Shared Mixed Reality Spaces:
Visualization and Interaction Techniques for Hybrid User Groups

Jann Philipp Freiwald

Dissertation zur Erlangung der Würde des Doktors der Naturwissenschaften der Fakultät für Mathematik, Informatik und Naturwissenschaften, Fachbereich Informatik der Universität Hamburg vorgelegt von Jann Philipp Freiwald aus Hamburg.

2022

Prüfungskommission

Vorsitz: Prof. Dr. Timo Gerkmann
Stellvertretender Vorsitz: Prof. Dr. Judith Simon
Erstgutachter: Prof. Dr. Frank Steinicke
Zweitgutachter: Prof. Dr. Ernst Kruijff

Datum der Disputation: 29.08.2022

Acknowledgements

I would like to acknowledge the people without whom this work would not have been possible. First of all, I would like to thank my supervisor, Prof. Dr. Frank Steinicke, for the guidance he has provided throughout my time as his Ph.D. student. He consistently allowed this thesis to be my own work, but steered me in the right direction whenever he thought I needed it. Furthermore, I would like to thank the members and alumni of the Human-Computer Interaction research group at the Universität Hamburg. Special thanks go to the people I have directly collaborated with (in alphabetical order): Dr. Omar Janeh, Dr. Nikolaos Katzakis, Dr. Paul Lubos, Dr. Fariba Mostajeran, Dr. Susanne Schmidt, and Dr. Oscar Javier Ariza Núñez. I would also like to thank the research assistants for helping conduct our user studies, and likewise every person who participated in them.

Finally, I want to express my very profound gratitude to my parents, my friends, and to my fiancé, Michaela, for providing me with unfailing support and continuous encouragement throughout my years of study, and through the process of researching and writing this thesis. I am incredibly fortunate to have such considerate and caring people in my life who experienced all of the ups and downs of my research. This dissertation stands as a testament to your unconditional love and encouragement. Thank you.

Abstract

The metaverse is commonly described as a plethora of multi-user mixed reality (MR) applications, in which users can interact with the virtual environment and each other through shared MR spaces. However, several aspects of the metaverse are still unclear, and fundamental research in human collaboration and interaction in MR hybrid setups is required. We use "hybrid setups" as umbrella term for the stark variance between individual users with regard to their technical setup, used interaction forms and virtual representations. This includes aspects such as different hardware devices, sitting and standing configurations, appearance of avatars, as well as locomotion and its visualization. This dissertation examines these components from both technical and user experience perspectives, including metrics such as usability, sense of presence and cybersickness.

First, we present techniques that enable effective hybrid collaborations between a variety of devices. More specifically, two types of MR devices require further research in terms of integration into shared MR spaces: (i) mobile devices and (ii) video see-through head-mounted displays. To integrate mobile devices, video streaming and camera-based tracking technologies are combined in a smartphone application called *VR Invite*, that acts as a handheld viewport to a local or remote shared MR space. The results of the accompanying experiment indicate that the opportunity for direct interaction positively influences the sense of presence of co-located mobile users. For video see-through head-mounted displays, an image reprojection algorithm named *Camera Time Warp* is introduced, that stabilizes virtual objects and virtual human representations in the real world reference frame. This technique addresses one of the major drawbacks of video see-through devices, the perceptual registration error caused by the inherent camera-to-photon latency. The user study demonstrates a positive effect of the *Camera Time Warp* technique on the participants' spatial orientation and subjective levels of cybersickness.

Second, we investigate visualization techniques within shared MR spaces to support spatial awareness as well as the sense of spatial presence and co-presence. In the context of human-to-human interactions, techniques to visualize a user's gaze direction and their points of interest are compared. The experiment shows that visualizations which are coupled to virtual human representations invoked the highest grade of co-presence, while task performance was highest when additional viewports were provided. In another user study, we contrasted virtual human representations, as well as the visualization of their locomotion from first- and third-person perspective. The results emphasize the importance of natural-looking human gait animations to increase the observer's spatial awareness and co-presence.

Third, we introduce novel techniques for locomotion and its visualization as implementations of embodied interactions. Based on findings of our prior experiments, the focus for these locomotion user interfaces was the visualization of human gait from first- and third-person perspectives. A locomotion user interfaces for seated MR experiences called *VR Strider* is presented, which maps cycling biomechanics of the user's legs to virtual walking movements. An experiment revealed a significant positive effect on the participants' angular and distance estimation, sense of presence and feeling of comfort compared to other established locomotion techniques. A second confirmatory study indicates the necessity of synchronized avatar animations for virtual locomotion.

Lastly, we present a toolkit for continuous and noncontinuous locomotion with matching human representations based on intelligent virtual agents. This toolkit includes two techniques each for avatar visualizations, i.e., *Smart Avatars* and *Puppeteer*, as well as locomotion, i.e., *PushPull* and *Stuttered Locomotion*. *Smart Avatars* deliver continuous full-body human representations for noncontinuous locomotion in shared VR spaces. They can smoothly visualize *Stuttered Locomotion*, a technique which transforms continuous user movement into a series of short-distance teleport steps. *Stuttered Locomotion* is also applicable to our motion-based *PushPull* technique, which employs a dynamic velocity multiplier based on the user's hand pose. An experiment indicates that *Smart Avatars* were preferred over regular virtual human representations, while remote controlling one's own virtual body through *Puppeteer* produced body ownership close to first-person interactions. The experiment evaluation shows that both *PushPull* and *Stuttered Locomotion* significantly reduce the occurrence of cybersickness. In summary, this work offers novel techniques and insight for the implementation of hybrid MR setups with respect to virtual interactions and their visualization.

Zusammenfassung

Das Metaverse gilt als ein Zusammenspiel von Multi-User Mixed Reality (MR) Anwendungen, in denen die Benutzer:innen in gemeinsam genutzen MR-Räumen mit der virtuellen Umgebung und miteinander interagieren können. Viele Aspekte des Metaversums bleiben unklar, und es besteht grundlegender Forschungsbedarf im Bereich der Zusammenarbeit und Interaktion in hybriden MR-Setups. Wir verwenden "hybride Setups" als Überbegriff für Beschreibung der Varianz zwischen einzelnen Nutzern hinsichtlich ihrer technischen Ausstattung, der verwendeten Interaktionsformen und virtuellen Repräsentationen. Dazu gehören Aspekte wie unterschiedliche Hardwaregeräte, Sitz- und Stehkonfigurationen, Aussehen der Avatare sowie die Fortbewegung und deren Visualisierung. In dieser Dissertation werden diese Komponenten sowohl aus technischer Sicht als auch aus der Perspektive der Benutzererfahrung untersucht, einschließlich Metriken wie Benutzerfreundlichkeit, Präsenzgefühl und Cyberkrankheit.

Zunächst stellen wir Techniken vor, die eine effektive hybride Zusammenarbeit zwischen verschiedenen Geräten ermöglichen. Insbesondere zwei Arten von MR-Geräten bedürfen weiterer Forschung im Hinblick auf die Integration in gemeinsam genutzte MR-Räume: (i) mobile Geräte und (ii) Videodurchsicht Head-Mounted Displays. Um mobile Geräte zu integrieren, werden Videostreaming- und kamerabasierte Tracking-Technologien in einer Smartphone-Anwendung namens *VR Invite* kombiniert, die als tragbarer Viewport für einen lokalen oder entfernten gemeinsamen MR-Raum fungiert. Die Ergebnisse des begleitenden Experiments deuten darauf hin, dass die Möglichkeit zur direkten Interaktion das Präsenzgefühl von mobilen Nutzern, die sich im selben physischen Raum befinden, positiv beeinflusst. Für Videodurchsicht Head-Mounted Displays wird ein Algorithmus zur Bildreprojektion namens *Camera Time Warp* eingeführt, der virtuelle Objekte und virtuelle menschliche Darstellungen im realen Referenzrahmen stabilisiert. Diese Technik behebt einen der Hauptnachteile von Videodurchsicht-Geräten, nämlich die durch die inhärente Kamera-zu-Photonen-Latenz verursachte Abweichung zwischen realen und dargestellten visuellen Informationen. In einer Nutzerstudie haben wir einen positiven Effekt der *Camera Time Warp*-Technik auf die räumliche Orientierung der Teilnehmer und das subjektive Ausmaß der Cyberkrankheit zeigen können.

Zweitens untersuchen wir Visualisierungstechniken in gemeinsam genutzten MR-Räumen, um das räumliche Bewusstsein sowie das Gefühl der räumlichen Präsenz und Ko-Präsenz zu unterstützen. Im Kontext von Mensch-zu-Mensch-Interaktionen werden Techniken zur Visualisierung der Blickrichtung der Benutzer:innen und ihrer Interessenpunkte verglichen. Die Ergebnisse des Experiments zeigen, dass Visualisierungen, die an virtuelle menschliche Repräsentationen gekoppelt sind, den höchsten Grad an Kopräsenz hervorrufen, während die Aufgabenleistung am höchsten ist, wenn zusätzliche Viewports bereitgestellt werden. In einer weiteren Nutzerstudie haben wir virtuelle Menschendarstellungen sowie die Visualisierung ihrer Fortbewegung aus der ersten und der dritten Perspektive miteinander verglichen. Die Ergebnisse unterstreichen die Bedeutung natürlich wirkender menschlicher Ganganimationen, um das räumliche Bewusstsein und die Kopräsenz des Betrachters zu erhöhen.

Drittens stellen wir neue Techniken für die Fortbewegung und ihre Visualisierung als Implementierungen verkörperter Interaktionen vor. Basierend auf den Erkenntnissen unserer früheren Experimente lag der Fokus für diese Lokomotions-Benutzerschnittstellen auf der Visualisierung des menschlichen Gangs aus erster und dritter Perspektive. Es wird eine Lokomotions-Benutzerschnittstelle für sitzende MR-Erfahrungen namens *VR Strider* vorgestellt, die die Biomechanik der Beine der Benutzer:innen auf virtuelle Gehbewegungen abbildet. Ein Experiment zeigte einen signifikant positiven Effekt auf die Winkel- und Entfernungseinschätzung der Teilnehmer, das Gefühl der Präsenz und den Komfort im Vergleich zu anderen etablierten Fortbewegungsmethoden. Eine zweite bestätigende Studie zeigt die Notwendigkeit synchronisierter Avatar-Animationen für die virtuelle Fortbewegung.

Schließlich stellen wir ein Toolkit für kontinuierliche und nicht-kontinuierliche Fortbewegung mit passenden menschlichen Repräsentationen auf der Grundlage intelligenter virtueller Agenten vor. Dieses Toolkit umfasst jeweils zwei Techniken für Avatarvisualisierungen, d.h. *Smart Avatars* und *Puppeteer*, sowie für die Fortbewegung, d.h. *PushPull* und *Stuttered Locomotion*. *Smart Avatars* liefern kontinuierliche menschliche Ganzkörperrepräsentationen für nicht-kontinuierliche Fortbewegung in gemeinsamen MR-Räumen. Sie können *Stuttered Locomotion*, eine Technik, die kontinuierliche Benutzerbewegungen in eine Reihe von Teleportationsschritten über kurze Distanzen umwandelt, fließend visualisieren. *Stuttered Locomotion* ist auch auf unsere bewegungsbasierte *PushPull*-Technik anwendbar, die einen dynamischen Geschwindigkeitsmultiplikator basierend auf der Handhaltung der Benutzer:innen einsetzt. Ein Experiment zeigt, dass *Smart Avatars* gegenüber regulären virtuellen menschlichen Repräsentationen bevorzugt wurden, während die Fernsteuerung des eigenen virtuellen Körpers durch *Puppeteer* zu einem Körperbesitz führte, der den Interaktionen aus erster Perspektive nahe kam. Die Auswertung des Experiments zeigt, dass sowohl *PushPull* als auch *Stuttered Locomotion* das Auftreten von Cyberkrankheit deutlich reduzieren.

