [111]. A potential solution to this issue is

to only render relevant stereo displays of 3D XR content at any given time. This will not only reduce the strain on the XR system, but will reduce the cognitive load on the user, as they won't have to be cognizant of less useful content [6].

Chapter 3

Problem Definition

The COVID-19 pandemic forced people to go about doing most of their daily work-related activities from home [81]. More than two years later, this trend still continues. In-person meetings around the world have been significantly reduced and most collaborative work has been moved online [108]. However, the subsequent overuse of virtual videoconferencing platforms has resulted in what we now know as “Zoom Fatigue” [68]. Users reported feeling tired, anxious, and worried as a result of overusing these remote collaboration platforms [112]. A major reason for “Zoom Fatigue” is because videoconferencing tends to disrupt the regular communication practices that we are accustomed to [68]. Videoconferencing also deprives us of a significant amount of contextual information (through the loss of body language cues and micro expressions) due to the physical separation between meeting participants.

Current remote work practices suffer from several of the issues presented above. Given the inherent limitations of videoconferencing, there is a need to make remote meetings more engaging, life-like and immersive. Browser-supported virtual environments like Mozilla Hubs and Virbela can serve as an effective medium to address some of the issues that affect video conferencing.

3.1 Research Questions

In this thesis we explore the following three research questions (RQs). RQ1 and RQ2 pertain to the comparison of online virtual environments with contemporary video conferencing platforms. RQ3 pertains to the evaluation of our proposed system design to support embodied interactions in XR environments.

3.1.1 Can online virtual environments can serve as a suitable medium for remote collaboration?

Online virtual environments like Mozilla Hubs and Virbela have been gaining popularity since the onset of the COVID-19 pandemic. These platforms have used to support large conferences and other social events (ISMAR 2020, IEEEVR 2020). While these platforms are capable of supporting large scale events, their potential as an alternative for video conferencing has not been adequately explored [91]. We seek to address this research gap by conducting a user study to explore the usability of online virtual environments for remote collaboration tasks. The SUS questionnaire [8] will be used to evaluate usability of the online virtual environment. The findings from this study can help in the design of better remote collaboration tools.

3.1.2 Can online virtual environments provide higher levels of presence and engagement for multi-user meetings when compared to contemporary video conferencing systems?

Virtual environments offer users an immersive an engaging platform to collaborate in. These environments contain special affordances like avatar customization and gestures to engage users and make them feel more present. Online virtual environments address the accessibility challenges posed by fully immersive XR platforms and can be an effective medium for remote collaboration. In this thesis we seek to explore how video conferencing tools and online virtual environments compare in facilitating user presence and engagement. User presence and engagement have been shown to have a strong correlation with directed attention [110]. We examine user attention (through eye-tracking software) while performing collaborative tasks on Zoom and Mozilla Hubs and compare user performance between the two platforms. A higher presence and engagement evaluation would indicate that online virtual environments can serve as a more immersive platform than video conferencing for remote collaboration.

3.1.3 Can 3D virtual environments (VEs) that support embodied interactions be used for remote synchronous meetings?

Our final research question seeks to address the feasibility of supporting embodied interactions in shared XR environments. Embodied interactions and gestures are intuitive and can enhance a user's sense of presence in a virtual environment [15, 52, 61]. To support these kinds of interactions, we proposed a system design in Section 4.2. This system design was implemented and network performance was evaluated. Network latency was used as the metric for this evaluation due to its important role in supporting user presence in shared virtual environments [20, 21, 35].

3.2 Research Challenges

The subject of presence in virtual environments has been studied for some time. Being a subjective quantity, presence can be difficult to measure. There is no consensus on the best way to measure presence for VR. However, over the last few decades researchers have come up with several means of trying to quantify what constitutes presence.

Quantifying presence into subjective and objective measures: Minsky defines presence as the quality of "being there" [67]. Schloerb [82] attempts to quantify telepresence in virtual environments by breaking it up into objective presence, the probability of a user completing a specified task and subjective presence, the probability that a user perceives themselves to be present in the virtual environment. Slater and Steed [87] present a similar breakdown for presence. They suggest evaluation methods including questionnaires for measuring the subjective phenomenon of presence and using observations (e.g. bio-signals) for measuring the objective phenomenon of presence.

Lombard and Ditton [53] present six components of presence: social richness, realism, transportation, immersion, presence as a social actor in the medium, presence of the medium as a social actor.

Farley and Steel [27] advocate for the utility of VR in education, by bringing up its ability to incorporate multiple sensory modalities. Students can have different learning styles: a) visual, b) auditory, c) kinesthetic, d) tactile/exploratory [29]. VR can provide students with experiences that cater to these different learning styles far more than the traditional presentation-based approach.

The user study proposed in this thesis incorporates both quantitative (eye-tracking) and qualitative measures (subjective, survey-based responses) for evaluating presence differences between video conferencing and virtual environments. A combination of these measures is used to identify whether online virtual environments can provide users with a greater sense of presence and engagement than contemporary video conferencing platforms.

Chapter 4

Proposed Approach

This chapter covers the methods used to evaluate the research questions posed in Chapter 3. Online virtual environments offer a promising solution to the affordability/accessibility challenge posed by XR hardware. These virtual environments, being browser-based can also help with reducing user fatigue that is more common with XR headsets. However, due to their novelty, online virtual environments have not been adequately compared and tested with other remote collaboration platforms [97]. In order to fill this research gap, this thesis seeks to explore whether online virtual environments can serve as suitable platforms for remote collaboration (RQ1). Additionally, we also compare user attention and engagement differences between traditional collaborative tools (video conferencing) and online virtual environments. Section 4.1 describes the user study that was conducted to evaluate RQ1 and RQ2.

Embodiment and interactions with digital content play a significant role in influencing user presence [83]. The third research question explored in this thesis involves designing and evaluating the feasibility of a system created to support platform-agnostic embodied interactions in XR (RQ3). This research question is addressed in Section 4.2

4.1 Exploring the differences in user attention and presence between video conferencing and browser-based virtual environments

This section describes the methods used to evaluate RQ1 and RQ2. We conducted a user study with 15 subjects in order to explore presence and engagement differences between Zoom and Mozilla Hubs. Presence in virtual environments has been defined as an "awareness phenomenon that requires directed attention" [110]. We base our approach to addressing RQ1 and RQ2 around this connection between presence and attention. In this study, we measure subject attention by using eye-tracking software and evaluating subject task performance.

4.1.1 Study Overview

The aim of this study is to compare subject attention and task performance in completing three similar tasks using a video conferencing (VCE) software and a 3D virtual environment (VE). The background scenario for all three tasks involves an research project involving a Advanced Trauma Life Support (ATLS) simulator.

During the study, subject attention will be measured objectively through task performance (correctness, involvement, speed) and eye-tracking. Presence will be evaluated as a subjective measure via the use of a standardized questionnaire (System Usability Scale [8]) once all the tasks have been completed. ANOVA analysis will be performed on the data (eye gaze duration, event recognition time and SUS scores) in order to determine if significant variations exist between the two platforms.

4.1.2 Study Tasks

1. Identifying the speaker in a multi-person meeting room:

Subjects will be asked to join a meeting room in a VCE and a VE. The meeting room will have other attendees. The subject will be asked to keep track of the conversation going on among the attendees. Conversations will never have more than one person speaking at a time. At pre-determined points during the conversation, a ding sound is played the subject is asked to identify the current speaker. The script for Task 1 can be found in Appendix A.3

Steps:

VCE: The researcher and the subject and other attendees begin in a Zoom meeting room. Only audio will be used for this task. subject switches to gallery view and is asked to keep track of the conversation. At a certain point the subject is asked to identify the most recent speaker (In gallery view, the current speaker's panel is highlighted with a green outline). This is repeated 3 more times.

VE: The researcher and the subject enter the Mozilla Hubs environment as avatars. Other meeting attendees are also present as avatars. The subject observes a conversation between the avatars. At a certain point the subject is asked to identify the most recent speaker. The current speaker's audio volume changes based on their distance from the subject. Additionally, there is a head-bobbing animation that accompanies speaker audio. A combination of these two can be used to identify the speaker. This is repeated 3 more times.

Measurements:

1. **Eye gaze duration:** Ths subject's eye gaze will be recorded for this task. The researcher will note the amount of time the subject's gaze was on the panel/avatar of the current speaker.

2. **Event recognition time:** How long does it take for the subject to respond with the name of the most recent speaker? The duration between when the ding sound is played

and when the subject responds will be recorded by the researcher.

3. **Accuracy:** Does the subject correctly identify the most recent speaker? The response provided will be noted by the researcher.

2. Getting the speaker's attention:

Raised hands are commonly used during classroom lectures/presentations, to draw the speaker's attention. In this task, the subject will be asked to present some information about the ATLS simulator using a VCE/VE. During the presentation, one or more attendees will raise their hand to ask the subject a question. The subject needs to respond to these attendees as soon as they see a raised hand. Speed of noticing and responding to attendee questions will be the objective measure evaluated in this task. The slides and questions for Task 2 can be found in Appendix A.4

Steps:

VCE: The researcher and the subject and other attendees begin in a Zoom meeting room. The subject is presented with a passage from a book and made to read it to the rest of the meeting attendees. Attendees will periodically raise their hands to ask a question. These raised hands will be seen as blue icons next to attendee names. The subject needs to respond to these attendees as soon as possible. If multiple hands are raised, subjects should try to answer questions in the order that the hands were raised. The subject can expect to receive simple questions like: "What is the last step of the ATLS?", "What is wrong with the patient?" The subjects should respond to these questions appropriately. subjects will be judged on their speed of reaction to the attendee question prompts.

VE: The researcher and the subject and other attendees begin in a VE. The setup is similar to that described in the VCE. The difference between them being that in the VE, the subject will see the raised hand icon above the attendee's avatar. The subject needs to address the attendee's questions appropriately and will be evaluated in a similar manner to the VCE task.

Measurements:

1. **Event recognition time:** How much time does the subject take to notice that a hand has been raised?

3. Space and Presence - speak up from left to right:

This task explores the spatial component of presence. In VCEs and VEs, meeting attendees occupy a certain "space", just like they do in real-world meetings. In a VE for instance, attendee avatars are situated at a certain position (x,y,z) within the VE world-space. In a VCE, attendees appear as panels on a screen. The space that an attendee occupies in a VCE/VE provides a spatial anchor for other meeting attendees to direct their attention towards, while addressing them. In real-world meetings, sometimes a moderator might ask attendees to speak up or present in a certain order based on their position in the meeting room (say, from left to right with respect to the moderator). This spatial affordance is limited in VCEs, while VEs come quite close in providing the same spatial presence to meeting attendees. The efficacy of this spatial presence for establishing an order (of speaking/presenting) is explored in this task. The prompts used for this task can be found in Appendix A.5

Steps:

VCE: The researcher, the subject and other attendees begin in a Zoom meeting room. The researcher provides a prompt to all of the meeting attendees and specifies the order in which everyone is supposed to respond. This order of response may be alphabetical or reverse alphabetical. This order is fixed for each prompt. All the meeting participants, including the researcher, then go around stating their response to the prompt. The study subject must also determine their turn to speak based on the order provided for the prompt.

VE: The researcher and the subject and other attendees begin in a VE. The setup is similar to

Fig: User study setup (Zoom).

Fig: User study setup (Mozilla Hubs).

Figure 4.1: User study setup: Zoom (Left), Mozilla Hubs (Right)

the VCE, except all the meeting attendee avatars are situated around a table. The researcher specifies the prompt and the response order. All the meeting participants, including the researcher, then go around stating their response to the prompt. The study subject must also determine their turn to speak based on the order provided for the prompt.

Measurements:

1. **Event recognition Time:** How much time does the subject take respond on their turn?

2. **Eye gaze duration:** How much of the time does the subject's gaze focus on the current speaker?

4.1.3 Study Setup

The entire study is expected to take about 60 minutes for each subject. Subjects will be expected to arrive at the study site at their allotted time slot. The researcher will take the subject to the 1st floor meeting room and will provide them with the computer that they will be use to complete the experiment. The subject will then be informed about and given

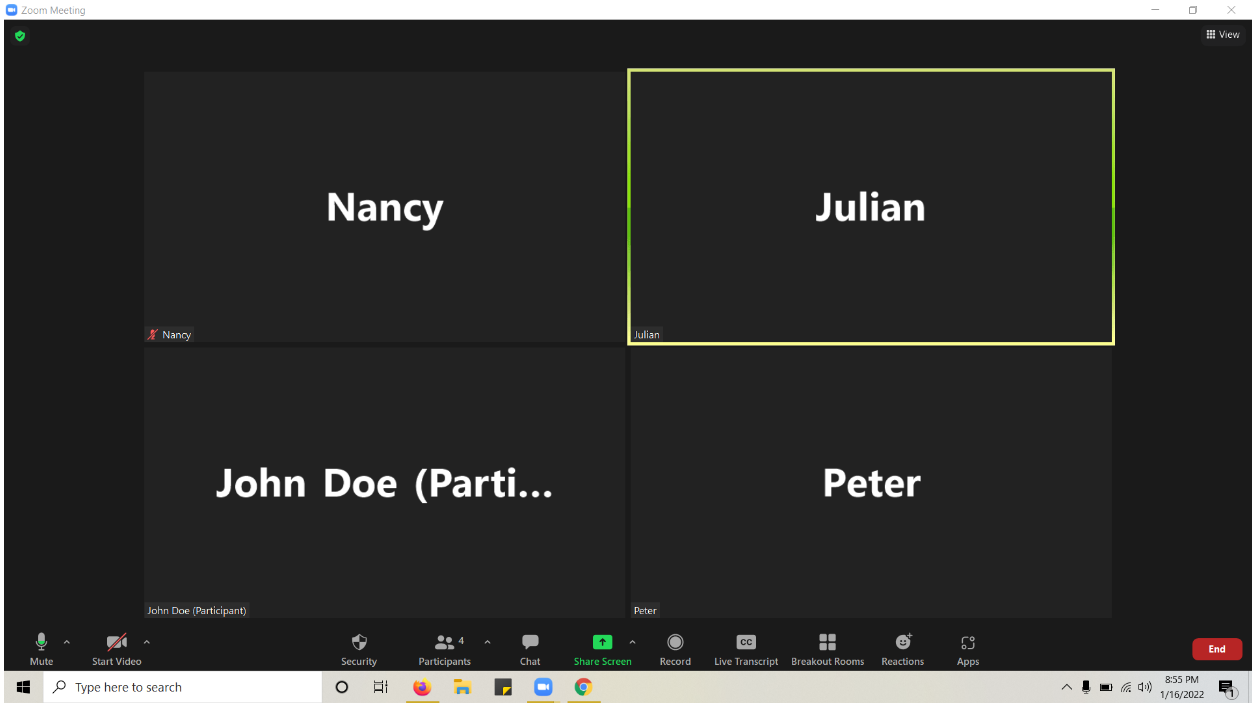


Figure 4.2: Participant point-of-view (Zoom)

directions to help them complete the consent process. Following this, the subject will be provided with a background questionnaire to complete, which we expect would take them under five minutes to finish. Shortly after this a member of the research study group will explain to the subject all the details about the VE/VE environments that they would need to know in order to successfully complete the study experiment. After this discussion, the subject will begin with the study and proceed to complete each of the three tasks listed in Section 4.1.2. The study setup for both Zoom and Mozilla Hubs tasks can be seen in Figure 4.1. The subject will complete the tasks for both the VCE (Figure 4.2) and VE (4.3) environments. Each task should take between 10-15 minutes to complete. Once the subjects have completed all three tasks, they will be asked to fill out a questionnaire asking them about their experience with the two platforms (VCE/VE). Finally, the subject will be thanked for their time and provided with the agreed compensation. This will conclude the study session for the subject.