Zusammenfassend bietet diese Arbeit neuartige Techniken und Erkenntnisse für die Implementierung von hybriden MR-Setups im Hinblick auf virtuelle Interaktionen und deren Visualisierung.

Contents

I Introduction

II Fundamentals

III Hybrid Collaborations

VI Conclusion

List of Figures

List of Tables

Abbreviations

AI - Artificial Intelligence
API - Application Programming Interface
ATT - Attractiveness
ATW - Asynchronous Time Warp
AR - Augmented Reality
AV - Augmented Virtuality
BO - Body Ownership
CamWarp - Camera Time Warp
CSSQ - Cybersickness Susceptibility Questionnaire
CVE - Collaborative Virtual Environment
DAQ - Device Assessment Questionnaire
FOV - Field of View
FPS - Frames Per Second
GSR - Galvanic Skin Response
HMD - Head-Mounted Display
HUD - Head-Up Display
HQ-I - Hedonic Quality - Identity
HQ-S - Hedonic Quality - Stimulation
IK - Inverse Kinematics
IMU - Inertial Measurement Unit
IP - Image Plane
IPQ - iGroup Presence Questionnaire
IVA - Intelligent Virtual Agent
LAN - Local Area Network
LUI - Locomotion User Interface
MR - Mixed Reality
NASA-TLX - National Aeronautics and Space Administration Task Load Index
NMSPI - Networked Minds Social Presence Inventory
OST - Optical See-Through
PQ - Pragmatic Quality
RQ - Research Question
SDK - Software Development Kit
SLAM - Simultaneous Localization and Mapping
SSQ - Simulator Sickness Questionnaire
SUS - System Usability scale
SUSP - Slater-Usoh-Steed Presence Questionnaire

VE - Virtual Environment
VR - Virtual Reality
VST - Video See-Through
WIP - Walking-in-Place
XR - Extended Reality

Introduction

1. Introduction

1.1 Motivation

Mixed reality (MR) is an emerging technology with enormous potential to enrich many areas of our society, ranging from work to education and leisure [39, 175, 183]. In MR, the real world and virtual world overlap, such that physical and virtual objects can co-exist and interact with each other in real time [28]. The real world can be augmented with computer-generated visualizations that become visible through head-mounted displays (HMDs). These visualizations can be shared between devices, which enable users to perform collaborative activities with virtual objects or even within virtual environments (VEs). MR has several practical applications, including collaborative design in 3D space, gaming and entertainment, remote training, as well as novel forms of working. One use case has received enormous media attention recently, which is referred to as the "metaverse" [77, 126]. This term does not refer to a specific application or product, but describes an abstract concept for a futuristic vision of a globally shared virtual world.

The metaverse is a contemporary subject of interest and debate amongst both consumers and developers of MR platforms, as well as in business-to-business and business-to-consumer markets. This surge in interest was likely caused by the recently announced rebranding of Facebook, Inc., who will continue business as "Meta". Mark Zuckerberg, Meta's CEO, motivated this transition by the potential of the metaverse [126]. According to Zuckerberg, the metaverse will be the successor of the mobile internet. To enter these platforms, users simply put on a virtual reality (VR) headset, and dive into a globally connected virtual world. While current implementations of this concept are limited by available technology, the metaverse is envisioned to be comparable to fictional pieces such as Steven Spielberg's 2018 "Ready Player One" [285], a film based on Ernest Cline's novel from 2011 [68], where users live a second life in an artificial world, including a functional global society, social interactions, but also the ability to influence the real world through online purchases or services. Associated virtual stores offer virtual and real products, while integrated services allow sharing of messages between users similar to social media. User can pursue virtual job opportunities, or have leisure time with their social circle of friends and family, regardless of their physical location.

The metaverse is an ambitious concept, whose technical definition, as of today, remains rather vague. Despite this ambiguity, companies like Meta and HTC use the term in marketing campaigns for their products [72, 195], which lead to several media outlets labeling the metaverse as contemporary marketing buzzword [77]. The term "metaverse" has originally been coined in Neal Stephenson's 1992 science fiction novel "Snow Crash" as a portmanteau of the Greek word "meta" (engl.: *beyond*), denoting a distinction from the conventional and concrete world, and "universe", the all-encompassing [208]. In this novel, people enter an alternative world and interact in a three-dimensional virtual space with each other and with intelligent virtual agents (IVAs) [293]. Depending on interpretation of the term, a social platform could thus be described as a metaverse. In fact, several software manufacturers like Epic Games and Meta already made claims that their multi-user applications are metaverses [141, 150]. While not all of those companies utilize immersive technology, their products provide social platforms with 3D visualizations of virtual worlds and forms of interactivity between users. For example, Epic Games, the developer of Unreal Engine, offers virtual concerts that can be accessed through a shared VE [150]. Epic Games, Inc. CEO Time Sweeney gave statements regarding the future of their gaming platforms and metaverse, denoting that the metaverse will be defined by whoever acquires one billion users on their platform first [167]. Contenders are companies such as Epic Games, Microsoft and Meta, which all aim to be the de facto leader in setting the standards. However, he emphasizes his belief that the concept of the metaverse is similar to the internet, in that no single entity can own it.

On the other hand, Meta portrays the metaverse as a platform for social multi-user MR applications, which is in line with the aforementioned vision of an immersive virtual world to live in [126]. Meta's concept sets itself apart from other social platforms and app stores through a unified centralized access point. Here, each user only has one account, one appearance and one centralized payment processor to access all the services and subplatforms across all providers. Regardless of the infrastructure of the metaverse, it is clear that companies strive to create metaverse platforms, and first steps in their creation have already been taken.

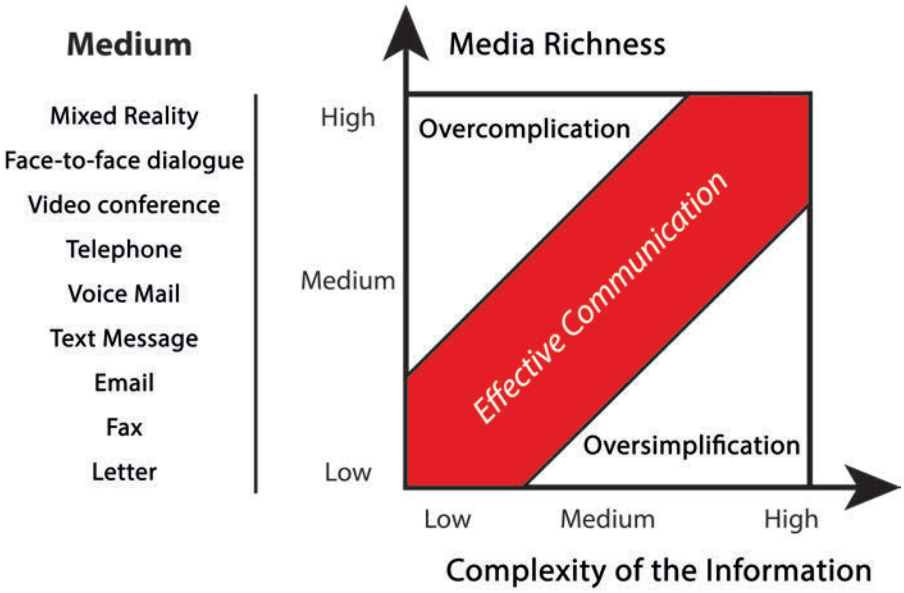


Figure 1.1: The media richness theory (illustration based on [75]).

It is yet unclear if immersive technology such as VR HMDs will be a criterion in the definition of a *true* metaverse. According to the media richness theory [74], depending on the complexity of the information to be conveyed, there is a suitable medium that can be used to communicate the content effectively (cf. Figure 1.1). In the original model, face-to-face meetings are seen as the pinnacle of human communication and information exchange [74]. However, In the context of social interactions, immersive technology has one key advantage over conventional two-dimensional mediums: it can provide a high degree of remote human-to-human information transmission, which may be on par or even exceed face-to-face meetings thanks to additional layers of information visualization.

This designation mainly stems from limitations in conventional remote synchronous interactions, as the possibility to convey information depends on the available human communication channels such as speech, body language, facial expressions and touch [132]. Immersive mediums can provide additional communication channels or increase the fidelity of existing ones through depth perception, body tracking capabilities and intricate haptic feedback. This technological immersion will allow users to experience the above illustrated futuristic vision of an alternative reality.

The prospect of a new era of human-to-human connection through digital worlds is an enticing concept. However, at the time writing this may be a distant future, and yet, the individual building blocks are already being developed. Devices for virtual, augmented and mixed reality such as Meta Quest [194] and Microsoft Hololens [209] represent the current state-of-the-art for immersive technology, and enable their users to virtually meet and interact in shared MR spaces. Connected through a sophisticated infrastructure through hubs or the metaverse, these spaces are at the center of each individual application that is provided to the user. Regardless of the success of the metaverse concept, multi-user MR applications offer a wide range of additional use cases to individuals, businesses and public institutions [39, 175, 183]. Their qualities in information transmission make them excellent tools for the "new work" approach, which stresses the importance of remote collaboration and distributed teams in the age of digitization [284]. For instance, remote mentoring could transform collaborative work as we know it. Through shared MR spaces, an expert could supervise remotely on a surgery procedure by providing guidance based on shared visual information. Mechanics of airplane engines, or general knowledge transfer in the educational and cultural area could benefit from this technology in a similar way. This is attractive from an economic standpoint, as the use of human resources can be optimized by minimizing personal travel. Less global travel is also ecological due to reduced CO2 emissions. If travel is required, MR could offer entertainment during transit in public transportation, in-flight or in autonomous cars. Evidently, people can experience a plethora of VEs and activities in shared MR spaces.

Their technical implementations have several components, which are shared amongst almost all of them. For instance, movement through the VE, known as locomotion, is an important activity [41]. Real walking is usually considered as the most intuitive way of locomotion in real and virtual worlds, and has been found to be more presence-enhancing compared to other forms of locomotion such as teleportation or joystick-based navigation [319]. However, there are limitations of using walking as locomotion user interface (LUI), for instance the limited size of the available walking space in the real world [331]. Due to this limitation, alternative interfaces and techniques such as the aforementioned teleportation and joystick-based navigation are often required. Their limitations, on the other hand, include a reduced sense of self-motion, limited sense of presence, inferior spatial cognition, or more frequent occurrences of cybersickness-related symptoms [54]. Therefore, one major contribution of this work is the introduction of novel LUIs that surmount those limitations for setups with limited available space.

Likewise, another major component is the visualization of connected users in the form of human representations, so-called avatars [98]. Their visual form can range from simple geometric shapes to highly detailed human 3D scans. Having an understanding of the spatial relation between users and points of interest in the VE is critical for effective collaboration [14], which includes anticipation of other users' imminent movement. While previous studies reported that the appearance of an avatar has an effect on the sense of co-presence and spatial awareness of observers [58, 244], the visualization of locomotion through embodied avatars and its effect on observers remains a largely unexplored field of research. In this dissertation the main components for embodied interactions in shared MR spaces will be discussed and several advancements of the state-of-the-art will be proposed. The technical hurdles that need to be overcome in order to achieve truly immersive shared MR spaces motivated this work, which will have a focus on the aforementioned components from both first and third person points of view.

1.2 Research Questions

To summarize, a true metaverse would likely consist of seamlessly connected shared experiences through multi-user MR applications. These applications have several critical components in common, independently of their use cases. However, several aspects of the metaverse proposal are still unclear, as missing pieces remain in the fundamental research of human collaboration and interaction in MR hybrid setups. In the context of multi-user MR applications, "hybrid setups" is an umbrella term for the stark variance between individual users with regard to their technical setup, used interaction forms and virtual representation. This includes aspects such as different hardware device types, sitting and standing configurations, avatar appearance, locomotion types and different visualizations of said locomotion. Accounting for all of these factors is a technical challenge for developers, especially so as some of them have yet to be researched in detail. This dissertation will examine these components from both technical and user experience viewpoints, including metrics such as usability, sense of presence and cybersickness. The contributions are grouped into three parts: (i) enabling hybrid collaborations, (ii) shared experiences, and (iii) interactions through embodied avatars. In particular, the following topics will be discussed:

i) With regard to enabling hybrid collaborations, two types of devices require further research in terms of multi-user MR integration: mobile devices and video see-through (VST) HMDs. The first part is dedicated to integrating them with VR and optical see-through (OST) augmented reality (AR) devices in order to broaden the accessibility of shared MR experiences. The following research questions (RQs) will be addressed:

- *RQ_1: How can bystanders and users of mobile devices be integrated into shared MR experiences, and how should they be represented?*
  Shared MR spaces are not necessarily occupied exclusively by HMD users. Such spaces can also be used for visualizations and interactions between co-located people, where one or multiple bystanders either observe an immersed HMD user or simply desire individual viewpoints within a VE and on its content. Conveying the content of shared MR spaces to observers and non-immersive users is challenging, as is providing direct participation.

- *RQ_2: Can the registration error in VST HMDs be decreased and will this have a positive effect on the emergence of cybersickness-related symptoms?*
  Another common type of device that enables participation in shared MR spaces is the VST HMD. Compared to real humans within one's field of view (FOV), virtual human representations are easy to identify in VST devices, even when using high resolution human 3D scans. Their position is not stable within the visible real world reference frame due to camera-to-photon latency, which is a detriment to both the technical immersion and psychological sense of presence.

ii) The second part will focus on visualization techniques with regard to shared spatial awareness as well as the sense of spatial and co-presence in shared MR spaces. The following questions are investigated:

- *RQ_3: How should the individual perspective of a user on different points of interest be communicated, independently of available avatar types and tracking capabilities, and what are the implications for an observer's spatial awareness?*
  Interactions with other people heavily rely on non-verbal communication. This is only possible to a lesser extent in shared MR spaces due to technical limitations, such as avatar

fidelity and accuracy of facial expression tracking, which may affect the expressiveness of body language. However, a shared understanding of spatial relations is important in many collaborative or competitive scenarios [14].

- *RQ_4: Does the depiction of continuous locomotion induce a stronger sense of spatial and co-presence and spatial awareness than abrupt teleportation movements?*
  The ability to induce a high sense of spatial and co-presence is a major benefit of shared MR spaces over traditional non-immersive online conferences. Previous studies reported that the appearance of an avatar has an effect on the sense of co-presence of observers [58, 244], but a yet mostly unexplored aspect of human representation is the effect of locomotion visualization on observers with regard to spatial awareness.

iii) The third part introduces novel techniques for locomotion and its visualization. In particular, the following questions will be answered:

- *RQ_5: How can a locomotion interface for seated users provide the illusion of walking? What are the implications for the sense of spatial and co-presence, and how should it be portrayed to observers?*
  Convenience of usage is an important aspect of any form of end-user technology. Prolonged sessions of HMD's quickly become tedious, as users typically have to stand stationary within their tracking space. Because of this, some experiences offer their users to remain seated. However, this has several implications, such as a lower point of view and potentially a reduced sense of presence or limitation of available locomotion types.

- *RQ_6: Can continuous locomotion be modified to incorporate merits of teleportation such as a high usability and reduction of cybersickness, and how can noncontinuous movements be visualized as continuous movements to observers?*
  Teleportation is one of the most-often provided locomotion types, even in shared MR spaces. This is likely due to a high degree of usability and efficiency while typically inducing only few cybersickness-related symptoms. However, continuous movements have been shown to be superior to teleportation with regard to both the sense of presence of its user and the co-presence of observers [54]. This is likely due to the abrupt visual nature of teleportation from both first- and third-person perspectives.

- *RQ_7: How can the user's point of view be transitioned to an outside perspective of their own body in order to replace self-movement?*
  MR experiences are often limited by their users' physical abilities. User interactions must be designed around the limitations of the human body, which prohibits intricate actions that can be found in non-immersive media, such as acrobatics performed by the protagonist in 3D games. However, remote controlling one's virtual body allows for super-natural user interactions.

The scope of this thesis does not include aspects of the metaverse such as economic factors, theories of computer-supported collaborative work, datacenter infrastructure or ethics of alternative virtual worlds. Its focus lies on shared MR spaces and the technical implementation of the above described components.

1.3 Outline

The descriptions of motivation and research questions above conclude *Part I* of this thesis. The remainder is structured as follows.

Part II describes fundamentals of this thesis as well as some generic tools developed for conducting user studies in MR. Chapter 2 describes relevant concepts such as avatars, locomotion, the sense of presence and cybersickness in VEs. Chapter 3 introduces the Cybersickness Susceptibility Questionnaire, which is followed by the description of the Remote Study Framework in Chapter 4.

Part III presents techniques that enable effective hybrid collaborations between a multitude of device types. Chapter 5 introduces a smartphone application that acts as a handheld window to a remotely connected VE. Chapter 6 describes an algorithm that improves image stabilization for VST devices to reduce the registration error.

Part IV explores shared experiences with a focus on visualization techniques. In Chapter 7 techniques to visualize the gaze direction of users are compared with regard to efficiency of information transmission. Chapter 8 evaluates different avatar and locomotion types and their impact on the sense of spatial and co-presence.

Part V reports insights on interactions through embodied avatars. Chapter 9 presents a novel LUI, which maps cycling motions of a sitting user to human gait animations. Chapter 10 introduces a toolkit for continuous and noncontinuous locomotion with matching human representations based on IVAs. Chapter 11 extends this toolkit with a technique to dynamically switch between first- and third-person perspectives.

Part VI discusses findings with respect to the research questions introduced above and concludes the thesis. Chapter 12 summarizes the main contributions of this work and presents a set of derived design guidelines. Chapter 13 presents opportunities for future research.

1.4 Publications

The main contributions of this dissertation have been published in peer-reviewed international journals and conferences, which are listed below:

1.4.1 Main Authorship

The following publications were mainly created by myself while co-authors contributed parts of the system implementation, writing of paper sections, or supervision.

1.4.1.1 Conference Papers

- ***Jann Philipp Freiwald***, Oscar Ariza, Omar Janeh, and Frank Steinicke. 2020. Walking by Cycling: A Novel In-Place Locomotion User Interface for Seated Virtual Reality Experiences. In Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1-12. [`https://doi.org/10.1145/3313831.3376574`]

- ***Jann Philipp Freiwald***, Nicholas Katzakis, and Frank Steinicke. 2018. Camera Time Warp: Compensating Latency in Video See-Through Head-Mounted-Displays for Reduced Cybersickness Effects. In Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST), pp. 1-7.
[`https://doi.org/10.1145/3281505.3281521`]

- ***Jann Philipp Freiwald***, Nicholas Katzakis, and Frank Steinicke. 2018. Camera Time Warp: Compensating Latency in Video See-Through Head-Mounted-Displays for Reduced Cybersickness Effects. In IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pp. 49-50.
[`https://doi.org/10.1109/ISMAR-Adjunct.2018.00032`]

- ***Jann Philipp Freiwald***, Sünje Gollek, and Frank Steinicke. 2021. VR Invite: A Project-Independent Smartphone App for VR Observation and Interactivity. In Human-Computer Interaction – INTERACT. Lecture Notes in Computer Science, vol 12932. Springer, Cham, pp. 352–370.
[`https://doi.org/10.1007/978-3-030-85623-6_22`]

- ***Jann Philipp Freiwald***, Julius Schenke, Nale Lehmann-Willenbrock, and Frank Steinicke. 2021. Effects of Avatar Appearance and Locomotion on Co-Presence in Virtual Reality Collaborations. In Proceedings of the Conference on Mensch und Computer, pp. 393-401.
[`https://doi.org/10.1145/3473856.3473870`]

- ***Jann Philipp Freiwald***, Lennart Diedrichsen, Alexander Baur, Oliver Manka, Pedram Berendjy Jorshery, and Frank Steinicke. 2020. Conveying Perspective in Multi-User Virtual Reality Collaborations. In Proceedings of the Conference on Mensch und Computer, pp. 137-144.
[`https://doi.org/10.1145/3404983.3405521`]

- ***Jann Philipp Freiwald***, Yvonne Göbel, Fariba Mostajeran, and Frank Steinicke. 2020. The Cybersickness Susceptibility Questionnaire: Predicting Virtual Reality Tolerance.In Proceedings of the Conference on Mensch und Computer, pp. 115-118.
[`https://doi.org/10.1145/3404983.3410022`]

1.4.1.2 Under Review

Chapters 10 and 11 are largely based on the following work, which is currently under review:

- ***Jann Philipp Freiwald***, Susanne Schmidt, Bernhard Riecke and Frank Steinicke. 2022. The (Non-)Continuous Toolkit: Rethinking Conventions for Locomotion and Its Visualization in Shared Virtual Reality Spaces. Submitted to SIGGRAPH ASIA 2022 and currently under review for publication in ACM Transactions on Graphics (TOG).

1.4.2 Co-Authorship

I was further involved in the creation of the following publications, where I contributed parts of the system implementation and writing of paper sections.