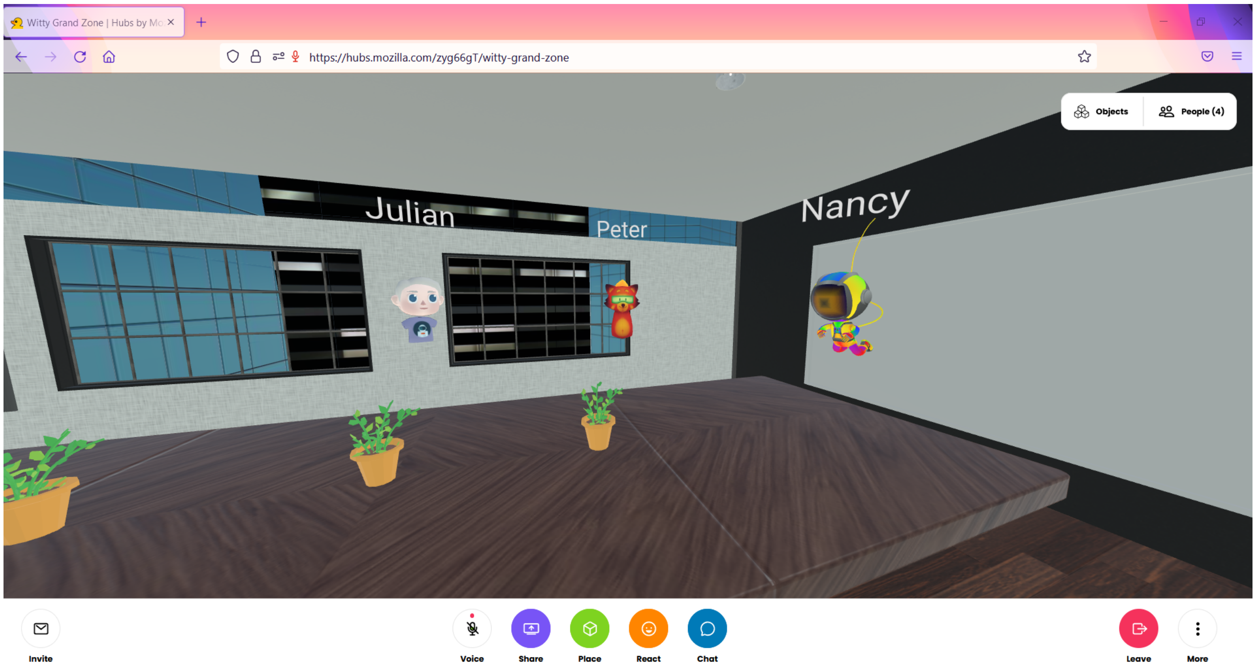


Figure 4.3: Participant point-of-view (Mozilla Hubs)

4.1.4 Evaluation

After all three tasks are completed, subjects will be asked to fill out a questionnaire to evaluate their experience with using both platforms by grading them on various parameters, including: usability, convenience, memorability and usefulness. This questionnaire will be based on the System Usability Scale (SUS) [8]. SUS is a widely used tool for measuring the usability of a variety of products and services. It consists of a ten-item questionnaire, each with five response options ranging from 'Strongly agree' to 'Strongly disagree'. Subject responses are given numeric values from 0-4 based on their selection.The sum of these values are then scaled to obtain a final score out of one hundred. A score of over eighty points out of one hundred will be considered as a favorable response.

We will collect both qualitative and quantitative data during the study. We will ask the subjects to complete a pre and post-questionnaire to collect demographic information and to obtain user feedback. Quantitative data collected will include the time taken by a subject to complete a task and the accuracy of responses. Eye-tracking software will be used to measure a subject's direction of attention during certain study tasks. Qualitative data collected through the study will explore participant levels of presence and immersion in the VCE/VE tasks. This data will allow us to compare and gauge the effectiveness of VEs over VCEs for remote collaboration.

4.2 Exploring the feasibility of 3D virtual environments with embodied interactions

Current remote work practices, using video conferencing, can be exhausting and impersonal. The lack of body language, eye gaze and gestural cues make it harder for users to focus in remote meetings. Given the inherent limitations of videoconferencing, our goal is to try and make remote meetings more engaging, life-like and immersive. Embodied interactions can help bridge the gap between virtual and in-person meetings. With the recent advances in XR infrastructure, motion-tracking sensors and network connectivity, it is possible to incorporate human body movement and gestures into meetings in virtual environments.

We describe an approach to designing collaborative virtual environments that can be used for a variety of purposes, including education and remote meetings [59]. Our proposed system architecture consists of three components: Body Tracking, Communication, and XR Rendering. Each of these components are flexible and are not bound to specific hardware. Figure 4.4 depicts the use of these the components to create a shared XR space for two users. This system can be further expanded to support additional users. The three components are described in more detail below.

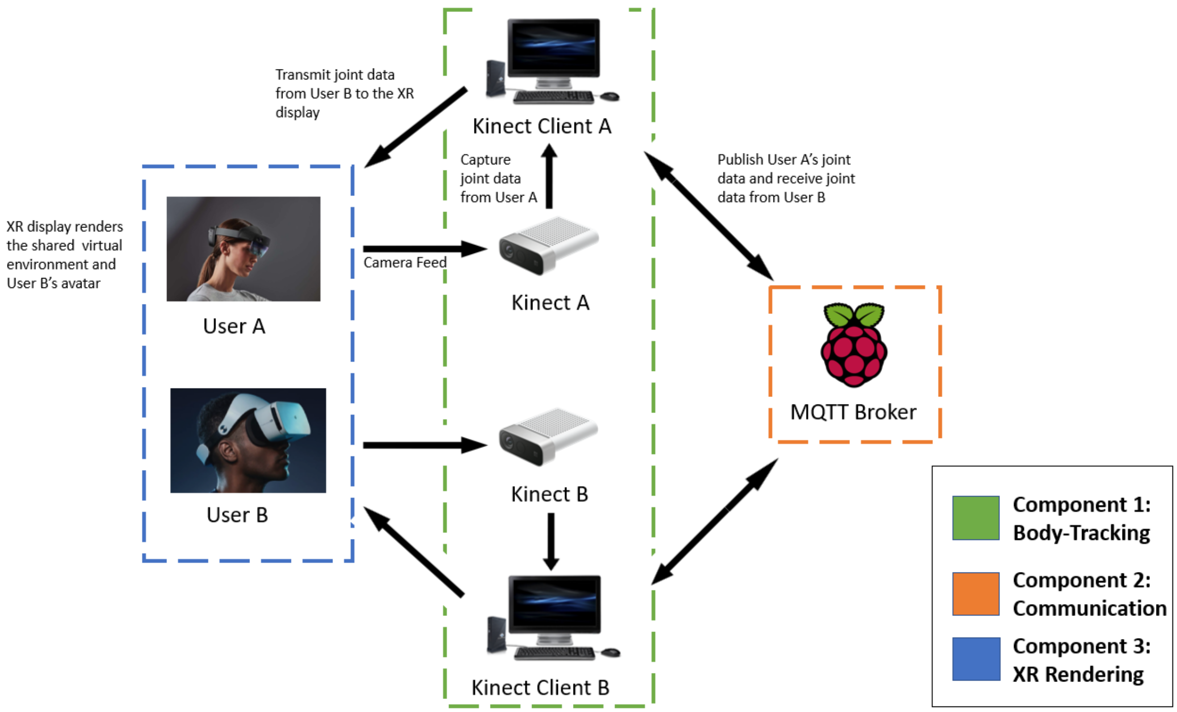


Figure 4.4: System Architecture Diagram. 1) Component 1: Body Tracking — captures a local user's body frames and identifies joints. 2) Component 2: Communication — transfers the messages from the Kinect client to other remote user Unity client applications. 3) Component 3: XR Rendering — Renders each user's avatar at their respective positions in the virtual environment. The XR device allows the user to see and interact with other users and the shared environment.

4.2.1 Component 1: Body Tracking

Advancements in computer vision and object-recognition technology have led to the commercial availability of several body-tracking sensors. These sensors are capable of tracking human skeleton joints in real-time. Real-time tracking can be implemented via marker-based (Qualysis, OptiTrack) or markerless (Microsoft Kinect, Intel RealSense) techniques. While marker-based body trackers provide more robust tracking [101] and have larger coverage areas, they are not practical for personal use at home or in a workplace. Marker-less body trackers are smaller, portable and more cost-effective overall.

Our system design makes use of marker-less body trackers to capture user skeleton data.

The skeleton joint data for user movements is repeatedly captured and used to rig virtual avatars that replicate their body movement in the XR application. Skeleton joint data for important body parts (head, arms, pelvis, legs, etc.) are continuously streamed to the XR applications of all the users in the meeting. This streaming is handled by our communication component.

4.2.2 Component 2: Communication

The skeleton joint data collected by the body tracking device includes the Cartesian coordinates (x, y, z) and the rotational quaternion coordinates (w, x, y, z) for each tracked body joint. Most marker-less body tracking devices collect skeleton joint data for all the joints present in the human body and use them to animate avatar models. However, this is a tremendous amount of data to send over a network. It can overload the communication network and result in significantly higher latency cost. To avoid this we recommend streaming only the most relevant user joint data through the communication channels.

For our design, we decided to use the MQTT network protocol [42] to support our communication component. MQTT is an OASIS standard messaging protocol for Internet of Things. It is a lightweight publish/subscribe network protocol that can be used to establish communication between remote devices.

Each user's skeleton joint data is streamed on a unique MQTT channel (topic) by their respective body-tracker client applications. The body-tracker client streams their skeleton joint data to a centralized MQTT broker. The broker then broadcasts this data to all the other users' XR client applications that are listening for updates on these channels. XR client applications run on each user's XR display component. Upon receiving new messages from the broker, these XR applications use the passed skeletal joint data to update each user's location within the shared remote meeting environment.

4.2.3 Component 3: XR Rendering

The XR component consists of an always-on application (developed on a game engine like Unity) that runs on the XR display hardware. This application displays the current state of shared virtual environment, along with the avatars of all remote meeting subjects. Virtual environments can be designed to replicate traditional collaborative environments (classrooms, conference rooms, office spaces, etc.). The XR application is responsible for controlling everything that a user sees and does in the shared virtual environment. The application updates the user's position based on their movement.

It also registers any embodied interaction between the user and an environment object and sends this to a designated channel on the communication component. For instance, if the user moves a chair in shared virtual classroom environment, this movement is also propagated to the XR applications of each other remotely connected user. Finally, the XR application also receives streamed data from the communication component (other user body movements and environment object interactions) and updates the state of each of these entities in the environment.

4.2.4 Implementation

In this section, we describe our proof concept implementation based on the approach described in the previous section. Given the impact of COVID-19 and the switch to online modes of education, we decided to go with a virtual classroom scenario for our implementation prototype. We envisioned this to be a space where students can join in remotely to attend lectures, work on group projects and interact with faculty members.

Figure 4.5: Left: The physical classroom at the Northern Virginia Center. Right: The virtual classroom environment Unity application.

4.2.5 Virtual classroom environment overview

The first virtual classroom that we designed was inspired by a real classroom located in the Northern Virginia Center, a satellite campus of Virginia Tech. This virtual classroom was designed using the Unity game engine. Images of the real and virtual classrooms can be seen in Figure 4.5.

After designing this classroom environment to closely resemble it's real-world counterpart, we re-configured its layout to allow users to interact in multiple environment settings. We created a total of four different classroom layouts:

- Small Classroom: The original classroom design from the Northern Virginia Center. This classroom environment is built to match the scale of its physical counterpart. It is best suited for small class sizes with more interpersonal engagement.
- Small Classroom post-COVID: This classroom environment was developed to reflect the"six-feet apart" social-distancing policy mandated at locations of in-person instruction, as a preventative measure against COVID-19.
- Conference Room: A modified classroom designed to resemble meeting rooms. The

desks in this layout have been brought together to create a large table in the middle of the room.

- Large Classroom: This virtual environment was developed to seat a large number of students. This scenario is appropriate for lecture halls with minimal group-based discussions.

To speed up future development, we also created a script that would allow developers to automatically generate their desired classroom layouts by specifying their required layout parameters in the form of a JSON input file. A demo video highlighting the navigation and interaction aspects of our prototype implementation can be found here [58].

4.2.6 Application Components

Our system design comprises of three major components: Body-Tracking, Communication and XR Rendering. For our reference implementation, we developed this system using the following devices:

Body-Tracking

We selected the Microsoft Azure Kinect as the body-tracking device for our implementation. Each remote user is equipped with a Kinect device and a dedicated client PC. Each user's Kinect captures their skeleton joints and updates the user's own avatar inside a Unity application running on the XR component. The joints tracked by the Kinect sensor are shown in Figure 4.6. The Kinect client also streams the user's joint information to all other user's within the same shared virtual classroom environment (via the communication component). In our prototype implementation, we tested this out by having two remote users in the same virtual classroom environment at the same time.

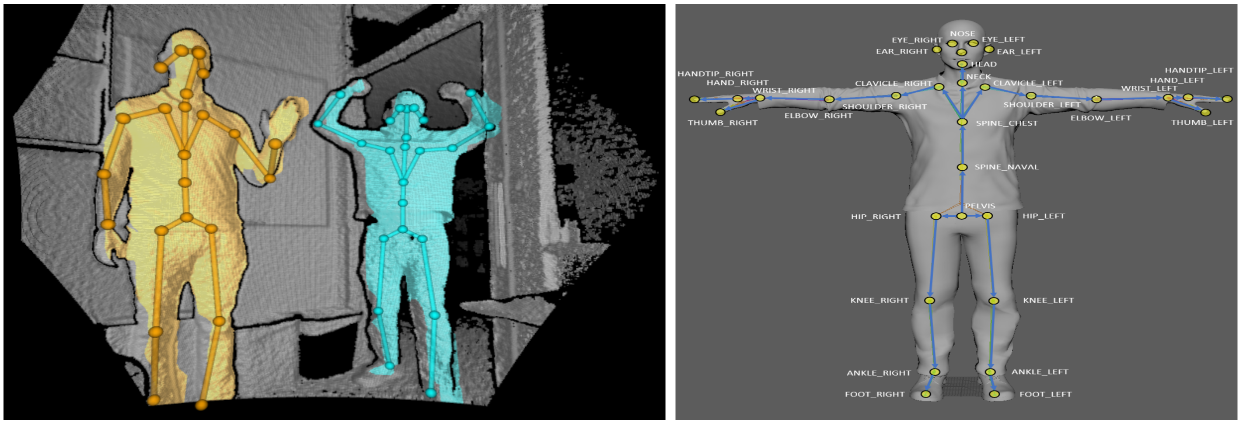


Figure 4.6: Left: Two user skeleton joints tracked by the Kinect. Right: All joints capable of being tracked by the Azure Kinect sensor [64].

Communication

The communication component serves as the backbone of this implementation. Due to the high volume of messages that need to be sent over the network, picking an efficient, lightweight messaging protocol is imperative. We used the MQTT protocol as the communication back-end for our implementation. The MQTT protocol follows a publish-subscribe messaging pattern. The MQTT broker is responsible for directing all of the shared virtual environment messages to their appropriate communication channels. A dedicated Raspberry Pi device served as our MQTT broker.