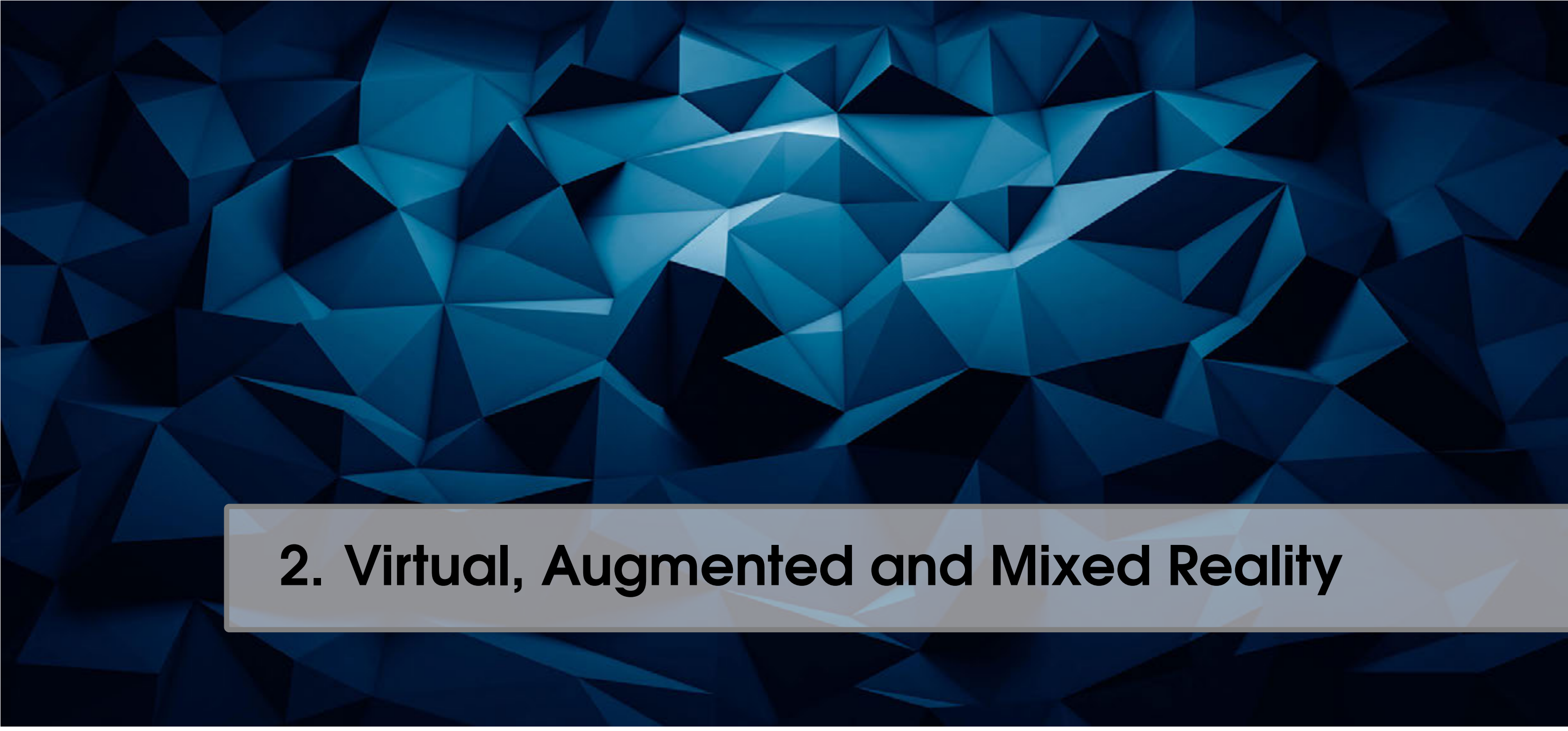
1.4.2.1 Conference Papers

- Oscar Ariza, ***Jann Philipp Freiwald***, Nadine Laage, Michaela Feist, Mariam Salloum, Gerd Bruder, and Frank Steinicke. 2016. Inducing Body-Transfer Illusions in VR by Providing Brief Phases of Visual-Tactile Stimulation. In Proceedings of the 2016 Symposium on Spatial User Interaction (SUI), pp. 61-68.
[`https://doi.org/10.1145/2983310.2985760`]

- Fariba Mostajeran, Nikolaos Katzakis, Oscar Ariza, ***Jann Philipp Freiwald***, and Frank Steinicke. 2019. Welcoming a Holographic Virtual Coach for Balance Training at Home: Two Focus Groups with Older Adults. In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 1465-1470.
[`https://doi.org/10.1109/VR.2019.8797813`]

II

Fundamentals



2. Virtual, Augmented and Mixed Reality

In this part, fundamental terms, technologies and concepts are described that will be referenced throughout the thesis. In particular, we explain the terms of reality-virtuality continuum, immersion and presence, MR technology, collaborative virtual environments, avatar representations, locomotion and cybersickness.

2.1 Reality-Virtuality Continuum

Our understanding of human-to-human interaction is largely based on the consideration of physical entities. However, this paradigm might shift due to the principle of extended reality (XR) and the interactions it enables. Before technical details and high level concepts can be discussed, the distinct terms virtual, augmented and mixed reality have to be clarified. The reality-virtuality continuum introduced by Milgram et al. [210] offers a taxonomy for these terms, illustrated in Figure 2.1.

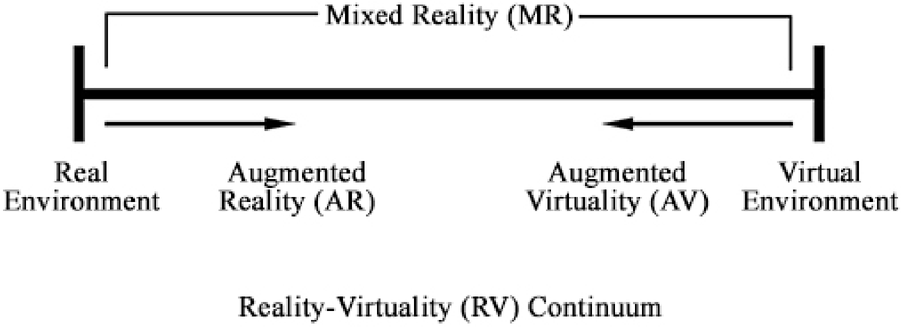


Figure 2.1: The reality-virtuality continuum (illustration by [275] based on [210]).

Milgram et al. portrayed these modalities as continuous spectrum, ranging from complete reality to complete virtuality. This continuum, therefore, encompasses all possible variations and compositions of real as well as virtual objects and environments. Four general segments are depicted on this scale, which blend into each other. These are:

- Real environment: the real world as we know it.
- Augmented reality: the addition of virtual information superimposed onto the real environment.

- Augmented virtuality: a VE enriched by real information from the real environment.
- Virtual reality: a fully computer-generated environment.

MR encompasses all variations where real and virtual elements are present, covering the entire continuum except for the extremes. This designation has recently been challenged by Skarbez et al. [275], who place VR in its state-of-the-art form (referred to as external VE) under MR. In their interpretation, only the futuristic vision of a "Matrix-Like" reality is beyond the spectrum of MR, a reference to the Wachowski siblings's 1999 film "The Matrix" and its computer-generated reality that is shared between all connected humans [327]. Their proposed taxonomy is illustrated in Figure 2.2. Both taxonomies focus on display technology; however, the reality-virtuality continuum is potentially also applicable to other modalities.

This work uses Skarbez et al.'s classification, and as such, shared MR spaces refers to multi-user applications with shared VEs that fall in the spectrum of AR, AV and VR. Additionally, the term "extended reality" refers to all real-and-virtual combined environments and human-machine interactions generated by computer technology and wearables [69]. Thus, XR is a superset, which includes the entire continuum from complete reality to complete virtuality.

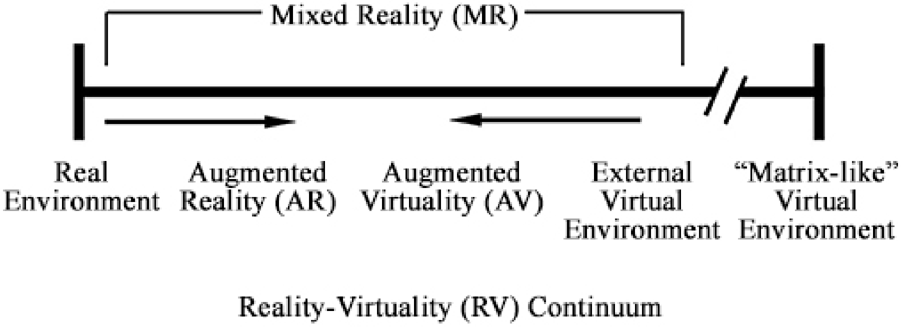


Figure 2.2: The reality-virtuality continuum revision [275].

2.2 Immersion and Presence

One major benefit of VR over traditional visualization media is the possibility for users to fully immerse themselves in an computer-generated environment, and be encapsulated in it. Hence, they can disconnect from the real world, and focus on the application. Through elaborate tracking and rendering techniques, spatial illusions are created, which may or may not operate by the same physical rule set as the real world. One major benefit of immersive technologies is the potential to invoke place and plausibility illusions for the users [278, 280]. This effect is achieved by employing devices that provide realistic and natural stimuli to the human sensory channels such as visual, auditory, haptic or even olfactory senses [55, 122, 164]. In this context, immersion denotes the technological capability of devices and VEs to establish the illusion of being physically present in an computer-generated space. Hence, immersion encompasses technical aspects such as image resolution, screen refresh rate, precision of 3D tracking or fidelity of haptic feedback. To a certain degree this applies to AR as well, where virtual objects are placed in the real world with ideally no visible distinction between real and virtual elements.

To which extent the provided place and plausibility illusions are accepted and interpreted as being part of the user's perceived reality, however, can strongly vary between individuals. As such, another term was coined to describe the degree of subjectively "being there", the *sense of presence* [138, 278]. In this work, the sense of presence, also referred to as "spatial presence", is measured through questionnaires, such as the Igroup Presence Questionnaire (IPQ) [268] or the Slater-Usoh-Steed Presence Questionnaire (SUSP) [277]. In these questionnaires, the user's perceived experience within the VE is usually contrasted with previous real world experiences [333]. For instance, the SUSP employs 7-point Likert scales ranging from 0 to 6 for items such as:

- "Please rate your sense of being in the VE, where 6 represents your normal experience of being in a place."
- "Did you perceive the VE more as images (0) that you saw or more as somewhere that you visited (6)?"

These questionnaires can be found in the appendix, Section 13. Presence can further be measured by employing physiological indices, behavioral feedback, or interviews [121]. According to Grassini and Laumann [121], physiological indices can be coarsely divided into brain-related and not brain-related. Among the brain-related measures, electroencephalography (EEG) is one of the most commonly used within the field of cognitive science and has found extensive use in relation to the sense of presence [306]. Non-brain-related indices include heart rate, galvanic skin response (GSR) and electromyography, which is a technique for recording the electrical activity produced by skeletal muscles. Furthermore, it is a sign of a higher level of presence when participants in a VE behave as if they are in a real environment. Examples of this behavioral feedback relate to conflicting multisensory cues that emanate from both real and simulated environments [276], postural sway [103], and responses to simulated stimuli [40]. Interviews are typically accompanied by questionnaires, which in tandem produce highly detailed data [142].

A number of perceptual factors have been reported to affect the perception of presence, ranging from sensory input to attentional, perceptual and other mental processes. Steuer et al. pointed out that presence has strong similarities with the phenomenon of distal attribution or externalization, which refers to the referencing of human perception to an external space beyond the limits of sensory organs themselves [294]. Similarly, Jerald et al. provided an analogous conclusion that presence is a subjective perception, which depends on external immersion and the user's current psychological state [157]. Subsequently, presence is a result which assimilates sensory stimuli with the current mental state and past experiences.

As presence depends on sensory stimuli, and the fidelity of that stimuli depends on the technical immersion, it can be assumed that immersion may limit the sense of presence [25, 44]. A higher level of immersion, e.g., by employing 3D sound and stereoscopic displays, has greater potential to elicit a higher sense of presence, as the provided sensory stimuli are close to previous real world experiences with regard to the stimuli's fidelity. A low level of immersion on the other hand might only offer abstractions or lower quality stimuli, causing a dissonance with previous experiences and therefore reducing the likelihood of the user accepting the presented place and plausibility illusions [6]. To summarize, immersion is an objective descriptor of a VE's technological characteristics, while presence is a subjective measure of experience in the VE.

2.2.1 Co-Presence and Social Presence

Besides the sense of (spatial) presence, *co-presence* denotes the sense of sharing a virtual space with other people. This term was originally coined in the work of Goffman who researched the social organization of real world gatherings [95]. He defined co-presence as a phenomenon between people that is best described as actively perceiving others, while in turn having the feeling that others also actively perceive them back. A high level of co-presence makes people appear approachable, available and interdependent [95]. In this sense, co-presence refers exclusively to a psychological connection to another person [222]. Zhao additionally distinguishes co-presence from social presence [342]. While co-presence and social presence partially overlap, social presence includes other factors such as psychological involvement, mutual understanding, and behavioral engagement [33]. According to Biocca et al. [32, 33], social presence has three main components:

(i) co-presence: the degree to which the users feel as if they are together in the same space.
(ii) psycho-behavioral interaction: the user perception of attention, emotional contagion, and

mutual understanding with their partner or participant.
(iii) (inter-)subjective symmetry: the degree to which the user perceives their level of social presence to be symmetrical or correlated with that of their partner's.