Kinect client PCs are always running an application that connects them to the MQTT broker and allows them to 'publish' (send) messages to specific communication channels. In our implementation, for example, coordinates for remote user A, were streamed to a channel called 'kinectCoords', while coordinates for user B were streamed to a channel called 'kinectCoords2'. Any device connected to the same MQTT broker can 'subscribe' (listen) to messages sent on these channels.

Whenever a user is tracked by the Kinect, their client PC applications publish their tracked skeletal joints along with the user's id to the MQTT broker on one of these channels. The broker in turn transmits these messages to all other devices that are listening to these channels. These include the XR applications running on each remote user's XR component that take the skeletal joint messages and update the corresponding user's avatar inside each user's local version of the shared virtual classroom environment.

XR Rendering

For each remotely connected user, the XR Rendering component is responsible for displaying the user's local view of the shared virtual environment. This component comprises of display hardware such as AR/VR/MR headsets. However, regular PC screens/displays can also be used to present the shared virtual environment (for improved accessibility). This component is always running a Unity application containing the current state of the shared virtual environment. The application is responsible for listening to messages sent via the MQTT broker and accurately updating the positions of all the users in the shared virtual environment. The positions of all interactable objects in the environment are also appropriately updated by this application. For our prototype application, the Kinect client PCs also ran the Unity XR application and displayed the local view of the shared environment for each remote user.

Chapter 5

Results and Discussion

This chapter presents the results obtained from implementing and testing the two approaches described in Chapter 4. Section 5.1 describes results pertaining to RQ1 and RQ2. Section 5.2 describes the results obtained by testing the system proposed in Section 4.2.

5.1 Comparing differences in presence and engagement between video conferencing and online virtual environments

The results presented in this section were obtained by conducting a user study with fifteen subjects. The study was conducted at 2202 Kraft Drive, Virginia Tech. The study sessions took approximately 1hr to complete for each subject. Quantitative and qualitative results from the study are presented in Sections 5.1.1 and 5.1.2.

Study Subjects: Study subjects were recruited by posting emails on several Virginia Tech email listservs. All subjects were Virginia Tech students who were above 18 years of age. Subject demographics are described in more detail in Table 5.1.

Table 5.1: Subject Demographics

Age		**Gender**		**Language**		**Major**	
21-25	26-30	M	F	English	Other	Computer Science	Other
11	4	10	5	6	9	7	8

Subject experience levels with Video Conferencing Tools and Virtual Environments are described in Table 5.2. The majority of the subjects had significant experience with Video Conferencing, but very few had used Virtual Environments for remote collaboration prior to the study.

Table 5.2: Subject Experience with Remote Collaboration Tools

	Video Conferencing			**Virtual Environments**		
	Very	Somewhat	Unfamiliar	Very	Somewhat	Unfamiliar
Familiarity	13	1	1	0	3	12
Hours used per week	< 2 hrs	2-8 hrs	> 8 hrs	< 2 hrs	2-8 hrs	> 8 hrs
Number of subjects	2	11	2	15	0	0

Results

5.1.1 Quantitative Analysis

This section covers the quantitative results that were obtained from running the study with 15 subjects. Subjects were asked to perform three identical tasks across Zoom and Mozilla Hubs in order to examine differences in subject attention and engagement across the two platforms. The quantitative parameters that were measured include: eye gaze duration and event recognition time. Task completion time was also recorded in order to provide context for the eye gaze and event recognition parameters. Box plots have been used throughout this section to present the data obtained from the study.

Box Plot Description: This section contains three box plots representing the data for task completion time, eye gaze duration and event recognition time. Data from the tasks on Zoom (video conferencing) is shown in blue, while data from Mozilla Hubs (virtual environment) is shown in orange. The x-axis for the box plots represent the tasks performed, while the y-axis of plots represents the amount of time taken for a task. The unit on the y-axis varies with the parameter being explored. For task completion time, this is time in minutes. For eye gaze duration, time is measured as a percentage of the task completion time. And finally for event recognition time, time is measured in seconds.

The elements of the box plot include [30, 99]:

- Q1: The lower bound of the box representing the 25th percentile value for the data.
- Q3: The upper bound of the box representing the 75th percentile value for the data.
- Q2: The solid line inside the box representing the median value for the data.
- Interquartile range (IQR): The side of the box, along the y-axis, representing the spread of the data between Q1 and Q3.
- Mean: The mean of the data by an ’X’.
- Whiskers: Lines extending above and below the box representing the theoretical maximum and minimum values for the data. The maximum is calculated as Q3 + (1.5 * IQR), while the minimum is calculated to be Q1 - (1.5 * IQR).
- : Outliers: Colored circles above or below the whiskers representing data values that lie beyond the theoretical maximum and minimum values.

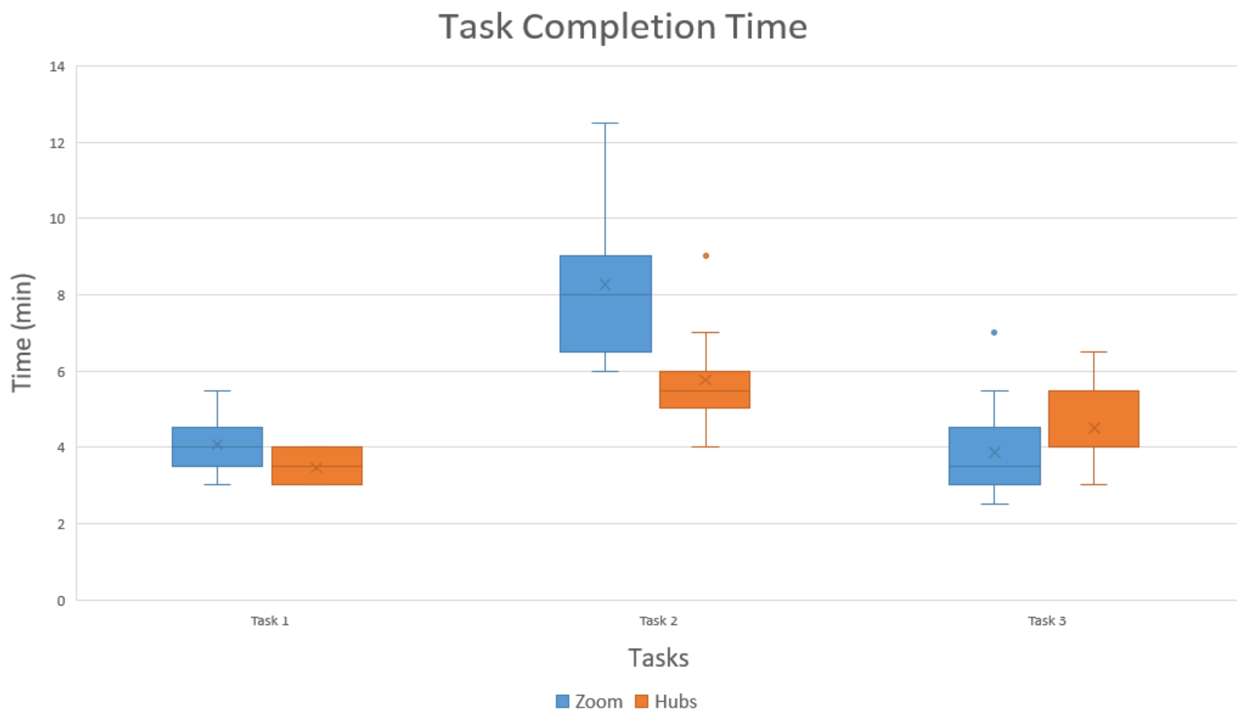


Figure 5.1: Task Completion time across Zoom and Mozilla Hubs.

The box plots in this section were created using Microsoft Excel. Excel uses the IQR method to calculate the theoretical maximum, minimum values and the outlier values [62]. These outliers can be seen in Figure 5.1 and Figure 5.5 when subjects exceeded the 75th percentile value by more than 1.5 times the IQR value. The following sections will further expand on the box plot values and the ANOVA analysis for each quantitative parameter that was measured during the study.

Task Completion Time: Task completion time was defined as the total amount of time that each subject spent completing a task on a specific platform. Average task completion time for all three tasks, across both platforms can be seen in Figure 5.1. Subjects took an average of 4.06 minutes to complete Task 1 on Zoom and 3.5 minutes to complete the same task in Mozilla Hubs. Task 2 took an average of 8.2 minutes in Zoom and 5.76 minutes on Mozilla Hubs. Task 3 took an average of 3.87 minutes on Zoom and 4.5 minutes on Zoom.

Subject eye gaze was used as the primary metric for measuring attention in Zoom (Video Conferencing) and Mozilla Hubs (Virtual Environment). Subjects wore a Microsoft HoloLens 2 headset while completing the three study tasks. The headset was calibrated for each subject prior to starting the three tasks. An eye-tracking application was developed to constantly monitor what the subject was looking at. The application denoted a subjects item of visual focus with a red dot. This red dot can be seen in Figure 5.2. The eye tracking application was developed using the Windows Mixed Reality Toolkit [63].

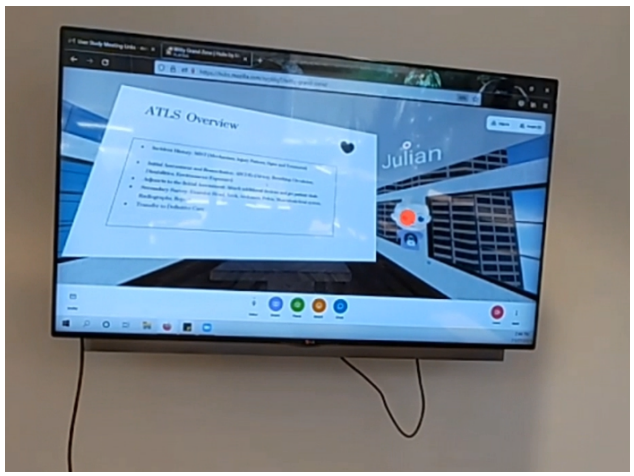


Figure 5.2: Subject POV through the HoloLens2. The red dot indicates the subject's current point of visual focus.

Eye tracking results

Subject eye gaze duration as a percentage of total time spent per task was recorded. Gaze duration was calculated based on the amount of time subjects spent looking at visual fixtures representing remote collaborators (i.e., user name panels on Zoom, user avatars on Mozilla Hubs). Higher gaze duration per task would indicate that subjects were looking at or paying closer attention to the events going on during the task. For example, in a Zoom meeting if a subject was actively looking at each collaborators speaker panel while they were speaking, they would receive a higher gaze duration value. Similarly on Mozilla Hubs, subjects received

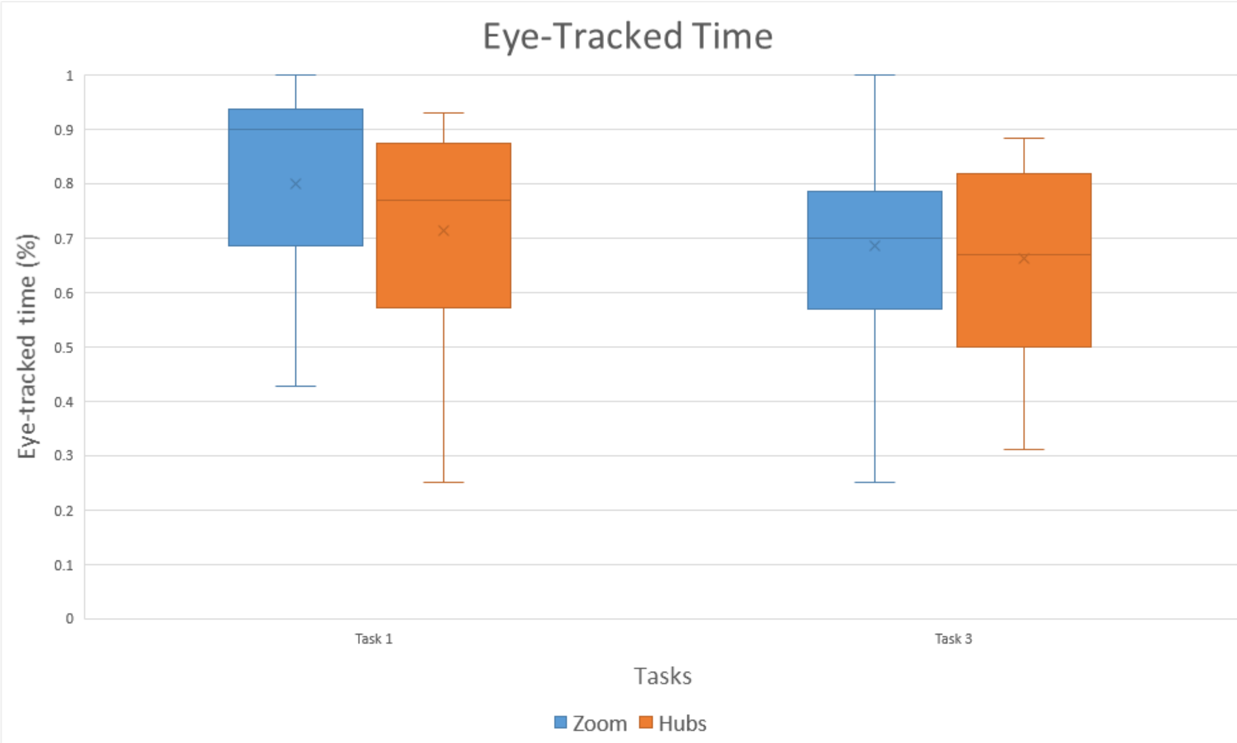


Figure 5.3: Eye gaze duration (%) for Zoom and Mozilla Hubs.

a higher gaze duration score if they spent more time looking at collaborator avatars, as they were speaking.

The average gaze duration per task of all fifteen subjects can be seen in Figure 5.3. Only eye-tracking data from Task 1 and Task 3 were used for this comparison, since Task 2 required subjects to present content from a set of slides (this would produce similar results for Zoom and Mozilla Hubs). The average eye gaze duration for Task 1 was 80% on Zoom and 71.4% on Mozilla Hubs. For Task 3, the average gaze duration was 68.5% on Zoom and 66.2% on Mozilla Hubs.

ANOVA Analysis: One-way ANOVA results and t-test results for gaze duration can be seen in Figure 5.4. The p-value that was obtained for this comparison was 0.28, which is much greater than our alpha value of 0.05. Therefore, we fail to reject the null hypothesis. ANOVA results indicate that there is no significant difference in gaze duration between the

Oneway Anova

Summary of Fit

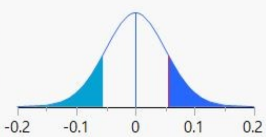
Rsquare	0.01989
Adj Rsquare	0.002992
Root Mean Square Error	0.196991
Mean of Response	0.715905
Observations (or Sum Wgts)	60

Pooled t Test

Zoom-Mozilla Hubs

Assuming equal variances

Difference	0.05518	t Ratio	1.084924
Std Err Dif	0.05086	DF	58
Upper CL Dif	0.15700	Prob > |t|	0.2824
Lower CL Dif	-0.04663	Prob > t	0.1412
Confidence	0.95	Prob < t	0.8588



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Platform	1	0.0456764	0.045676	1.1771	0.2824
Error	58	2.2507197	0.038806		
C. Total	59	2.2963961			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Mozilla Hubs	30	0.688313	0.03597	0.61632	0.76031
Zoom	30	0.743496	0.03597	0.67150	0.81549

Figure 5.4: ANOVA results for eye gaze duration (%).

two platforms for Task 1 and Task 3 of the study.

Event Recognition

Event recognition time was a secondary metric that was used to evaluate subject presence in both platforms. In all three user study tasks, subjects were required to respond to certain kinds of stimuli. In Task 1, subjects needed to respond by stating the name of the most recent speaker. In Task 2, subjects needed to notice questions from their collaborators. Subjects had to notice “raised hand” gestures on Zoom and Mozilla Hubs while presenting from the provided slides. Whenever a hand was raised by the subject’s collaborators, the subject was asked to pause their presentation and attempt to address the posed question as quickly as possible. This was the event for Task 2. Finally, for Task 3, subjects needed to note when it was their turn to respond to a prompt (according to a specific order that is specified before each prompt).

The average amount of time taken by all fifteen subjects to respond to the three task events

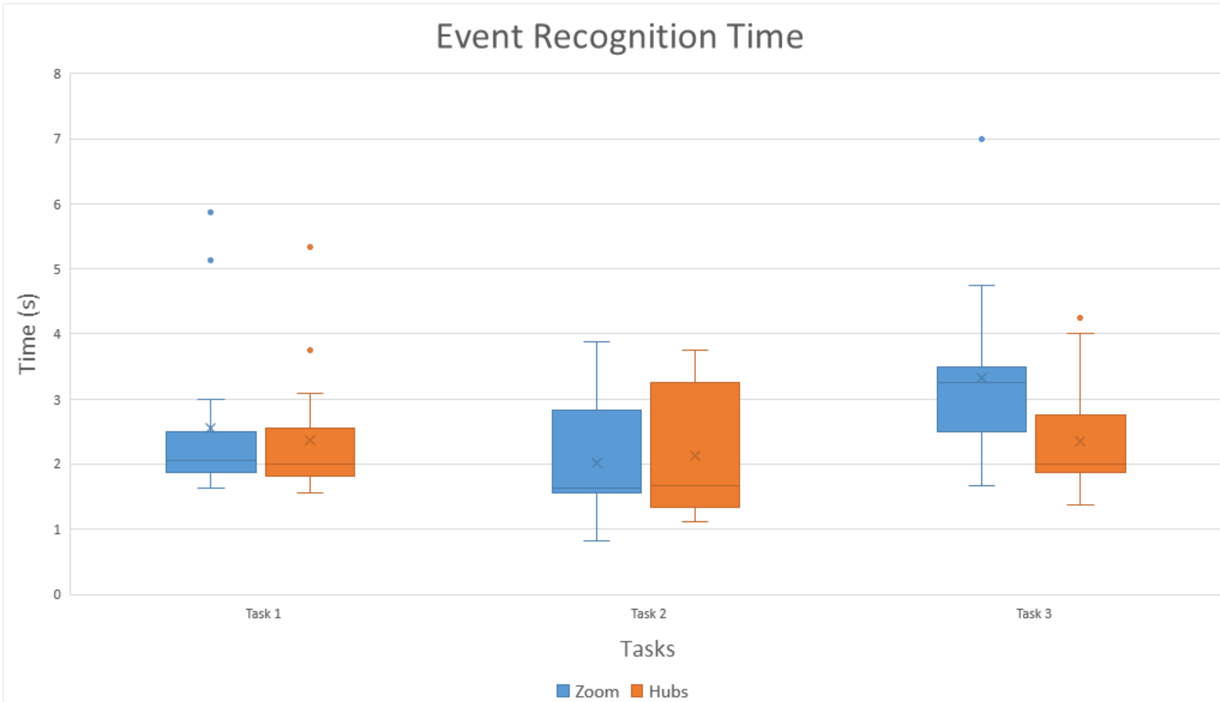
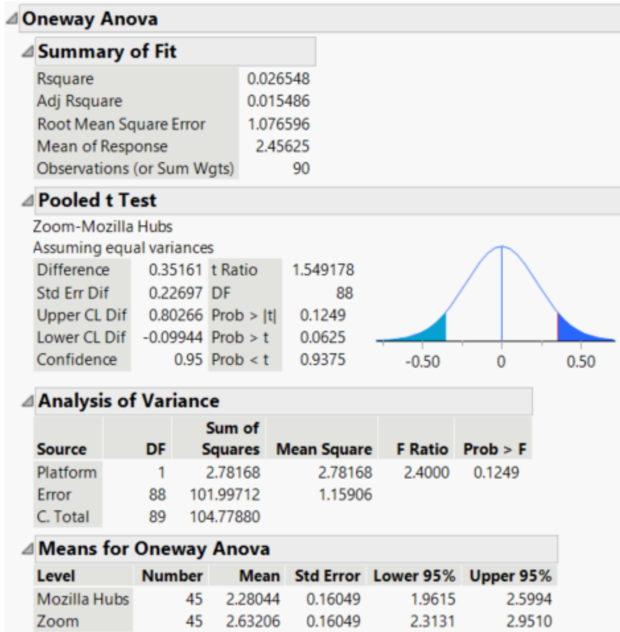


Figure 5.5: Event Recognition time for Zoom and Mozilla Hubs.

can be seen in Figure 5.5. The average event recognition time for Task 1 was 2.55 seconds on Zoom and 2.36 seconds on Mozilla Hubs. For Task 2, the average event recognition time was 2.01 seconds on Zoom and 2.18 seconds on Mozilla Hubs. For Task 2, the average event recognition time was 3.33 seconds on Zoom and 2.35 seconds on Mozilla Hubs.

Outliers: Outliers for event recognition time can be seen in Figure 5.5. Two large recognition time values can be seen in the plots for Task 1 (5.8s and 5.1s for Zoom, 5.73s and 3.75s for Mozilla Hubs). These outlier values were much larger in the case of Zoom, resulting in the mean (2.55s) being shifted upwards over the upper quartile of the box (2.5s). This variation is seen strongly because of the effect of this large outlier on a small dataset (15 data points). For Task 3, a single outlier can be seen in each of the plots (7s for Zoom and 4.25s for Mozilla Hubs). The large outlier value for Zoom has also shifted the mean much higher, as can be seen in the box plot (Mean value: 3.33s, Q3 value: 3.5s).



Oneway Anova

Summary of Fit

Rsquare	0.026548
Adj Rsquare	0.015486
Root Mean Square Error	1.076596
Mean of Response	2.45625
Observations (or Sum Wgts)	90

Pooled t Test

Zoom-Mozilla Hubs

Assuming equal variances

Difference	0.35161	t Ratio	1.549178
Std Err Dif	0.22697	DF	88
Upper CL Dif	0.80266	Prob > |t|	0.1249
Lower CL Dif	-0.09944	Prob > t	0.0625
Confidence	0.95	Prob < t	0.9375

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Platform	1	2.78168	2.78168	2.4000	0.1249
Error	88	101.99712	1.15906		
C. Total	89	104.77880			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Mozilla Hubs	45	2.28044	0.16049	1.9615	2.5994
Zoom	45	2.63206	0.16049	2.3131	2.9510

Figure 5.6: ANOVA results for Event Recognition time on Zoom and Mozilla Hubs.

ANOVA Analysis: One-way ANOVA results and t-test results for event recognition time can be seen in Figure 5.6. The p-value that was obtained for this comparison was 0.12, which is greater than our alpha value of 0.05. Therefore, we fail to reject the null hypothesis. ANOVA results indicate that there is no significant difference in event recognition time between the two platforms for Tasks 1, 2 and 3 of the study.

After completing all three study tasks, users were asked to fill out a post-study questionnaire A.2. This questionnaire is a modified version of the original System Usability Scale questionnaire developed by John Brooke in 1986 [8].

System Usability Scale

The System Usability Scale (SUS) is a standardized usability questionnaire used to evaluate software, websites and applications. It comprises of ten questions, each with five multiple choice options ranging from “Strongly Disagree” to “Strongly Agree”. Subjects are asked to respond to the ten questions based on their experience with the product being evaluated.

For this user study, the questionnaire was modified in the following manner:

- The questionnaire was split into two parts, each containing the original ten questions with the name of the product being evaluated changed to Zoom and Mozilla Hubs, respectively.
- An additional question was added to both parts: "I felt present when I was interacting with other people in Zoom/Mozilla Hubs".
- An extra free response question was added to both parts: "Please let us know about any feedback that you have about your experience with the Zoom/Mozilla Hubs portion of this study." The addition of this item enabled us to obtain specific feedback from the users that did not fit neatly into any of the other SUS question items.

SUS score calculation: Subject responses to each item on the SUS questionnaire are assigned raw score values out of 5 ("Strongly Disagree" = 1, "Strongly Agree" = 5). These raw scores are then converted into SUS Scores based on their sentiment. All odd numbered items are positive experience questions, while the even numbered questions are negative experience questions. Odd numbered question raw scores are reduced by 1 (i.e. Odd SUS Score = Odd Raw Score -1), while even numbered question raw scores are subtracted from 5 (i.e. Even SUS Score = 5 - Even Raw Score). The SUS scores from both odd and even questions are added together to get a composite score out of 40. This score is then scaled to 100 by multiplying the composite score by 2.5. A final score of over 68 indicates fair usability, while a score of over 80 indicates excellent usability.

The average SUS scores for Zoom and Mozilla Hubs obtained from the modified study questionnaire are given below in Table 5.3

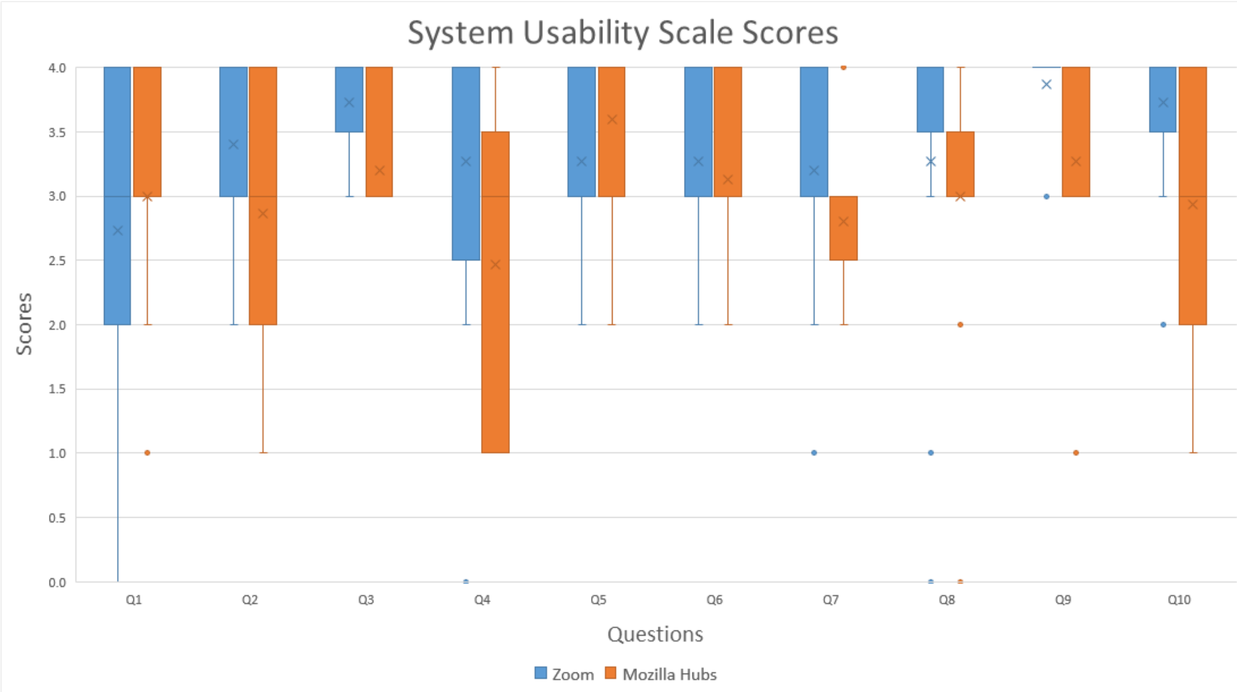


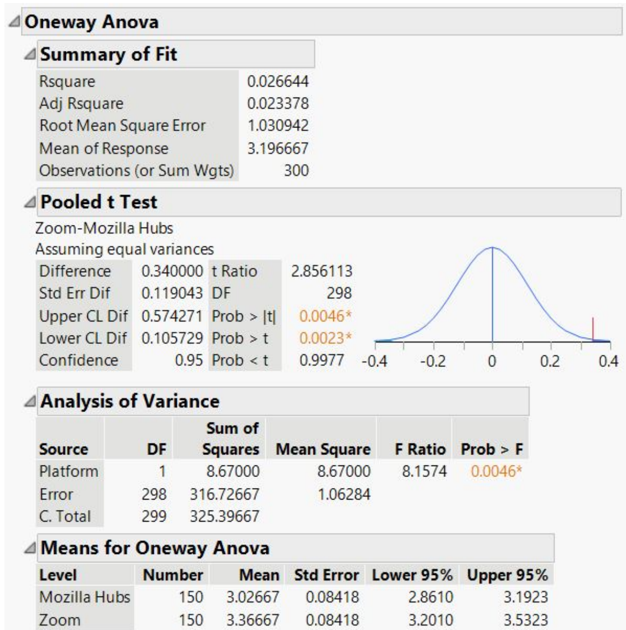
Figure 5.7: SUS score box plot for Zoom and Mozilla Hubs.

Table 5.3: Average SUS scores across both platforms

	Video Conferencing	Virtual Environments
SUS Scores	75	75.67

SUS Scores were also plotted for each of the ten orginal question items. These scores can be seen in Figure 5.7.

ANOVA Analysis: One-way ANOVA results and t-test results for the SUS scores can be seen in Figure 5.8. The p-value that was obtained for this comparison was 0.0046, which is much smaller than our alpha value of 0.05. Therefore, we can reject the null hypothesis. ANOVA results indicate that there is a significant difference between the usability scores provided by the study subjects, with subjects having a marginal preference for Mozilla Hubs



Oneway Anova

Summary of Fit

Rsquare	0.026644
Adj Rsquare	0.023378
Root Mean Square Error	1.030942
Mean of Response	3.196667
Observations (or Sum Wgts)	300

Pooled t Test

Zoom-Mozilla Hubs

Assuming equal variances

Difference	0.340000	t Ratio	2.856113
Std Err Dif	0.119043	DF	298
Upper CL Dif	0.574271	Prob > |t|	0.0046*
Lower CL Dif	0.105729	Prob > t	0.0023*
Confidence	0.95	Prob < t	0.9977

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Platform	1	8.67000	8.67000	8.1574	0.0046*
Error	298	316.72667	1.06284		
C. Total	299	325.39667			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Mozilla Hubs	150	3.02667	0.08418	2.8610	3.1923
Zoom	150	3.36667	0.08418	3.2010	3.5323

Figure 5.8: ANOVA results for the SUS scores.

over Zoom.

Presence subjective evaluation: In addition to the original ten SUS questions users were also asked to respond to the following question: “I felt present when I was interacting with other people in Zoom/Mozilla Hubs”. This was a subjective question that we added to directly address how present and immersed users felt while completing the three tasks across both platforms. The presence score for Zoom was 2.87 with a standard deviation of 1.25, while the average score for Mozilla Hubs was 4.73 with a standard deviation of 0.59. These values are given in Table 5.4

Table 5.4: Average subjective presence scores across both platforms

	Video Conferencing	Virtual Environments
Presence Score Average	2.87	4.73
Presence Score Std. Dev	1.25	0.59

5.1.2 Qualitative Analysis

User Comments: User feedback was collected from the free-response portion of the post-study questionnaire is listed below. Overall, users seemed to find Zoom to be more intuitive and easy to perform tasks in. Mozilla Hubs on the other hand was more interesting to use, with some subjects reporting increased immersion. This corroborates the results obtained through the subjective presence question item. Subject qualitative responses from the post-study questionnaire are listed below.