Therefore, social presence may refer to characteristics such as emotional perception and empathy, while co-presence is mostly limited to awareness of other people's presence. Similar to the sense of presence, co-presence can be limited by the degree of immersion. Additionally, factors such as avatar representations and the visualization of actions such as locomotion can influence co-presence [105]. It is usually measured through questionnaires such as the Networked Minds Social Presence Inventory (NMSPI) [32, 33]. This questionnaire is split into the three main components as described above, each divided into the perception of one self and the perception of others. For example, it includes the following pair of co-presence items:

- I often felt as if (my partner) and I were in the same room together.
- I think (my partner) often felt as if we were in the same room together.

In this work, the term *co-presence* is used in analogy to the presence definition of Slater et al. as the subjective feeling of being with and interacting with another person in a shared VE [267, 278]. Due to the potentially limited expressiveness of avatars, we will focus on co-presence in our user studies. When real others are visible in AR scenarios, we extend the questionnaires to incorporate aspects of social presence as well.

2.2.2 Relationships between Immersion and (Co-)Presence

Presence and immersion are fundamental not only to this work, but to MR in general. As such, these characteristics have been explored extensively.

Bowman et al. [44] established the dependence of immersion on the technical setup. According to them, both display hardware and software are major factors in determining the level of immersion. Components that could influence immersion include the FOV, display size, display resolution, stereoscopy, head-based rendering, realism of lighting, frame rate and refresh rate. Brown et al. [49] further suggested a fine-grained division of immersion into three subscales: engagement, engrossment and total immersion. Schuchardt et al. [269] examined the link between immersion and spatial awareness in VEs. Their results demonstrate that for certain tasks, immersion could significantly improve the participants' accuracy, task completion time, and spatial awareness.

McMahan et al. [206] argued that from a psychological viewpoint, the sensation of being immersed in an artificially generated VE is similar to being submerged in a large body of water. This experience takes over all attention and perceptual channels, and allows for exploration of new spaces, performing new actions and tasks, and being entertained by an interesting medium. The human brain is capable of isolating this experience, and the user's perceived reality takes place within this medium due to strong perceptual stimuli.

Besides the degree of immersion, subjective factors have been reported to influence the sense of presence. For instance, Kober et al. [172] investigated the relationship between personality variables and the sense of presence. The results of their user study suggest that the sense of presence correlates with personality traits such as absorption, mental imagination, perspective taking and immersive tendencies. Alsina-Jurnet et al. [11] support this notation, and add that the user's mental state likewise influences the sense of presence. Their study results indicate a correlation between presence and anxiety-inducing environments. Similarly, Riva et al. [258] elaborated on the effect of VEs on the user's mental state. VEs designed to purposefully induce anxiety or relaxation could produce a corresponding effect in participants. They concluded that the sense of presence was stronger in those scenarios, while presence conversely influenced the magnitude of an emotional response. Baños et al. [25] reported that both immersion and emotional content of VEs had a significant impact on the sense of presence in their user study. Their results indicate that immersion is a stronger factor for presence in non-emotional VEs than for emotional ones. These findings are supported by Schuemie et al.'s systematic literature review, which summarized correlations between the sense of presence and task performance, emotional responses, subjective sensations and cybersickness [270].

Bulu [53] investigated the relationship between spatial presence, social presence and co-presence. The results indicate that there are significant correlations between all of those characteristics, while social presence was the strongest predictor for participant satisfaction in the VE. Moreover, immersive tendencies were related to spatial and co-presence but not to social presence. Jo et al. [158] compared using real-world environments and VEs as space for remote collaboration in MR. Additionally, realistically reconstructed avatars were compared to simple visualizations. They concluded that the real world background and highly detailed human 3D scans promote a higher sense of co-presence.

2.3 Mixed Reality Technology

In this section, used technology for the implementation of novel concepts and the conduction of user studies is summarized. This includes the topics of immersive MR devices, the Unity engine and its networking stack.

2.3.1 Mixed Reality Devices

MR devices are headsets or handheld display devices for virtual objects and environments, which usually operate by providing computer-generated visualizations through video screens. The main principle is to render virtual elements in a 3D engine and display them in either the real world or in VEs, while the virtual viewpoint is constantly updated to align with the user's head pose. Depending on the degree of desired virtuality, the user's FOV has to be filled with either computer-generated or real visual information. For a fully immersive VR experience, it is necessary to remove all real-world information, which is usually achieved by covering the eyes with an opaque frame, housing a high pixel-density screen. On this screen, separate images for the left and right eye are displayed, which allow for the illusion of depth (so-called stereoscopy). Through concave lenses between eyes and screen, the user can focus on distant virtual objects. However, in order to see a clear image, the user's eyes have to constantly focus on the screen at a constant distance through accommodation, while their convergence has to adapt to the distance between eyes and virtual object. Therefore, accommodation and convergence do not necessarily coincide as they would in the real world. This phenomenon is referred to as the accommodation-convergence conflict and has been reported as one of the main factors in the occurrence of eye-strain or general cybersickness in VR headsets [199].

To add virtual elements into the real world, optical or video see-through approaches can be employed (OST and VST devices respectively) [266]. The central element of most OST displays (such as the Microsoft Hololens [209]) is an optical combiner with both transmissive and reflective characteristics, e.g., a semi-reflective mirror or a diffractive waveguide [19]. Due to its semi-transparent nature, the combiner lets light from the real environment pass through to the user's eyes. A projector or screen located at the side or on top of the lens supplies stereoscopic images that are reflected through the combiner, effectively superimposing virtual elements onto the real environment. Alternatively, transparent LCD or OLED displays in the user's FOV can be used to supply virtual information. The drawback of OST headsets is the semi-transparency of virtual objects, as full opacity requires total blockage of light from the real environment. The obvious disparity between real and virtual elements can be seen as a detriment to the technical immersion and psychological sense of presence.

VST headsets on the other hand rely on opaque video screens directly within the user's FOV, similar to VR HMDs. Many state-of-the-art VST solutions are in fact VR HMDs with additional cameras attached to their front side, for example the OVR Vision accessory [237]. Likewise, VR

HMDs such as the Oculus Quest and HTC Vive feature camera pass-through modes, where a video stream of the frontal camera setup is displayed in the headset. In essence, the partly transmissive lens of OST headsets is replaced by a video screen displaying the video stream of the stereo cameras. More precisely, the video stream is first routed into a computer that runs a 3D engine such as Unity. In this engine the video stream is used as background layer, while virtual elements are rendered on top before the final image is then transferred back to the headset's main screen. This approach allows virtual objects to completely occlude the real environment. Furthermore, VST devices enable augmented virtuality. By filtering the camera images, for example through computer vision and machine learning algorithms [9], parts of the video stream can be superimposed onto a VE. The entire visible FOV can be filled with either virtual or real elements, which is currently not the case for the OST approach. The major drawback of VST headsets is the camera-to-photon latency. Camera images have to be recorded, transferred to the rendering unit, be rendered, and transferred back to the screen before it can reach the user's eyes. The accumulative latency can reach up to 80ms, a noticeable delay [80, 237]. A potential solution to this drawback will be discussed in Chapter 6.

Not all AR displays are necessarily stereoscopic or head-mounted in nature. Rather, they can be classified according to their proximity to the user [31]. Displays can be placed directly in front of the user's eyes (i.e. HMDs), in arm-length distance (i.e., handheld devices), or at an arbitrary position inside the room (i.e., spatial displays). Handheld AR displays in particular gained enormous popularity in recent years, as the ubiquitous smartphone has put this technology right into people's pockets. Android and iOS devices offer developers software development kits (SDKs) for handheld AR solutions, namely Google's ARCore [119] and Apple's ARKit [149]. Smartphone AR is essentially an application of the VST technique, passing the phone's camera feed through to the AR SDK and subsequently to the screen. These SDKs can be accessed directly by developers or through engines such as Unity. Popular applications of this technology are Pokémon GO [153], Google Lens [151] and IKEA place [171], which have garnered millions of users globally [117]. As evidenced by the reality-virtuality continuum, handheld AR is a form of MR. Thus, the smartphone's ubiquity is an opportunity to incorporate handheld AR users into shared MR spaces. How this incorporation could be achieved will be discussed in Chapter 5.

2.3.1.1 Tracking solutions

As mentioned above, virtual elements are rendered from a constantly updated point of view to coincide with the user's perspective. Most MR devices feature inertial measurement units (IMUs) in every tracked component, e.g., headset and handheld controllers, to determine their pose. IMUs provide acceleration data, which can be translated to movements and rotations in 3D space. However, IMUs are prone to drifting due to double-integration of error. A shift in perspective that does not coincide with the user's proprioception might cause symptoms of cybersickness, which necessitates that the tracking has to be stabilized [188]. To counteract the drifting behaviour, an additional tracking system is usually required for accurate pose estimation in 3D space. There are three main approaches for this supplementary tracking: (i) camera-based outside-in tracking, (ii) laser-based outside-in tracking, and (iii) camera-based inside-out tracking.

Camera-based Outside-In Tracking The Oculus Rift was one of the first available consumer VR HMDs. Its tracking system relies on an external infrared camera and infrared LEDs distributed over the surface of the headset and controllers. Figure 2.3 (left) shows the pattern of LEDs on an Oculus Rift DK2, taken with an infrared camera. The stationary camera's angular aperture defines the maximum available tracking space. Using computer vision algorithms, the diodes are detected as individual markers and their constellation is reconstructed. The pose of objects with at least three markers in a distinctive arrangement can be calculated through a process referred to as rigid body tracking [111]. Figure 2.3 (right) shows a successfully reconstructed arrangement.

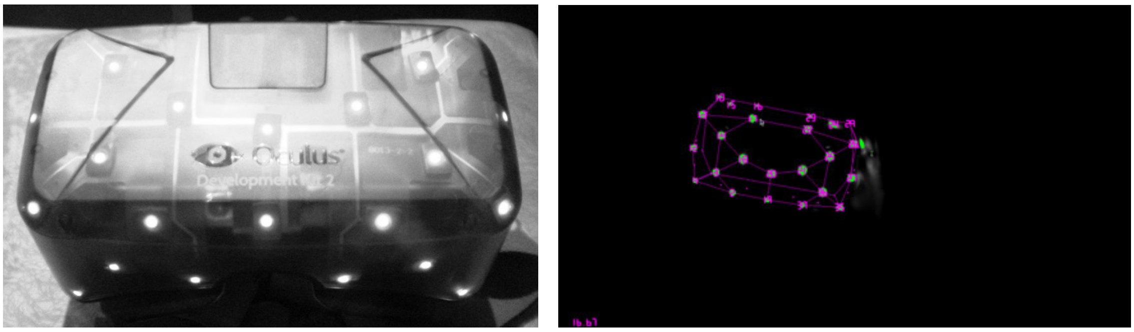


Figure 2.3: Infrared LEDs on an HMD (left) and pose estimation (right) [1, 85].