Zoom:

- “I found the tasks of experience easy to follow. The Zoom system makes it easy to track the person who speaks even if an external distraction occurs at the same”
- “It was good, but doesn’t seem immersive”
- “I have been using it for sometime now. So it just felt as usual to me”
- “I found Zoom to be pretty easy to use and effortless. But the interaction didn’t feel as nice as the VR version”
- “It was good”
- “After 2 yrs of being online I will not sign up to do zoom if necessary I do not mind using it”
- “I feel like i was not that present during the zoom meetings, i was only paying attention to when my turn will come and what to say”

Mozilla Hubs:

- “Initially I did not feel very confident using Mozilla. However, after I was explained how it works and how to operate it, it became easier and really interesting”
- “I liked it, i felt like i was present there”
- “It was my first time with hubs and I liked the concept behind it”
- “I liked using this platform. It felt like I was on Club Penguin server but fancier”
- “It’s a pretty cool way to interact with the immersive interface”
- “With mozilla hubs I feel seeing the speaker talk made me pay attention to them and the presentation also was as similar to real life as possible, which made me go into the professional meeting zone mentally. The alphabetical order thing was hard to do in hubs than zoom”

We collected 7 qualitative responses from the study subjects for Zoom and 6 responses for Mozilla Hubs. Zoom responses can be used to classify subjects into three groups based on their affect towards Zoom:

- Group 1: Those accustomed to and happy with Zoom for remote collaboration. 3 out of the 7 responses fall under this group.
- Group 2: Those who were neutral about using Zoom. 3 out of the 7 responses fall under this group.
- Group 3: Those who disliked using Zoom for remote collaboration. 1 out of the 7 responses fall under this group.

A similar grouping can be created to classify subjects based on their affect towards Mozilla Hubs:

- Group 1: Those who were happy with Mozilla Hubs and can see themselves using it for remote collaboration. 4 out of the 6 responses fall under this group.
- Group 2: Those who were neutral about using Mozilla Hubs. 2 out of the 6 responses fall under this group.
- Group 3: Those who disliked using Mozilla Hubs for remote collaboration. None of the responses fall under this group.

Qualitative feedback from the subjects suggest an overall positive sentiment towards Mozilla Hubs and a more neutral sentiment for Zoom. These results, combined with the SUS score results from Section 5.1.2 and the subjective presence question item, indicate that subjects had a moderate preference for using online virtual environments for remote collaboration over video conferencing.

5.2 Embodied interaction in XR environments

Embodied interaction can enhance a user's sense of presence in an XR environment [15, 36, 52]. In this section we present the evaluation for the system design that we proposed in Section 4.2 to support embodied interactions in XR. We tested network latency for the system across three different configurations and observed a maximum latency of 120ms.

5.2.1 Performance Evaluation

To evaluate the feasibility of our implementation for real-time remote collaboration, we conducted a few validation tests. Network latency was selected our test parameter because

Figure 5.9: Left: The avatars of the two remote user's interacting. Right: The avatar of a remote user as seen from another user's point-of-view.

of its important role in supporting distributed XR collaboration [20, 21, 25, 33]. Two remote users (A and B) met and interacted with each in our shared virtual classroom environment. The user's were made to perform simple actions like moving around and waving their hands (Figure 5.9). The network latency for these interactions was measured for three remote user configurations. Equation 5.1 was used to calculate the travel time of 10000 MQTT messages between the Kinect client and XR applications of user A and user B. The round-trip travel time for the messages was computed and averaged in order to negate the effects of minor system clock time differences between each user's client PC. The results from running these latency tests are summarized in Table 5.5.

$$t_{Travel} = \frac{(t_{rec} - t_{sent})}{2} \tag{5.1}$$

where:

- t_{Travel} is the time taken to send a message from user A's Kinect client application to user B's XR application client over the internet using MQTT (one-way average network delay).
- t_{sent} is the timestamp of when user A's Kinect client application sent out a message to user B's XR application via MQTT.

Table 5.5: Communication latency for different distances between locations.

Communication Type	Distance (miles)	Communication Time(ms)
Same building, local area network	0	56
Short-distance, same town	3	60
Long-distance, different country	4710	120

- t_{rec} is the timestamp of when user A's Kinect client application receives confirmation of having received t_{sent} from user B's XR application.

In the first test configuration, both the users were located in separate rooms within the same building at our research lab facility (Virginia Tech, Blacksburg, VA, USA). As both of these users were on the same network, the latency of the communication component was a minimal 0.056s. As a result, avatar movements were updated very fast for both users.

In the second test configuration, user A was located at our lab facility, while user B joined the virtual classroom environment from the university library, three miles away from the lab. In this test, we noticed a delay of 0.06s, which also allowed for fast avatar position updates. In our final test configuration, user A was located in Blacksburg, VA, USA, while user B joined the virtual classroom from Zagreb, Croatia. In this test, we noticed a larger delay of 0.120s per message. There was also more jitter in the avatar movements observed, but the movements of the avatars still matched the body movements of the users. Also, given the large physical distance between the two users (4710 miles), we believe this to be a tolerable amount of latency for real-time collaborative applications [51].

5.3 Results Summary

In Chapter 3, we defined the research questions that we sought to address through this thesis. We defined these research questions as the following:

- RQ1: Can online virtual environments can serve as a suitable medium for remote collaboration?
- RQ2: Can online virtual environments provide higher levels of presence and engagement for multi-user meetings when compared to contemporary video conferencing systems?
- RQ3: Can 3D virtual environments (VEs) that support embodied interactions be used for remote synchronous meetings?

In this section we cover the results from implementing the approaches described in chapter 4. Section 4.1 describes the approach taken to address RQ1 and RQ2, while Section 4.2 covers the approach taken to address RQ3. The results from implementing both of these methods were described in Section 5.1 and Section 5.2. A summary of these results are described in this section.

With RQ1, we aim to establish that web-based 3D VEs can serve as a suitable medium for remote collaboration. Web-based VEs are uniquely situated between currently used video conferencing tools and fully immersive XR. However, their enhanced accessibility make them ideal platforms to provide users with immersive remote collaboration experiences. RQ2 builds on RQ1, but hypothesizing that in addition to being suitable platforms for remote collaboration, VEs can provide users with a richer collaboration experience. We conjectured that VEs would accomplish this by providing users with greater levels of presence and immersion with the collaborative medium. Finally RQ3 explores the feasibility of incorporating real-time body tracking in XR environments. We proposed a system design in Section 4.2

and implemented and tested it in 5.2. Results indicated a maximum latency of 120ms which supports the use of the proposed system in real-time collaborative XR applications.

5.3.1 Research Question 1:

RQ1 examined the usability of web-based virtual environments (Mozilla Hubs). After analyzing the results from the user study 5.1, we found that Web-based virtual environments perform comparably with videoconferencing platforms. SUS results even suggest that subjects had a slight preference for Mozilla Hubs over Zoom (Mozilla Hubs SUS Score = 75.66, Zoom SUS Score = 75). User comments were also favorable an indicated positive inclination towards using web-based VEs like Mozilla Hubs.

5.3.2 Research Question 2:

The quantitative results corresponding to RQ2 were less conclusive. ANOVA analysis over the eye-tracking data gave a p-value of 0.28, which led us to fail to reject the null hypothesis (there is a significant between subject eye gaze between the two platforms). ANOVA analysis on event recognition data was also inconclusive with a p-value of 0.12. Therefore, we failed to reject the null hypothesis for this case as well. However, the System Usability Scale, user comments and the subjective presence evaluation question demonstrated positive results. Based on the SUS Scores, subjects were seen to have a marginal preference for Mozilla Hubs over Zoom. The subjective presence question asked users about how immersed in the environment they felt while performing the study tasks. Results from this question indicated a strong preference for Mozilla Hubs over Zoom (Average Mozilla Hubs Score = 4.73, Std. Dev = 0.59; Average Zoom Score = 2.87, Std. Dev = 1.25). However, since the eye tracking and event recognition time results were inconclusive, further work needs to be done to quantifiably evaluate presence and engagement differences between the two platforms.

5.3.3 Research Question 3:

To evaluate RQ3, we proposed, implementation and tested the system design from Section 5.2. Results from testing latency between the remote collaborators XR applications suggested positive results. Testing was done by sending 10000 MQTT messages (representing collaborator avatar movements) between the XR applications of two remote collaborators. When the collaborators were connected over the same network, latency was 56ms. When they were 3 miles away from each other, latency was 60ms. A final test with the collaborators separated by a distance of 4710 miles revealed a latency of 120ms. This maximum latency values supports our claim for RQ3 and established that XR platforms are capable of supporting embodied interactions for remote collaboration.

5.4 Discussion

5.4.1 User study results and the limitations of the existing state-of-the-art

We compared user attention and task performance across Zoom (video conferencing) and Mozilla Hubs (Virtual Environment). We conducted a user study with fifteen subjects who were very familiar with Zoom and unfamiliar with Mozilla Hubs. The goal of this study was to determine the current feasibility of web-based virtual environments for remote collaboration. With the recent increased attention in Facebook's Metaverse, immersive technology is currently in the spotlight. XR is a promising space that seeks to permeate most industry domains. However, immersive technology is very much hardware dependent. The relatively high price point of the hardware makes the commercial use of XR difficult. Web-based virtual environments seek to address these accessibility challenges, by giving users multiple options to attend meetings in immersive spaces. People with access to XR hardware can join meetings using their personal XR devices, while those without the hardware can join immersive meetings through their web browser. This study is quite useful in this context as it explores user engagement differences between the current state-of-the-art in remote collaboration (video conferencing) with web-based virtual environments.

We conducted quantitative and qualitative analysis of our user study data. The results from the study can be found in Section 5.1. For quantitative results, we compared user eye gaze duration and event recognition times. Eye gaze duration represents the amount of time (as a percentage of total task time) spent by subjects looking at items of interest. Event recognition time represents the amount of time (in seconds) that it took users to recognize important task events (the current speaker in a conversation, collaborators wanting to ask questions, knowing when it is their turn to speak). The graphs and ANOVA analysis for both of these metrics can be found in Section 5.1. The results that we obtained depicted comparable results for Zoom and Mozilla Hubs. The ANOVA analysis concluded that there was no significant difference in these two metrics between the two platforms. However, the fact that the results were comparable is reassuring, as it demonstrates that while web-based platforms aren't undeniably better than video conferencing tools, they perform at least on par with them.

Qualitative results showed moderate subject preference for the Web-based virtual environment over video conferencing. Subjects gave Zoom an average SUS score of 75, while Mozilla Hubs received a slightly higher score of 75.67. Subjects gave Mozilla Hubs a significantly higher average score of 4.73 (std dev: 0.89) in the subjective presence evaluation, while Zoom was given an average score of 2.87 (std dev: 1.29). Finally, subjects commented that while Zoom was more familiar and intuitive, Mozilla Hubs was more fun and felt more immersive. These qualitative results suggest that users are likely to enjoy performing remote collaborative tasks in web-based virtual environments.

Web-based virtual environments are more affordable, easy to access and are less tiring than fully-immersive virtual environments. They are more stimulating than video conferencing tools and can support avatar gestures, facial expressions and diverse background environments. All of these features make online virtual environments a great intermediary platform for remote collaboration, until fully immersive XR becomes more mainstream. The findings of this thesis are positive as they suggest that online virtual environments perform comparably with video conferencing tools for remote collaboration. However, online virtual environments are still very new and are being used by a small set of users. As these environments evolve to support additional functionality, their utility for remote collaboration will continue to increase.

5.4.2 Embodied cognition and presence in remote collaboration

Embodied cognition in shared XR environments can be enhanced by ensuring that the latency between the movement of a user's body and their virtual avatar is minimal. In our evaluation, we measured the communication latency in updating avatars for two remote users in three scenarios. We observed that the communication time between updating user avatar positions corresponds to the physical distance between the users. We also noticed some jitter in the avatar movements, which was more pronounced in the case of large physical distance among users. This jitter could be attributed to packets of data containing user skeleton joints, not reaching user XR applications fast enough. This can be further optimized in future iterations of this implementation to improve the efficiency of the embodied experience.

Embodied cognition and interaction in remote collaboration can be of special relevance to online classrooms. Many students are unable to pay as much attention in online classes as they could in physical classrooms. For students suffering from Attention deficit hyperactivity disorder (ADHD) focusing during online lectures is especially challenging. Research shows that ADHD is not simply a set of mental functions, but rather a range of bodily dynamics through which humans engage with their environment [56].The embodied cognition of remote collaboration has the potential to significantly help students with attention disorders pay more in online classrooms.

Virtual-environment-based classroom sessions can also slowly be incorporated into existing Zoom-based online classrooms. During these sessions, students and the instructor can log into a shared virtual environment. The classroom can serve as a homeroom for the instructor to provide lesson overviews, introductions and set expectations for what students will encounter in that day's virtual environment. Virtual classroom sessions can range from simple sessions, where the instructor presents virtual models to the class, to transportation into completely virtual landscapes [84].

For example, the doors of the classroom could be made to be portals that teleport the students to Africa, where they can interact with local flora and fauna and gaze upon Mount Kilimanjaro. The students can use the doors/portals to move between classrooms for different subjects. Students can be allowed to dynamically interact with the objects in a room to change their state. Virtual objects can also be implemented to have textual and audio properties that play when interacted with [11]. The interactive nature of virtual environments can captivate and hold student attention for much longer than in a typical lecture session. For this reason virtual environments can be a great supplement to existing online education. However, a primary challenge to incorporating embodied interactions in shared virtual environments is making these environments accessible for all users. While XR headsets are slowly becoming more affordable, they are still a few years away from being used as personal devices for work (like laptops). Body-tracking sensors are also typically used for niche areas like for gaming, animation or research. The lack of commercial access to specific hardware can be a barrier to the widespread use of embodied interaction and XR.

A potential solution to this problem is democratize access by also making the shared virtual environments available as web-enabled experiences. Users would be able to open up a browser window and log into the virtual environment, just like they would to attend a Zoom meeting. Users would interact with the same virtual content through their browser as they would through an XR headset. The primary difference between the two being that XR users would experience the environment from a first-person perspective, while the online users would experience it in third-person. Both types of users can create custom avatars to represent themselves in the shared virtual environment. XR users could navigate and interact with the environment using their controller, while online users would use their keyboard and mouse to move their avatars around in a virtual environment. This kind of implementation would be accessible to a much wider range of users and can open up exciting new avenues for platform-agnostic collaboration in shared virtual environments.

Chapter 6

Conclusion

Extended Reality presents some interesting possibilities and opens up new modes of collaboration among individuals. It has the potential to make distances smaller and enhance user interaction with smart spaces. Remote-collaboration between engineers and on-site workers can reduce cognitive load for on-site workers. Interactive tutorials can provide students with novel ways of learning material.
The objectives of this thesis were: 1) To explore presence and engagement differences between video conferencing and virtual environments and 2) To evaluate the feasibility of embodied interactions in collaborative XR. We observed no significant differences in attention and event recognition time between the two platforms (based on the eye-tracking data). However, our qualitative analysis revealed slight subject preference of Mozilla Hubs over Zoom. These results are still quite positive overall, considering that subjects were very familiar with Zoom and inexperienced with Mozilla Hubs prior to the study. Virtual Environments are still in their infancy and a lot of their features are still in the prototype stage. However, given their current trajectory and renewed excitement about the Metaverse, online virtual environments should see significant improvements to usability within the next five to ten years.