Laser-based Outside-In Tracking Another approach for 3D pose estimation is HTC's "Lighthouse" system, which operates similarly to a time-of-flight sensor [51]. Instead of placing the infrared emitters on the headset and using a stationary camera, the placement of those components is switched. Stationary infrared emitters (cf. Figure 2.4 (left)) are placed outside the boundaries of the tracking space, while the surface of the headset is covered in photo-sensors (cf. Figure 2.4 (right)). The emitters, known as "lighthouses", send out two types of light pulses: an omnidirectional pulse from an array of infrared LEDs, and a strong directional pulse from rotating infrared lasers.

When the headset detects an LED flash, a timer is started for each individual photo-sensor. based on the constant speed of light and the elapsed time between LED flash and the laser sweep hitting the photos-sensor, the distance of each sensor to the lighthouse can be calculated. The headset's pose within the tracking space is then estimated through the constraints of distances from all sensors to the emitter. Two lighthouses operating at slightly different frequencies are usually used in tandem to minimize occlusion within the tracking space.

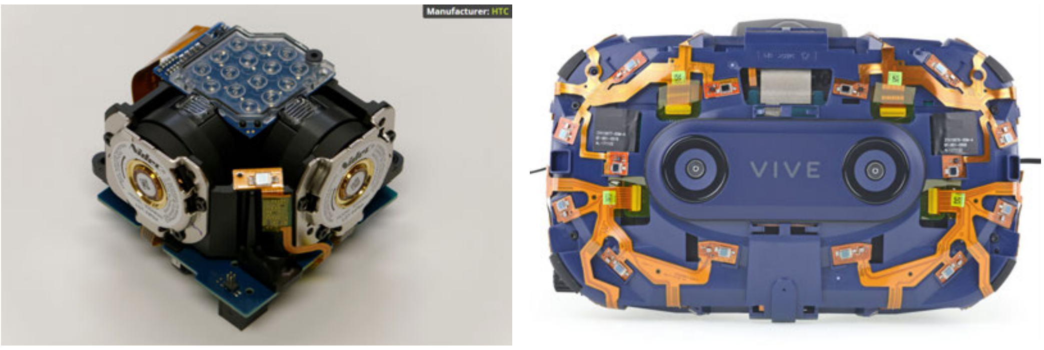


Figure 2.4: HTC Lighthouse (left) and Vive Pro (right) tracking components [201, 229].

Camera-based Inside-Out Tracking Lastly, mobile MR devices such as the Oculus Quest or Microsoft Hololens employ a tracking system without stationary components, so-called camera-based inside-out tracking. These types of headsets feature arrays of RGB and depth camera sensors that are spread out over their surface (cf. Figure 2.5). Through these cameras, a 3D reconstruction of the user's surroundings can be calculated. Computer vision algorithms for simultaneous localization and mapping (SLAM) are employed to determine the user's position within a previously established tracking space. This approach is computationally expensive and therefore taxing for mobile processors, which are usually optimized for power consumption over performance. Additionally, occlusion is a frequent problem. Hand gestures or handheld controllers are tracked through the headset's inside-out camera setup, and thus their pose can only be accurately calculated while in front of the user. Otherwise, the tracking defaults to IMU sensor input data.

Figure 2.5: The mobile MR headset Oculus Quest 2 with visible camera array [194].

2.3.2 Unity Engine

All presented concepts and user studies in this work were implemented in Unity, a cross-platform 3D engine first released by Unity Technologies in 2005. It is one of the most commonly used proprietary engines in both industry and research, with a total market share of more than 50% of all released applications, thanks to merits such as free licensing for small teams and an extensive list of supported devices [304]. Build-targets include platforms such as Windows and MacOS personal computers, game consoles, android and iOS mobile devices as well as MR headsets. It grants access to graphics application programming interfaces (APIs) such as DirectX and OpenGL, various shaders and other rendering tools, as well as physics simulations.

Game engines operate on the concept of an update-loop, where discrete states of a simulation are calculated based on a rule set of physics and predefined object behaviours. Each virtual object's state is determined by its last state and current influence factors such as gravity or user inputs. In each iteration of the update-loop, all virtual objects are updated once. Afterwards, the graphics API is used to render the scene from a given point of view, e.g. the detected pose of the MR headset. Once this image, referred to as frame, has been rendered, it is displayed on the device's screen and the update-loop is repeated.

Objects in Unity, officially named *GameObjects*, consist of several components. They all share a single base component, the transform component. It determines the object's pose in 3D space through a 3D vector for the position, a quaternion for its rotation, and another 3D vector for scale on each axis. There are several premade components that can be added to objects, such as rigid-body physics, audio emitters and listeners, light sources, or containers for 3D meshes and animations. Most importantly, however, developers can define their own custom components through C# scripts. These scripts contain C# class definitions, which will determine the logic and behaviour of their assigned object.

These *GameObjects* can be nested within each other, forming a hierarchy. Therefore, objects may consist of several sub-objects, each with their own set of components. Unity further offers their so-called prefab system to reuse object definitions, for example, when multiple user configurations are required. Prefabs are essentially a blueprint of an object and all of its components and sub-objects, which can be instantiated at runtime through scripts. This framework offers developers full access to the .NET environment, and hence allows for immensely complex applications. Typical programming patterns in Unity include singletons, observers, states and object pools.

2.3.2.1 Unity Networking Stack

Several networking stacks are available for Unity, such as the Mid-Level Networking API, Photon and Mirror, of which the latter was used for the implementations in this thesis. Mirror is an open source stack based on Unet, Unity's internal networking module. It is a multi-layered system of connection mechanisms, messaging services and network managers, as depicted in Figure 2.6.

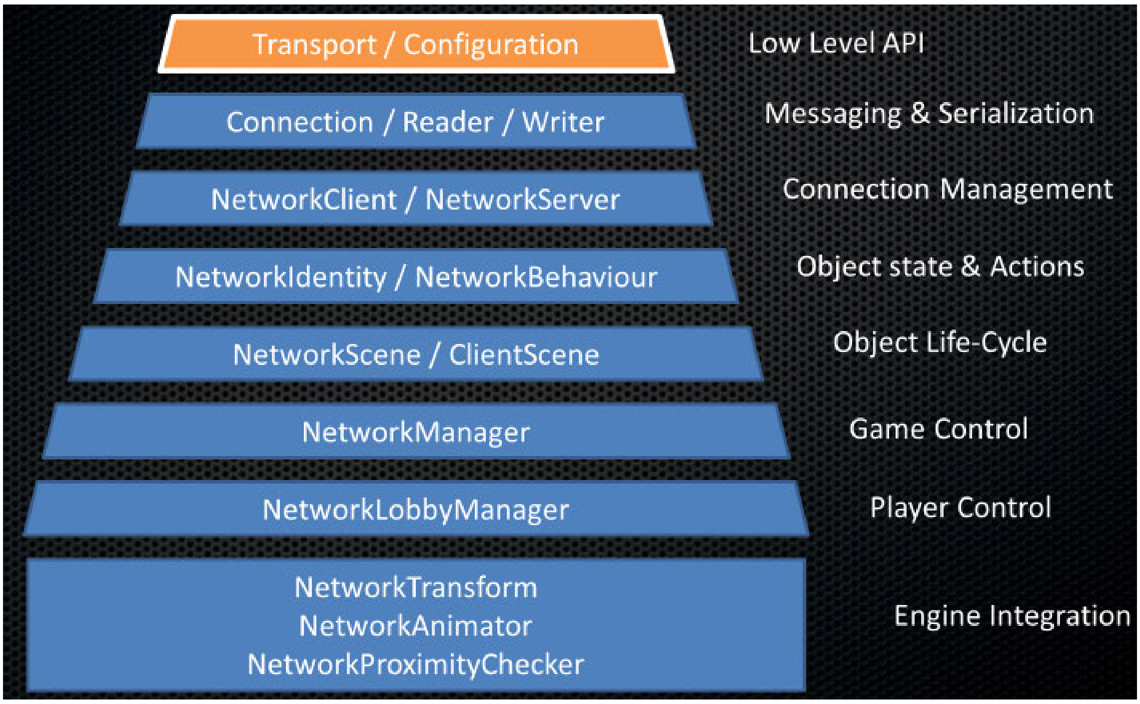


Figure 2.6: The Unity networking layers (illustration by [303]).

Starting with the Low Level API, Mirror provides connections and message exchange between multiple Unity instances in a network through protocols such as UDP. Readers and writers then abstract individual packages to readable messages, which can be used to create and manage sessions between host and clients. The following layers are then concerned with synchronization of Game Objects between server and client worlds.

Each synchronized object has to exist on each connected computer. Therefore, objects are instantiated on each computer at the same time with a shared identifier, the unique *Network Identity*. Through this identity, changes to one instance can be propagated to all other instances on every client. However, in the underlying authoritative server-client architecture only one client can have authority over each object, which is usually the host. Therefore, clients have to send change requests to the host, who then applies the state changes and synchronizes the result back to all clients. Change requests are handled through so-called *NetworkBehaviours*, which are custom components offering automatic synchronization of variables as well as client-to-server commands and server-to-client remote procedure calls.

NetworkScenes handle the life-cycle of synchronized objects on all clients at the same time. Before the shared *NetworkScene* is entered, participants can be gathered in a lobby through the *NetworkLobbyManager*. Within the *NetworkScene*, the *NetworkManager* acts as central control unit that keeps track of shared data and connections between host and clients. Lastly, predefined network components such as *NetworkTransforms* can be attached to object prefabs, making them synchronize states such as position and orientation automatically.

2.4 Collaborative Virtual Environments

Collaborative Virtual Environments (CVEs) are computer-based systems that enable collaboration and communication in a shared VE between users who may either use the same physical work space or remotely connect through the internet [67, 123]. In this context, VEs refer to a wide spectrum of visualization tools, ranging from simple two-dimensional canvases to elaborate MR spaces. According to Singhal and Zyda [343], the main features of a CVE are defined as follows:

- **Shared sense of spatial presence:** All users perceive the illusion of being in the same place.
- **Shared sense of co-presence:** Participants perceive a virtual representation of themselves and of others.
- **Real-time interaction:** Multiple users located in different physical locations can interact with each other in real-time.
- **Communication methods:** The VE enables some form of communication between participants. Communication can be implemented, for example, via gestures, text, or speech.

Their focus evidently lies on the aspect of human-to-human interaction. Another definition was presented by Sung et al. [299], which incorporates technical aspects as well. According to them, the following criteria must be met for an application to classify as a CVE:

- **Awareness:** Every user must be aware of the other participants.
- **Responsiveness:** Every action of the user must be communicated to the other participants as quickly as possible.
- **Consistensy:** The world must appear coherent and conform to expectations even though multiple users are actively manipulating and interacting with objects.

They further argued that consistency and responsiveness are opposing goals, while the focus too often lies on consistency. According to Pečiva [239], CVE applications maintain the data set that represents the VE, while the VEs are domain of computer graphics and VR. Essentially, CVEs extend regular VEs by shared data, which is consistent across all connected clients. These definitions are not limited to a work context and apply to any form of computer-based human-to-human interaction. Therefore, any synchronous application that fulfills these criteria, such as the discussed use cases of shared MR spaces, can be categorized as CVEs.

Manninen [202] further compiled and described the top level interaction form categories in immersive CVEs. These are avatar appearance, kinesics, occulesics, facial expressions, environmental details and language-based communication. Avatar appearance contains all forms of indirect interaction that is emanating from, e.g., the clothes, hair, physique or adornments of the avatar while kinesics describes postures, body movement, head movement or gestures. Because eye movement, eye contact and line of sight are such important aspects of human-to-human interaction, they have their own category, called occulesics. Finally, the environmental details category contains all forms of interaction with objects or the VE itself.