The outbreak of COVID-19 also forced people to switch to video conferencing platforms for remote work. The feeling of immersion, engagement, and presence in such video conferencing platforms is low. This can result in reduced attention and increased fatigue among remote collaborators. To address some of the challenges, we proposed and implemented an approach to conducting virtual meetings, while leveraging embodied interaction. These embodied interaction techniques can help facilitate better remote collaboration in shared virtual environments. Our approach comprises of a body tracking component, a communication component, and an XR rendering component. We developed a virtual classroom scenario as our prototype implementation and connected two locations, over 4000 miles apart. The maximum latency that we observed during our usability evaluation was 120ms, suggesting the potential for this kind of approach in facilitating real-time remote collaboration.

In summary, XR technology has come a long way in the past two decades and is currently poised to radically change the way we interact with digital content and each other. Further developments in XR will open up new exciting avenues for data visualization, training and remote collaboration.

Future Work

Video conferencing tools like Zoom and Skype are currently the dominant platforms for remote collaboration. However, Virtual Environments like Mozilla Hubs and VirBELA are also seeing increased interest, especially as platforms to hold large group events. These platforms are more immersive than video conferencing tools, however, they still fall short of fully-immersive VR. The next iteration of this study will also compare subject task performance in fully-immersive VR with web-based virtual environments and video conferencing tools. This comparison between the three platforms would provide a richer understanding of the commercial feasibility of immersive technology.

Additionally, our future work will involve porting and testing the usability of the system with standalone XR headsets. We also intend to enhance our system design by exploring more efficient data management and communication protocols that can reduce jitter in avatar movements and allow several remote users to join the virtual environment. These additional components will be implemented and validated by a user study in our future work.

Bibliography

[1] Md Ashraful Alam and Khalid Al Riyami. Shear strengthening of reinforced concrete beam using natural fibre reinforced polymer laminates. *Construction and Building Materials*, 162:683–696, 2018.

[2] M. Bauer, G. Kortuem, and Z. Segall. ”where are you pointing at?” a study of remote collaboration in a wearable videoconference system. In *Digest of Papers. Third International Symposium on Wearable Computers*, pages 151–158, Oct 1999. doi: 10.1109/ISWC.1999.806696.

[3] M. Billinghurst, J. Bowskill, M. Jessop, and J. Morphett. A wearable spatial conferencing space. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, pages 76–83, Oct 1998. doi: 10.1109/ISWC.1998.729532.

[4] M. Billinghurst, H. Kato, and I. Poupyrev. The magicbook - moving seamlessly between reality and virtuality. *IEEE Computer Graphics and Applications*, 21(3):6–8, May 2001. ISSN 1558-1756. doi: 10.1109/38.920621.

[5] Mark Billinghurst and Hirokazu Kato. Collaborative augmented reality. *Commun. ACM*, 45(7):64–70, July 2002. ISSN 0001-0782. doi: 10.1145/514236.514265. URL http://doi.acm.org/10.1145/514236.514265.

[6] D. A. Bowman, S. Coquillart, B. Froehlich, M. Hirose, Y. Kitamura, K. Kiyokawa, and W. Stuerzlinger. 3d user interfaces: New directions and perspectives. *IEEE Computer Graphics and Applications*, 28(6):20–36, Nov 2008. ISSN 1558-1756. doi: 10.1109/MCG.2008.109.

[7] Neil A Bradbury. Attention span during lectures: 8 seconds, 10 minutes, or more? *Advances in physiology education*, 40(4):509–513, 2016.

[8] John Brooke et al. SUS—a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[9] Barry Brown, Ian MacColl, Matthew Chalmers, Areti Galani, Cliff Randell, and Anthony Steed. Lessons from the lighthouse: Collaboration in a shared mixed reality system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 577–584, New York, NY, USA, 2003. ACM. ISBN 1-58113-630-7. doi: 10.1145/642611.642711. URL http://doi.acm.org/10.1145/642611.642711.

[10] Ozan Cakmakci and Jannick Rolland. Head-worn displays: a review. *Journal of display technology*, 2(3):199–216, 2006.

[11] Andrea Casu, Lucio Davide Spano, Fabio Sorrentino, and Riccardo Scateni. Riftart: Bringing masterpieces in the classroom through immersive virtual reality. In *Proceedings of the Smart Tools and Apps for Graphics — Eurographics Italian Chapter Conference*, pages 77–84, 2015.

[12] Thomas Caudell and David Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. volume 2, pages 659 – 669 vol.2, 02 1992. ISBN 0-8186-2420-5. doi: 10.1109/HICSS.1992.183317.

[13] Sicheng Chen, Miao Chen, Andreas Kunz, Evren Yantac, Mathias Bergmark, Anders Sundin, and Morten Fjeld. Semarbeta: Mobile sketch-gesture-video remote support for car drivers. 03 2013. doi: 10.1145/2459236.2459249.

[14] Kun-Hung Cheng and Chin-Chung Tsai. Affordances of augmented reality in science learning: Suggestions for future research. *Journal of Science Education and Technology*,

22(4):449–462, Aug 2013. ISSN 1573-1839. doi: 10.1007/s10956-012-9405-9. URL https://doi.org/10.1007/s10956-012-9405-9.

[15] Adrian David Cheok, Wang Weihua, Xubo Yang, Simon Prince, Fong Siew Wan, Mark Billinghurst, and Hirokazu Kato. Interactive theatre experience in embodied + wearable mixed reality space. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, ISMAR '02, pages 59–, Washington, DC, 2002. IEEE Computer Society. ISBN 0-7695-1781-1.

[16] Meia Chita-Tegmark. Social attention in asd: A review and meta-analysis of eye-tracking studies. *Research in Developmental Disabilities*, 48:79–93, 2016. ISSN 0891-4222. doi: https://doi.org/10.1016/j.ridd.2015.10.011. URL https://www.sciencedirect.com/science/article/pii/S0891422215001821.

[17] Archi Dasgupta, Nicole Buckingham, Denis Gračanin, Mohamed Handosa, and Reza Tasooji. A mixed reality based social interactions testbed: A game theory approach. In Jessie Y.C. Chen and Gino Fragomeni, editors, *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry*, pages 40–56, Cham, 2018. Springer International Publishing. ISBN 978-3-319-91584-5.

[18] Archi Dasgupta, Mark Manuel, Rifat S. Mansur, Nabil Nowak, and Denis Gračanin. Towards real time object recognition for context awareness in mixed reality: A machine learning approach. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 262–268, 2020.

[19] Archi Dasgupta, Samuel Williams, Gunnar Nelson, Mark Manuel, Shaoli Dasgupta, and Denis Gračanin. Redefining the digital paradigm for virtual museums. In *International Conference on Human-Computer Interaction*, pages 357–373. Springer, 2021.

[20] Declan Delaney, Tomás Ward, and Seamus McLoone. On consistency and network latency in distributed interactive applications: A survey—part I. *Presence: Teleoperators and Virtual Environments*, 15(2):218–23, April 2006.

[21] Declan Delaney, Tomás Ward, and Seamus McLoone. On consistency and network latency in distributed interactive applications: A survey—part II. *Presence: Teleoperators and Virtual Environments*, 15(4):465–482, August 2006.

[22] Gwyneth Doherty-Sneddon and Fiona G Phelps. Gaze aversion: A response to cognitive or social difficulty? *Memory & cognition*, 33(4):727–733, 2005.

[23] Paul Dourish and Sara Bly. Portholes: Supporting awareness in a distributed work group. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '92, pages 541–547, New York, NY, USA, 1992. ACM. ISBN 0-89791-513-5. doi: 10.1145/142750.142982. URL http://doi.acm.org/10.1145/142750.142982.

[24] Romina Druta, Cristian Druta, Paul Negirla, and Ioan Silea. A review on methods and systems for remote collaboration. *Applied Sciences*, 11(21), 2021. ISSN 2076-3417. doi: 10.3390/app112110035. URL https://www.mdpi.com/2076-3417/11/21/10035.

[25] Carmine Elvezio, Frank Ling, Jen-Shuo Liu, and Steven Feiner. Collaborative virtual reality for low-latency interaction. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings*, UIST '18 Adjunct, page 179–181, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450359498. doi: 10.1145/3266037.3271643. URL https://doi.org/10.1145/3266037.3271643.

[26] FE Fadzli, MS Kamson, AW Ismail, and MYF Aladin. 3D telepresence for remote collaboration in extended reality (xR) application. *IOP Conference Series: Materials Science and Engineering*, 979:012005, November 2020.

[27] Helen Farley and Caroline Steel. A quest for the holy grail: Tactile precision, natural movement and haptic feedback in 3d virtual spaces. In *Proceedings ASCILITE 2009: 26th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Same Places, Different Spaces*, pages 285–295. Australasian Society for Computers in Learning in Tertiary Education (ASCILITE), 2009.

[28] Morten Fjeld, Kristina Lauche, Martin Bichsel, Fred Voorhorst, Helmut Krueger, and Matthias Rauterberg. Physical and virtual tools: Activity theory applied to the design of groupware. *Computer Supported Cooperative Work (CSCW)*, 11(1-2):153–180, 2002.

[29] Neil Fleming and David Baume. Learning styles again: VARKing up the right tree! *Educational developments*, 7(4):4–7, 2006.

[30] Michael Galrnyk. Understanding boxplots. https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51, 2018. [Last accessed 21-Dec-2021].

[31] S. Gauglitz, B. Nuernberger, M. Turk, and T. Höllerer. World-stabilized annotations and virtual scene navigation for remote collaboration. pages 449–460, 2014. doi: 10.1145/2642918.2647372. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-84912054997&doi=10.1145%2f2642918.2647372&partnerID=40&md5=20e3a7bd6db7dbe2346ef088833268e2.

[32] Jane Goldman. Kingsman: The secret service. https://www.foxmovies.com/movies/kingsman-the-golden-circle, 2014. [Last accessed 14-May-2019].

[33] Denis Gračanin, Yunxian Zhou, and Luiz DaSilva. Quality of service for networked virtual environments. *IEEE Communications Magazine*, 42(4):42–48, April 2004.

[34] J. Grudin. Computer-supported cooperative work: history and focus. *Computer*, 27(5):19–26, May 1994. doi: 10.1109/2.291294.

[35] Felix G. Hamza-Lup and Jannick P. Rolland. Scene synchronization for real-time interaction in distributed mixed reality and virtual reality environments. *Presence: Teleoperators & Virtual Environments*, 13(3):315–327, June 2004.

[36] Mohamed Handosa, Hendrik Schulze, Denis Gračanin, Matthew Tucker, and Mark Manuel. Extending embodied interactions in mixed reality environments. In Jessie Y.C. Chen and Gino Fragomeni, editors, *Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation*, pages 314–327, Cham, 2018. Springer International Publishing. ISBN 978-3-319-91581-4.

[37] Mohamed Handosa, Archi Dasgupta, Mark Manuel, and Denis Gračanin. Rethinking user interaction with smart environments—a comparative study of four interaction modalities. In *International Conference on Human-Computer Interaction*, pages 39–57. Springer, 2020.

[38] Iris Herbst, Anne-Kathrine Braun, Roderick McCall, and Wolfgang Broll. Timewarp: interactive time travel with a mobile mixed reality game. *In the Proceedings of MobileHCI 2008, Amsterdam, Netherlands*, 2008.

[39] Roy S Hessels, Gijs A Holleman, Tim HW Cornelissen, Ignace TC Hooge, and Chantal Kemner. Eye contact takes two–autistic and social anxiety traits predict gaze behavior in dyadic interaction. *Journal of Experimental Psychopathology*, 9(2):jep–062917, 2018.

[40] Tobias Höllerer, Steven Feiner, Drexel Hallaway, Blaine Bell, Marco Lanzagorta, Dennis Brown, Simon Julier, Yohan Baillot, and Lawrence Rosenblum. User interface management techniques for collaborative mobile augmented reality. *Computers And Graphics*, 25(5):799–810, 2001.

[41] C. E. Hughes, C. B. Stapleton, D. E. Hughes, and E. M. Smith. Mixed reality in

education, entertainment, and training. *IEEE Computer Graphics and Applications*, 25(6):24–30, Nov 2005. ISSN 0272-1716. doi: 10.1109/MCG.2005.139.

[42] Urs Hunkeler, Hong Linh Truong, and Andy Stanford-Clark. MQTT-S—A publish/-subscribe protocol for wireless sensor networks. In *Proceedings of the 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE'08)*, pages 791–798. IEEE, 2008.

[43] Hans-Christian Jetter, Harald Reiterer, and Florian Geyer. Blended interaction: Understanding natural human—computer interaction in post-wimp interactive spaces. *Personal Ubiquitous Comput.*, 18(5):1139–1158, June 2014. ISSN 1617-4909. doi: 10.1007/s00779-013-0725-4. URL http://dx.doi.org/10.1007/s00779-013-0725-4.

[44] R. Johansen. Teams for tomorrow (groupware). In *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences*, volume iii, pages 521–534 vol.3, Jan 1991. doi: 10.1109/HICSS.1991.184183.

[45] Vlasios Kasapakis, Damianos Gavalas, and Dzardanova Elena. Robust outdoors marker-based augmented reality applications: Mitigating the effect of lighting sensitivity. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 423–431. Springer, 2018.

[46] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, pages 85–94, Oct 1999. doi: 10.1109/IWAR.1999.803809.

[47] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana. Virtual object manipulation on a table-top ar environment. In *Proceedings IEEE and ACM Interna-*

tional Symposium on Augmented Reality (ISAR 2000), pages 111–119, Oct 2000. doi: 10.1109/ISAR.2000.880934.

[48] Adam Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26:22–63, 1967.

[49] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pages 3–12, Oct 2010. doi: 10.1109/ISMAR.2010.5643530.

[50] Philipp Ladwig and Christian Geiger. A literature review on collaboration in mixed reality. In Michael E. Auer and Reinhard Langmann, editors, *Smart Industry And Smart Education*, pages 591–600, Cham, 2019. Springer International Publishing. ISBN 978-3-319-95678-7.

[51] Valerie D Lehner and Thomas A DeFanti. Distributed virtual reality: Supporting remote collaboration in vehicle design. *IEEE Computer Graphics and Applications*, 17 (2):13–17, 1997.

[52] Robb Lindgren and J. Michael Moshell. Supporting children's learning with body-based metaphors in a mixed reality environment. In *Proceedings of the 10th International Conference on Interaction Design and Children*, IDC '11, pages 177–180, New York, 2011. ACM. ISBN 978-1-4503-0751-2.

[53] Matthew Lombard and Theresa Ditton. At the heart of it all: The concept of presence. *Journal of computer-mediated communication*, 3(2):JCMC321, 1997.

[54] MagicLeap. Magic leap | the next era of computing. https://www.magicleap.com/. [Last accessed 14-May-2019].

[55] N. Magnenat-Thalmann and U. Bonanni. Haptics in virtual reality and multimedia. *IEEE MultiMedia*, 13(3):6–11, 2006. doi: 10.1109/MMUL.2006.56.

[56] Michelle Maiese. Rethinking attention deficit hyperactivity disorder. *Philosophical Psychology*, 25(6):893–916, 2012.

[57] Hassan Mansour and Gustav Kuhn. Studying "natural" eye movements in an "unnatural" social environment: The influence of social activity, framing, and sub-clinical traits on gaze aversion. *Quarterly Journal of Experimental Psychology*, 72(8):1913–1925, 2019.

[58] Mark Manuel. Video: Virtual classroom. https://youtu.be/XGzxcCI3RbA. [Last accessed 12 February 2021].

[59] Mark Manuel, Poorvesh Dongre, Abdulaziz Alhamadani, and Denis Gračanin. Supporting embodied and remote collaboration in shared virtual environments. In *International Conference on Human-Computer Interaction*, pages 639–652. Springer, 2021.

[60] Joshua McVeigh-Schultz and Katherine Isbister. A "beyond being there" for vr meetings: envisioning the future of remote work. *Human–Computer Interaction*, 0(0): 1–21, 2021. doi: 10.1080/07370024.2021.1994860. URL https://doi.org/10.1080/07370024.2021.1994860.

[61] Brian E Mennecke, Janea L Triplett, Lesya M Hassall, Zayira Jordán Conde, and Rex Heer. An examination of a theory of embodied social presence in virtual worlds. *Decision Sciences*, 42(2):413–450, 2011.

[62] Microsoft. Box plots with outliers. https://www.real-statistics.com/descriptive-statistics/box-plots-with-outliers/, 2016. [Last accessed 21-Dec-2021].

[63] Microsoft. Microsoft hololens (1st gen). https://docs.microsoft.com/en-us/windows/mixed-reality/hololens-hardware-details, 2018. [Last accessed 14-May-2019].

[64] Microsoft. Azure kinect body tracking joints, 2020. URL https://docs.microsoft.com/en-us/azure/kinect-dk/body-joints. [Last accessed 12 February 2021].

[65] Paul Milgram and Fumio Kishino. A taxonomy of mixed reality visual displays. 1994.

[66] Shinya Minatani, Itaru Kitahara, Yoshinari Kameda, and Yuichi Ohta. Face-to-face tabletop remote collaboration in mixed reality. In *Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, ISMAR '07, pages 1–4, Washington, DC, USA, 2007. IEEE Computer Society. ISBN 978-1-4244-1749-0. doi: 10.1109/ISMAR.2007.4538823. URL https://doi.org/10.1109/ISMAR.2007.4538823.

[67] Marvin Minsky. Telepresence, 1980.

[68] Betsy Morris. Why does zoom exhaust you? science has an answer. *The Wall Street Journal*, May 2020. [Last accessed 6 June 2021].

[69] Jens Müller, Roman Rädle, and Harald Reiterer. Remote collaboration with mixed reality displays: How shared virtual landmarks facilitate spatial referencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 6481–6486, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4655-9. doi: 10.1145/3025453.3025717. URL http://doi.acm.org/10.1145/3025453.3025717.

[70] Mohammad Nazim and Bhaskar Mukherjee. *Knowledge management in libraries: Concepts, tools and approaches.* Chandos publishing, 2016.

[71] S. Ohl. Tele-immersion concepts. *IEEE Transactions on Visualization and Computer Graphics*, 24(10):2827–2842, Oct 2018. ISSN 1077-2626. doi: 10.1109/TVCG.2017.2767590.

[72] Oyewole Oyekoya, Ran Stone, William Steptoe, Laith Alkurdi, Stefan Klare, Angelika Peer, Tim Weyrich, Benjamin Cohen, Franco Tecchia, and Anthony Steed. Supporting interoperability and presence awareness in collaborative mixed reality environments. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, VRST '13, pages 165–174, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2379-6. doi: 10.1145/2503713.2503732. URL http://doi.acm.org/10.1145/2503713.2503732.

[73] Chris Panou, Lemonia Ragia, Despoina Dimelli, and Katerina Mania. Outdoors mobile augmented reality application visualizing 3d reconstructed historical monuments. In *GISTAM*, pages 59–67, 2018.

[74] E. Peters, B. Heijligers, J. de Kievith, X. Razafindrakoto, R. van Oosterhout, C. Santos, I. Mayer, and M. Louwerse. Design for collaboration in mixed reality: Technical challenges and solutions. In *2016 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES)*, pages 1–7, Sep. 2016. doi: 10.1109/VS-GAMES.2016.7590343.

[75] Ulrich J Pfeiffer, Kai Vogeley, and Leonhard Schilbach. From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience & Biobehavioral Reviews*, 37(10):2516–2528, 2013.

[76] Ronald Poelman, Oytun Akman, Stephan Lukosch, and Pieter Jonker. As if being there: mediated reality for crime scene investigation. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 1267–1276. ACM, 2012.

[77] Tobii Pro. What does eye tracking data tell us? https://www.tobiipro.com/learn-and-support/learn/steps-in-an-eye-tracking-study/interpret/what-does-eye-tracking-data-tell-us/, 2018. [Last accessed 15-Aug-2020].

[78] Meta Quest. Oculus quest (2nd gen). https://www.oculus.com/quest-2/, 2021. [Last accessed 20-Dec-2021].

[79] G. Reitmayr and D. Schmalstieg. Mobile collaborative augmented reality. In *Proceedings IEEE and ACM International Symposium on Augmented Reality*, pages 114–123, Oct 2001. doi: 10.1109/ISAR.2001.970521.

[80] Patrick Renner and Thies Pfeiffer. Evaluation of attention guiding techniques for augmented reality-based assistance in picking and assembly tasks. In *Proceedings of the 22Nd International Conference on Intelligent User Interfaces Companion*, IUI '17 Companion, pages 89–92, New York, 2017. ACM. ISBN 978-1-4503-4893-5.

[81] Giuseppe Riva, Brenda K Wiederhold, and Fabrizia Mantovani. Surviving covid-19: The neuroscience of smart working and distance learning. *Cyberpsychology, Behavior, and Social Networking*, 24(2):79–85, 2021.

[82] David W Schloerb. A quantitative measure of telepresence. *Presence: Teleoperators And Virtual Environments*, 4(1):64–80, 1995.

[83] Ulrike Schultze. Embodiment and presence in virtual worlds: a review. *Journal of Information Technology*, 25(4):434–449, 2010.

[84] Sharad Sharma, Ruth Agada, and Jeff Ruffin. Virtual reality classroom as an constructivist approach. In *2013 Proceedings of IEEE Southeastcon*, pages 1–5. IEEE, 2013.

[85] Sandeep Singhal and Michael Zyda. *Networked Virtual Environments: Design and Implementation*. ACM Press SIGGRAPH Series. Addison Wesley, Reading, Massachusetts, 1999.

[86] David Sirkin, Gina Venolia, John Tang, George Robertson, Taemie Kim, Kori Inkpen, Mara Sedlins, Bongshin Lee, and Mike Sinclair. Motion and attention in a kinetic videoconferencing proxy. In *IFIP Conference on Human-Computer Interaction*, pages 162–180. Springer, 2011.

[87] Mel Slater and Anthony Steed. A virtual presence counter. *Presence: Teleoperators And Virtual Environments*, 9(5):413–434, 2000.

[88] Robert Speidel, Achim Schneider, Jasmin Körner, Claudia Grab-Kroll, and Wolfgang Öchsner. Did video kill the xr star? digital trends in medical education before and after the covid-19 outbreak from the perspective of students and lecturers from the faculty of medicine at the university of ulm. *GMS Journal for Medical Education*, 38 (6), 2021.

[89] Anmol Srivastava. Enriching student learning experience using augmented reality and smart learning objects. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI 2016, pages 572–576, New York, 2016. ACM. ISBN 978-1-4503-4556-9.

[90] Adam Stefanile. The transition from classroom to zoom and how it has changed education.

[91] Frank Steinicke, Nale Lehmann-Willenbrock, and Annika Luisa Meinecke. A first pilot study to compare virtual group meetings using video conferences and (immersive) virtual reality. In *Symposium on Spatial User Interaction*, pages 1–2, 2020.

[92] Norbert A Streitz, Jörg Geißler, and Torsten Holmer. Roomware for cooperative buildings: Integrated design of architectural spaces and information spaces. In *Proceedings of the International Workshop on Cooperative Buildings*, pages 4–21. Springer, 1998.

[93] Norbert A. Streitz, Jörg Geißler, Torsten Holmer, Shin'ichi Konomi, Christian Müller-Tomfelde, Wolfgang Reischl, Petra Rexroth, Peter Seitz, and Ralf Steinmetz. i-LAND: An interactive landscape for creativity and innovation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 120–127, New York, 1999. ACM. ISBN 0-201-48559-1.

[94] Ivan E. Sutherland. 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), pages 757–764, New York, NY, USA, 1968. ACM. doi: 10.1145/1476589.1476686. URL http://doi.acm.org/10.1145/1476589.1476686.

[95] Z. Szalavári, D. Schmalstieg, A. Fuhrmann, and M. Gervautz. "studierstube": An environment for collaboration in augmented reality. *Virtual Reality*, 3(1):37–48, Mar 1998. ISSN 1434-9957. doi: 10.1007/BF01409796. URL https://doi.org/10.1007/BF01409796.

[96] John C. Tang and Scott L. Minneman. Videodraw: A video interface for collaborative drawing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 313–320, New York, NY, USA, 1990. ACM. ISBN 0-201-50932-6. doi: 10.1145/97243.97302. URL http://doi.acm.org/10.1145/97243.97302.

[97] Chryssa Themelis and Julie-Ann Sime. From video-conferencing to holoportation and haptics: How emerging technologies can enhance presence in online education? In *Emerging technologies and pedagogies in the curriculum*, pages 261–276. Springer, 2020.

[98] Meng-Jung Tsai, Huei-Tse Hou, Meng-Lung Lai, Wan-Yi Liu, and Fang-Ying Yang. Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers And Education*, 58(1):375–385, 2012. ISSN 0360-1315. doi: https://doi.org/10.1016/j.compedu.2011.07.012. URL https://www.sciencedirect.com/science/article/pii/S0360131511001709.

[99] Penn State University. Elementary statistics. https://online.stat.psu.edu/stat200/lesson/3/3.2. [Last accessed 21-Dec-2021].

[100] Filemon A Uriarte. *Introduction to knowledge management: A brief introduction to the basic elements of knowledge management for non-practitioners interested in understanding the subject*. Asean Foundation, 2008.

[101] Eline van der Kruk and Marco M. Reijne. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *European Journal of Sport Science*, 18 (6):806–819, 2018.

[102] DWF Van Krevelen and Ronald Poelman. A survey of augmented reality technologies, applications and limitations. *International journal of virtual reality*, 9(2):1–20, 2010.

[103] Shannon Vandenberg and Morgan Magnuson. A comparison of student and faculty attitudes on the use of zoom, a video conferencing platform: A mixed-methods study. *Nurse Education in Practice*, 54:103138, 2021.

[104] Hana Vrzakova, Mary Jean Amon, McKenzie Rees, Myrthe Faber, and Sidney D'Mello. Looking for a deal? visual social attention during negotiations via mixed media videoconferencing. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3): 1–35, 2021.

[105] Mark Weiser. Ubiquitous computing. *Computer*, 26(10):71–72, 1993.

[106] Pierre Wellner. Interacting with paper on the digitaldesk. *Commun. ACM*, 36(7): 87–96, July 1993. ISSN 0001-0782. doi: 10.1145/159544.159630. URL http://doi.acm.org/10.1145/159544.159630.

[107] Brenda K Wiederhold. Beyond zoom: the new reality, 2020.

[108] Brenda K Wiederhold. Connecting through technology during the coronavirus disease 2019 pandemic: Avoiding "Zoom fatigue". *Cyberpsychology, Behavior, and Social Networking*, 23(7):437–438, July 2020.

[109] Jason Wither, Stephen DiVerdi, and Tobias Höllerer. Annotation in outdoor augmented reality. *Computers And Graphics*, 33(6):679–689, 2009.

[110] Bob G Witmer and Michael J Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[111] Matthias M. Wloka and Brian G. Anderson. Resolving occlusion in augmented reality. In *Proceedings of the 1995 Symposium on Interactive 3D Graphics*, I3D '95, pages 5–12, New York, NY, USA, 1995. ACM. ISBN 0-89791-736-7. doi: 10.1145/199404.199405. URL http://doi.acm.org/10.1145/199404.199405.

[112] Carolyn Reinach Wolf. Virtual platforms are helpful tools but can add to our stress, 2020. URL https://www.psychologytoday.com/ca/blog/the-desk-the-mental-health-lawyer/202005/virtual-platforms-are-helpful-tools-can-add-our-stress. [Last accessed 12 February 2021].

[113] Jianghao Xiong, En-Lin Hsiang, Ziqian He, Tao Zhan, and Shin-Tson Wu. Augmented reality and virtual reality displays: emerging technologies and future perspectives. *Light: Science & Applications*, 10(1):1–30, 2021.

Appendices

Appendix A

User Study Documents

A.1 Pre-Study Questionnaire

This section contains all of the questions that were included in the pre-study questionnaire.

1. Subject Code:
2. How old are you?
3. What is your gender?
    - Male
    - Female
    - Other (please specify)
    - Do not wish to provide
4. What is your native language?
    - English
    - Other (please specify)
    - Do not wish to provide
5. What is the highest level of school you have completed or the highest degree you have received?

- Less than high school degree
- High school degree or equivalent
- Some college but no degree
- Associate degree
- Bachelor's degree
- Graduate degree
- Do not wish to provide

6. If applicable, what is your major of study?

7. What is your ethnicity?

   - White
   - Black or African-American
   - American Indian or Alaska Native
   - Asian
   - Native Hawaiian or Pacific Islander
   - From multiple races
   - Some other race (please specify)
   - Do not wish to provide

8. Do you play computer games?

   - Yes
   - Sometimes
   - No

9. How many hours/week do you play computer games?

    - 0 hrs < 2 hrs
    - 2 hrs <= 8hrs
    - > 8 hrs

10. How familiar are you with immersive technology devices like the Microsoft HoloLens or Oculus Rift?

    - No experience
    - Minimal experience
    - Considerable experience

11. How familiar are you with Videoconferencing technology (Zoom, Skype, e.t.c)?

    - Very familiar
    - Moderately familiar
    - Somewhat familiar
    - Not very familiar

12. How many hours/week do you use videoconferencing technology?

    - 0 hrs < 2 hrs
    - 2 hrs <= 8hrs
    - > 8 hrs

13. How familiar are you with Web-based Virtual Environments (Mozilla Hubs, VirBELA, Spatial.io, e.t.c)?

    - Very familiar

- Moderately familiar
- Somewhat familiar
- Not very familiar

14. How many hours/week do you use Web-based Virtual Environments?
    - 0 hrs < 2 hrs
    - 2 hrs <= 8hrs
    - > 8 hrs

A.2 Post-Study Questionnaire

This section contains all of the questions that were included in the post-study questionnaire. This questionnaire comprises a modified version of the System Usability Scale (SUS) questionnaire [8].

1. Subject Code
2. I think I would like to use Zoom frequently.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
3. I found Zoom unnecessarily complex.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
4. I thought Zoom was easy to use.

- 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree

5. I think that I would need the support of a technical person to use Zoom.
   - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
6. I found the various functions in Zoom were well integrated.
   - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
7. I thought there was too much inconsistency in Zoom.
   - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
8. I would imagine that most people would learn to use Zoom very quickly.
   - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
9. I found Zoom very cumbersome (awkward) to use.
   - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
10. I felt very confident using Zoom.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
11. I needed to learn a lot of things before I could get going with Zoom.