2.4.1 Data Synchronization

Since MR-based CVEs are often built upon proprietary 3D engines like Unity, their networking architecture usually relies on design principles that were established for their originally intended use case, i.e. 3D games [245]. In fact, many multiplayer 3D games fulfill the criteria to classify as CVEs, leading to several companies issuing claims that this makes their games qualify as implementations of the metaverse concept [150, 167]. Regarding the synchronization of data within CVEs, Pečiva's systematic literature review classified four categories of general networking schemes [239].

Figure 2.7 (A) schematically shows the first category of architectures, called "centralized primaries". One computer holds the primary instances of all objects (the "primaries", shown in red) and thus acts as the authoritative server. All other clients can make update requests, whereby the server manipulates the primary instances and then transmits the new state to all participants. The server has complete control over the session, guaranteeing consistency. This model is the de facto standard for state-of-the-art 3D engines, as evident by its implementation in every major Unity

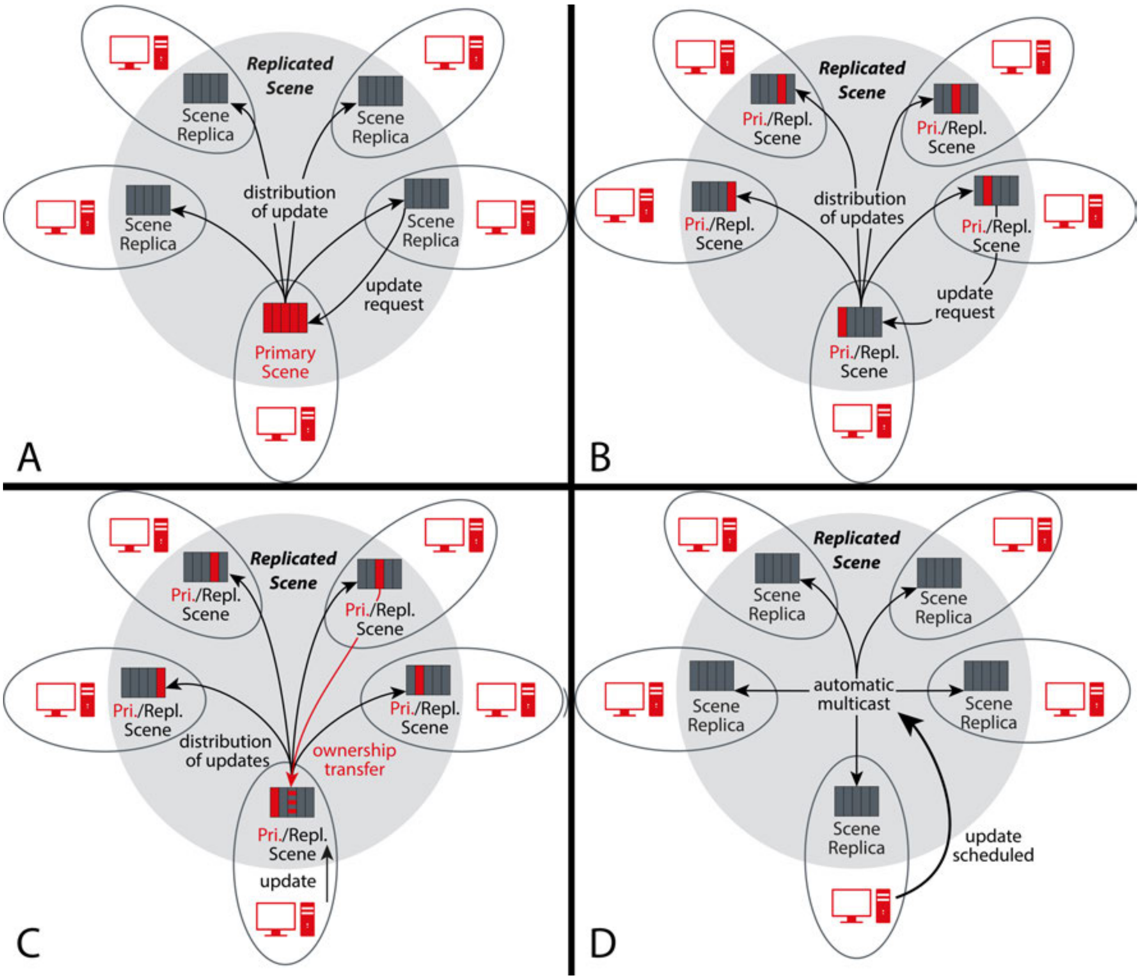


Figure 2.7: Cagetories of networking models (illustration based on [239]).

networking stack. However, Mirror allows the server to delegate authority over individual objects to clients, which increases the perceived responsiveness for user-controlled objects. In this case, changes are made locally first and then transmitted to the server, who then distributes them to the remaining clients.

This is similar in concept to the "distributed primaries" architecture (cf. Figure 2.7 (B)). Here, every client has authority over a portion of synchronized objects. clients broadcast change requests to all other participants directly, eliminating the need for a dedicated server. In this pessimistic concurrency control model, the requestor has to wait for confirmation from all receivers before a new request may be sent. As a result, the system's responsiveness may suffer. If changes are not confirmed but assumed, the model is instead referred to as optimistic concurrency. This is the case for the "Data Ownership" architecture, where authority over objects is passed between clients when changes are requested (cf. Figure 2.7 (C)). The fourth category, "Active Replication", is a peer-to-peer implementation (cf. Figure 2.7 (D)). Here, change requests are collected and broadcast on a timed update schedule. This approach, however, has no mechanism for solving conflicting instructions and therefore is exclusive to CVEs where each client may change only a specific subset of objects. According to Pečiva, the guarantee of synchronization in the centralized primaries approach and its robustness against malicious manipulation of local data makes it the most wide-spread architecture for CVEs [239].

2.5 Avatars

Avatars are a core component of immersive CVEs and subsequently shared MR spaces [299, 343]. An avatar is the representation of a human user in a VE, whose appearance may vary from simple geometric shapes to highly detailed human 3D scans. Through personalized appearances, they help identify a person and distinguish them from others [329]. While avatars are often limited to virtual hands in single-user MR experiences, more realistic representations have been reported to be beneficial for spatial awareness and interactions between multiple users in shared MR spaces [226].

2.5.1 Technical Implementation

In MR, avatars have a wide variety of implementations. As the above described tracking solutions only determine the position and orientation of the MR headset and its handheld controllers, it is difficult to accurately reconstruct the full body pose of the user. Therefore, many applications are limited to visualizations of head and hands, optionally with a torso that is placed directly beneath the head [226]. Additionally, devices such as the HTC Vive Facial Tracker offer real-time reconstruction of facial expressions [71]. However, configuring arms and legs for full body avatars, which are important factors of anthropomorphism and realism, remains a challenge for developers [227].

From a technical viewpoint, full body avatars are definitions for how animations affect the transform components of a humanoid character [272]. To animate a humanoid character, a 3D mesh and a skeleton are required. In this context, a skeleton is a hierarchical assembly of bone objects, imitating the structure of a real human skeleton. Each bone in this skeleton defines which vertices of the humanoid mesh are affected by it, usually based on a weight-by-distance function. Once a mesh is mapped to the bone structure and thus to the animation rig, animations that change the skeleton deform the character mesh.

A character usually has a collection of animations, including animations for walking, running, jumping, strafing, squatting, climbing and idling. In order to manage switching from one animation to another, Unity offers the Animator component. This component is a visual state machine, which works as a flow-chart of animation clips and transitions. While the Animator is in a certain state, the associated animation plays either once or on loop. Then, when the conditions of a transition pointing away from this state are met, for example when the animation has finished, or an external trigger has been set via script, the transition is executed and the state is switched. Transitions between states may be abrupt or smoothed through interpolation between the associated animations of current and next state at runtime.

Individual sections of the skeleton can be masked, which will inhibit an animation from changing this section of the 3D mesh. This property can be leveraged in conjunction with the Animator's multiple layer system, where each layer consists of a flow-chart, and all layers operate at the same time. Thus, one layer can be used to animate the top half of a body, and another layer the lower half. This allows for procedurally generated animations, for example, through inverse kinematics (IK).

2.5.1.1 Inverse Kinematics

IK is the use of kinematic equations to determine the joint parameters of a manipulator so that the end effector moves to a desired position. IK can be applied in many areas, including robotics, engineering, computer graphics and video games [16]. As joint parameters for a character are given in forward kinematics, e.g., by applying a predefined animation, the resulting position of each element from a kinematic chain, e.g., between shoulder and hand, can typically be calculated directly using multiple applications of trigonometric formulas. However, the reverse operation is, in general, much more challenging. Given a target pose for the end effector of a kinematic chain, IK finds the solution for all joint values to achieve this pose. Effectively, IK can be applied to a character's skeleton to place, for example, the hand in a given position and orientation relative to the shoulder. Through a series of constraints, such as limited elbow bending on certain axes, realistic body poses are formed. There are three general approaches for solving IK problems: (i) analytical solutions, (ii) numerical solutions and (iii) heuristic solutions:

Analytic Solutions Analytic solutions are closed-form expressions that take the end effector pose as input and provide joint positions as output. Analytical IK solvers can be significantly faster than numerical solvers and provide more than one solution, but only a finite number of solutions, for a given end effector pose.

Numerical Solutions Numerical solutions provide methods of modelling and solving IK problems. The most flexible of these methods typically rely on iterative optimization to seek out an approximate solution, due to the difficulty of inverting the forward kinematics equation and the possibility of an empty solution space. The core idea behind several of these methods is to model the forward kinematics equation using a Taylor series expansion, which can be simpler to invert and solve than the original system.

Heuristic Methods Heuristic methods perform simple, iterative operations to gradually lead to an approximation of the solution. The heuristic algorithms have low computational cost (return the final pose very quickly), and usually support joint constraints. Unity employs a heuristic algorithm called *Forward And Backward Reaching Inverse Kinematics* (FABRIK) [15]. FABRIK avoids the use of rotational angles or matrices, and instead finds each joint position via locating a point on a line. It divides the problem into two phases, a forward and backward reaching approach. Thus, it converges in few iterations, has low computational cost, and produces visually realistic poses, which appear smooth over multiple frames.

This technique can be applied to correctly position feet on uneven ground and slopes, or have an avatar mimic a user's tracked real world hand movements. Applications of this approach in shared MR spaces will be discussed in Chapters 9 and 10.

2.5.1.2 Real-Time Human 3D Reconstructions

A different approach to human representation is the use of real-time 3D reconstruction. Users are depicted similarly to video conferencing systems, but as 3D projections rather than 2D videos [236]. In the context of this dissertation, we implemented a system that is capable of constructing a 3D mesh of users through a setup of RGB and depth cameras. The main idea of this reconstruction is to take a reference image of an empty room, and compare it to the live video stream for each camera. Through this comparison, the background can be subtracted from each frame, leaving only the RGB and depth information of the user. We placed three Microsoft Kinect V2s beneath the ceiling of a room, approximately with a 120° between them and pointing towards the room's center. Each Kinect is connected to an associated computer that processes the background subtraction through a compute shader, which is a parallel processor for each pixel in an image, performed on the graphics card. The results are then transferred via LAN to a central server, which assembles the images into a 3D mesh.

Through known camera parameters such as intrinsic and extrinsic calibration, this server arranges a number of virtual projectors in a Unity scene. Each projector consists of a highly subdivided plane. For each pixel on this plane's material, the color from the RGB image is applied, while pixels that are marked as background are made transparent. Then, a vertex shader is applied based on the depth camera image, positioning the vertex of the associated pixel on the plane's Z axis. As a result, a 2.5-dimensional projection of the user is achieved. This process is performed simultaneously for all three projectors, effectively reconstructing the full 3D mesh.