- 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree

12. I felt present when I was interacting with other people in Zoom.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
13. Please let us know about any feedback that you have about your experience with the Zoom portion of this study.
14. I think I would like to use Mozilla Hubs frequently.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
15. I found Mozilla Hubs unnecessarily complex.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
16. I thought Mozilla Hubs was easy to use.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
17. I think that I would need the support of a technical person to use Mozilla Hubs.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
18. I found the various functions in Mozilla Hubs were well integrated.

- 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree

19. I thought there was too much inconsistency in Mozilla Hubs.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
20. I would imagine that most people would learn to use Mozilla Hubs very quickly.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
21. I found Mozilla Hubs very cumbersome (awkward) to use.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
22. I felt very confident using Mozilla Hubs.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
23. I needed to learn a lot of things before I could get going with Mozilla Hubs.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree
24. I felt present when I was interacting with other people in Mozilla Hubs.
    - 1) Strongly disagree, 2) Somewhat disagree, 3) Neither agree nor disagree, 4) Somewhat agree, 5) Strongly agree

25. Please let us know about any feedback that you have about your experience with the Mozilla Hubs portion of this study.

A.3 Task 1 Materials

The following script was used in Task 1 of the video conferencing/Virtual Environments study. The solid lines indicate locations where the subject was required to identify to most recent speaker.

Task 1: Script

Trauma Project Group Meeting: Debrief

Cast: 2 computer science graduate students working on developing a mixed-reality software application.

Nancy: Collaborator 1

Peter: Collaborator 2

Scenario: You are about to witness a team meeting involving four graduate students who are working on an emergency room simulator. The team has just finished meeting with their stakeholders – a medical doctor and his attending resident doctor. The team decided to meet to review feedback from the doctors and to prepare for their next meeting.

ZOOM MEETING

Peter: I think that meeting went pretty well. Nancy: Yes, they seemed to be really happy with our progress from the last meeting. They liked the new hand interactions that we added in. Peter: Yeah, thanks for adding those grabbable scripts to the stethoscope and the pulse oximeter. It was really fun to move them around using the MR headset. Nancy: Of course! I had fun writing and testing out the hand-tracking. I'm glad we were able to get that resolved. After all those rollbacks that we did, it's great that we were able to come up with a working build. Peter: Just in the nick of time too. Man… Mixed-Reality development can be tough sometimes. We spent all day yesterday pushing out builds to the headset. I wonder how hard it must be for those developers at Microsoft.

Nancy: Haha yeah it can be a pain sometimes. It just takes a lot of trial and error with some of these builds. But we did good today, so great job all! Now on to the agenda for this meeting. Do you have the feedback that they gave us during the meeting? Peter: Yes, I have it here. Nancy: Great, let's get to it then. Peter: Okay, so the first item that I have here is the placement for pulse indicators on the patient. Dr. Jones said that we might have redo the markers for the lower abdomen and the arms. Nancy: He also wanted us to add markers for measuring pulse around the patient's carotid, near their neck, and the back of their foot.

Peter: Got it. I can take a look at those. Do we have any images that we can use for reference? Nancy: I emailed Kimberly about that today morning. She's sent over a few images. I'll add these to the team drive, under the scenarios folder. Peter: Great, thanks, Nancy! Nancy: Sure thing! What else do we have? Peter: Breathing sounds need to be updated. We're still using the old audio assets from our last project.

Nancy: I found a database for lungs sounds earlier this week. I think some of them were really good. Let me share the link with you all. Peter: Some of these are paid. Nancy: Yes, let's use the free ones for now. We can check with Dr. Novak and pick up some of the better ones later this week. I can take care of this. Peter: Awesome! Thanks, Nancy! So, those were some of the items that they pointed out at the meeting today. Did I miss anything? Nancy: They also wanted us to switch up the patient from female to male to better fit the scenario. Any updates on this? Peter: Yes, the female model is a placeholder till we finalize the male model. We currently have two models that we are considering. They're both kind of pricey, so we'll have to get Dr. Novak's okay on whichever one we pick.

Nancy: I personally like the model that you showed everyone at the meeting. It's got a hospital robe on and it's rigged so we can animate it if we want. Peter: Yes, that would be useful for future scenarios where we need to patient to interact with the user. Let me check with Dr. Novak and get back to you all on this. Nancy: Sounds good. Peter: Great. Thanks for running us through all of the feedback, Nancy. How about we move on to setting up actionable items for our next meeting? (Everyone agrees) Everyone: express approval Narrator: The rest of this meeting will take place in a Mozilla Hubs meeting room. Everyone switches over to Mozilla Hubs.

MOZILLA HUBS MEETING ROOM

Peter: Okay, so to quickly recap all of the work that we've done so far. Nancy, could you please pull up the slides and give us a rundown of the items that we've completed over the past 3 months. Nancy: Yes, of course. (Pulls up a power point slide in Hubs) Nancy: So far, from the scenario, we've completed the introductory MIST step, Airway, Breathing and Circulation steps. For mixed reality features, we've added hand-tracking, voice commands, long-range item interaction, audio sources and we've started working on text displays. The text displays for the Circulation step should be ready by our next meeting.

Peter: Great. Thanks, Nancy! So, it looks like we want to focus on completing the Disability step by next meeting. I'll finalize the male patient asset and we can present it to Kimberly and Dr. Jones with the Disability step completed. Nancy: That sounds good. I can work on porting over some of our existing animations over onto the new model. Peter: Yes, the blinking animations that you added would be great to have. Nancy: And the toe wiggling. (All Laugh) Peter: Yes, I'll be sure to include the toe-wiggling too.

Nancy: Great. And Peter, you said that you wanted to work on the new pulse indicators? Peter: Yes, I can have these done by the end of the week. Nancy: Great! Now, let's see… are we forgetting anything? Peter: Dr. Novak talked about building an application that Kimberly and Dr. Jones could easily deploy onto their Mixed Reality headset. It would be nice if we could provide them with something like this at our next meeting. Nancy: Yes, I remember reading an article about something like this. Let me see if I can pull it up. (Takes a few seconds and then shares a browser tab with the article) Peter: This could work. Nancy, do you think you could look into this some more and give us an update next week?

Nancy: Yes, of course. Peter: Thanks! Peter: Okay, great. So, before we end for today, I just want to make sure that everyone knows what they need to get done before our next meeting. Let's all go around the room (clockwise -> Peter, Nancy) and state our tasks for the week. I can start. Peter: For this week, I will finalize the male patient model, import it into Unity along with the other new assets and work on deploying a test build for the MR headset. Nancy? Nancy: I'll fix the breathing sounds for the patient and try to get more information on easily deploying applications to the headset. Peter: Awesome! Thanks! I'll see you in-person in next Tuesday and we can work on testing out the new features. Nancy: At the usual time? Peter: Yeah at 2:30pm, as usual. Nancy: Great, see you then! Bye!

End of Meeting

A.4 Task 2 Materials

The following questions were used in Task 2 of the video conferencing/Virtual Environments study:

Task 2: Script

Scenario: The user made to take on the role of Team Lead for the project from Task 1. They are given a set of slides to present from. They are given the following instructions: "You are the Team Lead for the Project Group meeting that you just witnessed. You and your team have been working on creating an ATLS (Advanced Trauma Life Support) simulator in mixed reality for over four months. At this point, you have completed about 25% of the project and expect it to be completed by December 2021. Imagine you are presenting a status report to a potential client interested in hiring your team for a new assignment. Provide an overview of the project and elaborate your progress so far (you will be given a set of slides to present from and are welcome to improvise during your presentation). Be on the lookout for any questions that the clients might ask. Clients will raise their hands when they want to ask questions. You want to respond to these as quickly as possible. Feel free to improvise with your answers. Take a few moments to familiarize yourself with the slides and let know when you are ready to proceed."

Slides:

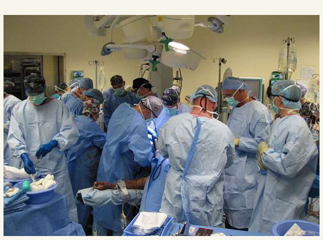
ATLS Simulator

Here is where our presentation begins

ATLS

- Advanced Trauma Life Support® (ATLS) is a training program for medical providers in the management of acute trauma cases, developed by the American College of Surgeons. Its courses provide a safe and reliable method for immediate management of injured patients.
- The ATLS program is designed to teach medical practitioners a systematic, concise approach to the care of a trauma patient.
- It also covers how to arrange for a patient's inter-hospital transfer and assure that optimum care is provided throughout the process.

ATLS



Project Goals:

- Improve the diagnosis accuracy, coordination and simplify multitasking
- Evaluate the use of the AR checklist in ATLS.
- Create reusable medical training scenarios.
- Facilitate learning and mastery of medical techniques in a non-clinical settings.

ATLS Overview

- Incident History: MIST (Mechanism, Injury Pattern, Signs and Treatment)

- Initial Assessment and Resuscitation: ABCDEs (Airway, Breathing, Circulation, Disabilities, Environment/Exposure).
- Adjuncts to the Initial Assessment: Attach additional devices and get patient vitals.
- Secondary Survey: Examine Head, Neck, Abdomen, Pelvis, Musculoskeletal system, Radiographs, Report.
- Transfer to Definitive Care

Selected Scenario

Bike Accident:

Patient is rolled into trauma bay by EMS provider and report is given. "Patient is a 19 year old male who fell off his dirt bike, injuring his right leg." Patient complains of severe right leg pain. Learner assesses right leg and notices deformity in lower right leg. There is no open wound present. Learner asks for x-rays of right knee, right leg and right ankle. Xrays are shown and reveal fractured lower tibia and fibula. Learner applies splint and calls orthopedics for a consult. Secondary survey fails to identify any additional injuries.

Scene Breakdown

ATLS:

1) Step 1: Patient is rolled into the trauma bay.
   - Room is equipped with a bed, pulse oximeter, stethoscope, and cardiac monitor.
1) Step 2: MIST/Background: 19-year old male, fell off dirt bike, injured right leg.
2) Step 3: ABCDE assessment
   a) Airway: Ask patient to state name.
   b) Breathing: Listen to breath sounds bilaterally with a stethoscope.
   c) Circulation: Examine blood pressure and SaO2 levels
   d) Disability: Evaluate patient's GCS score.
   e) Environment/Exposure: Assess right leg and apply splint.
4) Secondary survey - does not reveal any new information

Scene Breakdown

Step 1:

The Patient is rolled into the trauma bay.

Scenario / Pre step:
There is a room that is equipped with two display monitors, pulse oximeter, ECG machine, Stethoscope and few drawers for the medicines.

Step:
The patient is lying on a stretcher with wheels and is rolled into the trauma bay. Take a moment to look around the room.

Post Step:
The patient and the student are in the trauma bay and the process of assessment begins.

Step 2

Display the MIST of the patient to the student.

Scenario / Pre step:
The patient is in the trauma bay.

Step:
The MIST information for the patient is read out to the patient.

	M- Mechanism		I- Injury Pattern	S- Signs	T- Treatment
Scenario	Patient Info	Case			
Patient 1 - Easier	Male, 19 years	Automobile Injury, fell off dirt bike	Injured Right Leg		.

Post Step:
The Student is ready to start the ABCDEs of evaluation.

Step 3a

Begin ABCDEs

Scenario / Pre step:
The patient is in the Trauma Bay. Open AR Checklist. Note first task - Airway

Step a: Airway
Assess Airway by asking patient to state their name. The patient clearly states “Ryan”. Record “patient airway” in notes and mark completed on checklist.

Post Step:
Examine checklist for next task - Breathing.

Step 3b

Step b: Breathing
Listen to breath sounds bilaterally with a stethoscope. Blue bubbles indicate where the stethoscope needs to be placed. There are equal sounds on both sides. Note: “bilateral breath sounds present.”. Mark Completed on checklist.

Post Step:
Examine checklist for next task - Circulation.

Step 3c

Step c: Circulation
Ensure that the patient is connected to the cardiac monitor with pulse oximetry in place. Ask for blood pressure to be obtained. The simulation provides the following vital signs, “Blood pressure 110/85, heart rate 98, respirations 16, and oxygen saturation 99% on room air.”

Assess pulses (radial, femoral, posterior tibial and dorsalis pedis) bilaterally - indicated with blue bubbles. Note and record “Pulses are 2+ bilaterally in all locations.” Mark Completed.

Post Step:
Examine checklist for next task - Disability.

Step 3d

Step d: Disability
Ask the patient to state where he is and to wiggle his toes. Patient states "in the hospital" and wiggles his toes on both sides.

Note: "GCS is 15." Mark Completed.

Post Step:
Examine checklist for next task - Environment/Exposure.

Step 3e

Step e: Environment/Exposure
Patient complains of severe right leg pain.

Assess the right leg and notice deformity in the lower right leg. Note no open wound present. Ask for x-rays of the right knee, right leg and right ankle. X-Rays are shown and reveal fractured lower tibia and fibula. Apply splint and call orthopedics for a consult. Mark Completed.

Post Step:
Examine checklist for next task - Secondary Survey.

Step 4

Secondary Survey - Examine Head, Neck, Abdomen, Pelvis, Musculoskeletal system, Radiographs, Report.

Scenario / Pre step:
The patient is lying on the bed. They have a splint on their right leg.

Step:
Examine Head, Neck, Abdomen, Pelvis, Musculoskeletal system, Radiographs. No additional injuries.

Post Step:
- Debrief, Diagnose and send to Secondary Care.

Suggestions & Questions

MOZILLA HUBS MEETING ROOM:

1. Collaborator 1 (Slide 2): What was the name of the organization that created the ATLS program?

   Answer: *The American College of Surgeons.*

2. Collaborator 2 (Slide 3): Are Emergency Rooms usually this crowded?

   Answer: *Yes, there are a lot of medics in the room when a patient is brought in.*

3. Collaborator 1 (Slide 4): What is the most important goal of your project?

   Answer: *To evaluate the use of the AR checklist in ATLS and create reusable medical training scenarios.*

4. Collaborator 2 (Slide 5): What is the final step for ATLS? **Answer:** *Transfer to Definitive Care.*

ZOOM MEETING ROOM:

1. Collaborator 1 (Slide 7): What are some of the most important steps in this breakdown?

   Answer: *Step 3: ABCDEs are the most important aspect.*

2. Collaborator 2 (Slide 9): What is wrong with the patient?

   Answer: *They fell off their bike and badly injured their right leg.*

3. Collaborator 2 (Slide 11): How will the user know where to examine the patient for breathing or to check their pulse?

   Answer: *Markers provided in the MR application.*

4. Collaborator 1 (Slide 15): Do you also take or ask for X-Rays during the secondary survey?

   Answer: *Not typically. Only if they are needed (based on injury)*

A.5 Task 3 Materials

The following prompts were used in Task 3 of the video conferencing/Virtual Environments study:

Task 3: Script

Begin in the Zoom meeting room. "After the presentation. You are having a casual chat with the other team members."

ZOOM MEETING ROOM:

Questions:

1. What was the most interesting aspect of this project? (subjects speak in alphabetical order)

2. What kind of experience have you had with Immersive Technologies? (reverse alphabetical order)

MOZILLA HUBS MEETING ROOM

Questions:

1. Where do you see immersive technologies like AR/VR going in the future? (reverse alphabetical order)

2. Would you like to work on/participate in other AR/VR projects like the Trauma project? (alphabetical order)