This technique was employed in the study of Mostajeran et al. [214], where a reconstructed person was compared to realistic and abstract avatars for the visualization of a remote supervisor. The results indicate that users were generally positive towards remote supervisors. The reconstructed visualization was rated favorably over abstract avatars, with similar ratings to realistic ones. The developed video streaming technology was additionally used for the smartphone application that acts as a handheld window to a remotely connected VE, which will be discussed in Chapter 5.

2.5.2 Related Work on Avatars

Avatars enable effective spatial relations and interactions between multiple users in shared MR spaces [226]. In this context, Steed et al. have shown a positive effect of the appearance of an avatar as a user's self-representation on the sense of presence and user experience in single-user scenarios by providing virtual avatars to replace a user's physical body [289]. Moreover, Mohler et al. [211] have shown that displaying a virtual body improves the user's overall distance estimation, as it gives a sense of scale and relation.

According to Nowak et al. [226], avatars facilitate interactions with other users, IVAs or the VE itself and can shape computer-mediated communication experiences. For instance, users make judgments and attributions based on the names, appearance, and behaviours of others' avatars. Moreover, avatars may influence self-perception as well as both their owner's and observer's behaviours in interactions [189, 215, 228]. This phenomenon is referred to as the *Proteus effect* [339]. An example of this effect has been reported by Lee et al. [340], showing evidence that participants, who embody taller avatars, were more confident and aggressive when negotiating with another person. Avatars have been shown to have significant effects on several aspects of communication in VEs, such as interpersonal, group, environmental, non-verbal and organizational communication [7, 26, 139, 176, 320, 321, 328].

Through a systematic literature review, Nowak et al. [226] accumulated determining factors for an avatar's characteristics. These include user preferences, appearance, social norms or behaviors that are reflective of human capacities, experiences within the VE, and technological affordances or constraints of the system [34, 224, 297, 338]. They argue that observers employ similar processing strategies between real and virtual bodies due to their comparable characteristics as a perceivable representation of real others. According to Bailenson et al. [20], among the first judgments made of an avatar is determining *agency* or humanity, which leads to the assessment of the social potential of the represented user [216, 224]. In this context, hamilton and Nowak [135, 227] demonstrated that more human-like appearances were perceived to have greater social potential. In a VE, there may not be a clear visible distinction between IVAs and human users. Because of this, observers may not be able to distinguish agency [169, 223, 225]. However, previous works have shown that users try to determine whether an avatar is controlled by a real person or an IVA, and base their responses in interactions on this perceived agency [34, 35].

In shared MR spaces, the degree to which an avatar's appearance is similar to that of a real person is referred to as *realism*. This characteristic includes metrics such as the level of fidelity, fluidity of motion, or adherence to anthropomorphism [21, 227]. The degree of realism has been reported to have a significant influence on social and co-presence, communication satisfaction, and engagement of users interacting with the avatar [21, 46, 58]. For example, Casaneuva et al. showed evidence that gestures and facial expressions, or realistic human-like avatars in general, can increase co-presence in multi-user scenarios [58]. Waltemate et al. [329] further reported that personalized humanoid avatars made through photogrammetry received favorable ratings regarding body ownership and sense of presence compared to generic humanoid avatars. In addition to the appearance (mesh and texture), animations can be used to make the avatars look even more realistic or human-like, referring to the fluidity of motion characteristic. This includes realistic animations using IK [124] or full body tracking [286].

several MR social platforms offer full body avatar animation systems that are compatible with both IK and full-body tracking solutions. For example, VRChat's current Avatar 3.0 implementation uses Unity's Mecanim animation system to provide a customizable animation state machine with locomotion, gesture and IK layers to its users [324]. Lastly, avatar systems are not limited to shared VR spaces; besides typical non-immersive online games and social experiences, they can also be employed in MR telepresence setups [243]. Here, systems such as Holoportation or Remote Fusion

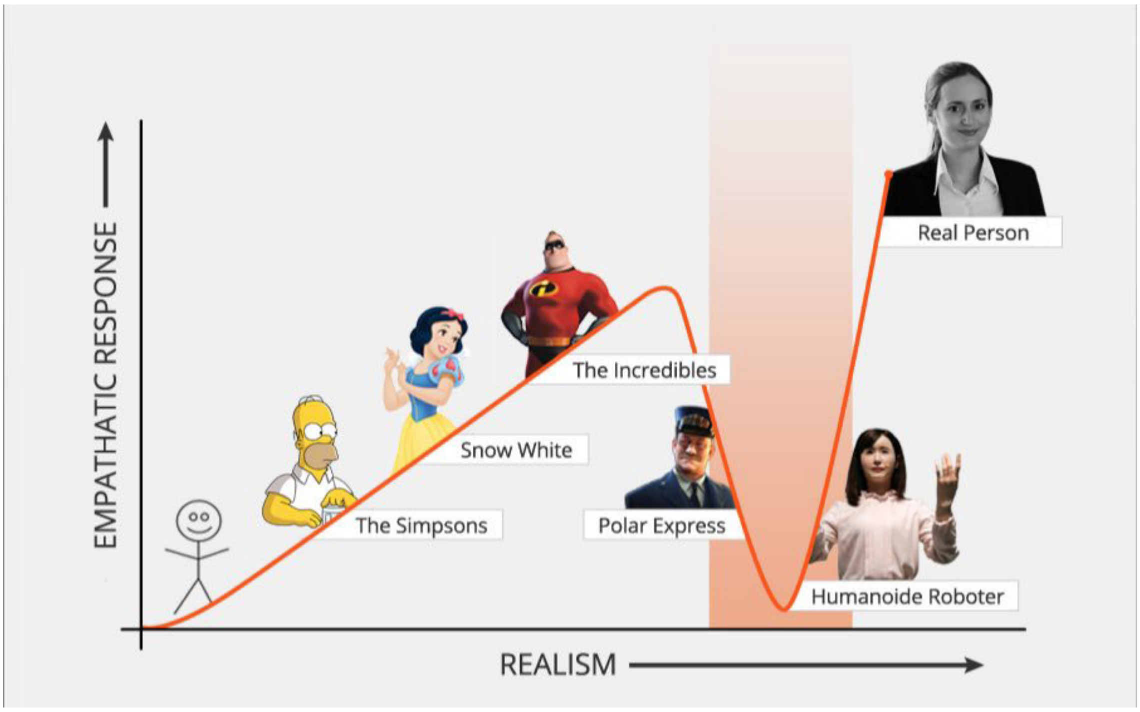


Figure 2.8: The uncanny valley (illustration by [168] based on [213]).

allow the use of real-time 3D human reconstructions in addition to generic or procedural avatar animations [2, 236, 305].

As discussed in Section 1.1, MR enables the enhancement of human communication channels to transmit more information than what is possible in the real world. This applies to avatars as well. For instance, gaze may be augmented to allow every receiver to see consistent eye contact from the same source at the same time [20]. Modifications to the avatar's shape are also conceivable; Won et al. [334] showed that additional body parts such as extra limbs may be utilized to expand non-verbal communication. This, of course, reduces anthropomorphism and thus realism.

To which extend such modified human representations may affect an observer can be derived from the uncanny valley effect. This phenomenon is best described as a function of an avatar's perceived pleasantness relative to its degree of realism [213]. Generally, the higher an avatar's realism, the more pleasant it is perceived to be (cf. Figure 2.8). However, a strong dip in this function occurs close to maximum realism, denoting that avatars which appear to be almost human-like, yet are perceivably not, are seen as unpleasant or even frightening. For this reason, avatars that fall into this uncanny valley are more likely to be rejected by users [125, 271]. Moreover, users expect anthropomorphic bodies to be animated authentically, again referring to the fluidity of motion characteristic [78]. This aspect of human representation will be further explored in Chapters 8, 9 and 10.

2.6 Locomotion

MR experiences typically allow users to interact with virtual objects, often through direct manipulation with their hands. However, not every object is within the user's personal space and thus their immediate reach. Therefore, designers of such experiences must offer at least one of the following options: (i) an interaction method for distant objects, or (ii) a LUI that lets the user move to the interactible object:

Distant Interactions In some cases, especially single-user scenarios, interactions can be provided that bring the interactible object to the user. For example, the VR game Half-Life: Alyx [73] offers so-called gravity gloves, which allow the user to aim at a distant interactible object and pull it towards them. Alternatively, stationary experiences such as Beat Saber [110] use a technique where every interactible objects appears at a distance, and automatically approaches the user in order for them to interact with them. Other techniques involve direct selection or manipulation from a distance, such as employing gaze tracking to select the interactible object, and then remotely control it through button presses or gestures [242].

Locomotion User Interfaces In many applications the MR interaction space is larger than the physical tracking space. Depending on the scenario, not all interactible objects can be moved to the user or be interacted with from a distance, for example, in shared MR spaces that feature a large-scale VE, where geometry occludes interactible objects. Some experiences employ location-based interactions, such as story-driven games where the user needs to travel in the VE to reach an objective [73]. These cases necessitate the use of user locomotion. While distant interactions can be partially employed in shared MR spaces, repositioning within the VE is often required to change perspective on points of interest, as is moving through a VE that is more expansive than the physical tracking space. Evidently, movement through a VE is an important activity in MR, including shared MR spaces [41].

In a systematic literature review, Boletsis [37] compared the most commonly used or studied locomotion techniques for MR applications. He distinguishes four different locomotion types:

- **Room-scale-based locomotion:** This motion type uses physical motion to provide interaction and supports continuous motion. However, unlike other types, the MR interaction space is limited by the size of the real environment and thus offers limited possibilities.
- **Teleportation-based locomotion:** The user's viewpoint is teleported to a predefined position without time delay as non-continuous motion.
- **Joystick-based locomotion:** Joysticks are used to continuously move the user's avatar and viewpoint in the VE.
- **Motion-based locomotion:** Techniques of this type provide continuous motion in VEs using some type of physical movement of users in the real world. Examples include swimming, climbing, flying, or walking-in-place.

Real walking within the user's tracking space, referred to as room-scale-based according to Boletsis, is enabled by the discussed tracking solutions in Section 2.3.1.1. The position of the HMD is projected into the MR interaction space, allowing the user to freely move within the VE, bound only by the physical limitations of the available tracking space. Real walking is usually considered as the most intuitive way of locomotion in real and virtual worlds, and has been found to be more presence-enhancing compared to other forms of locomotion such as teleportation or joystick-based navigation [319]. However, the physical limitations of the tracking space confine the effective interaction space within the VE [331]. Subsequently, the user's imminent reach in the VE is extended only by a short distance. Due to this limitation, additional or alternative interfaces and techniques are often required.

One of the most-often employed techniques to extend the user's reach through locomotion is teleportation [37]. It is offered in most major SDK provided by manufacturers of VR headsets such as the Oculus SDK [196] and SteamVR integration [315]. To trigger a teleportation, the user first needs to specify the target location. This is usually implemented through a laserpointer approach, either visualized as a straight line or a parable with limited reach, as a point-and-click metaphor [196]. The target destination is defined as the intersection of this laserpointer with a walkable surface area. The user confirms the destination by pressing a predefined button. Then, the user's viewpoint is moved instantaneously to the target location, usually keeping their orientation intact. A common technique to reduce cybersickness in this instance is to quickly fade the HMD screen to black before the movement, and fading back to normal vision afterwards [248]. In a literal blink of an eye the user has traveled to a distant location. This approach is highly efficient; it has been shown numerous times to have strong usability ratings, while being efficient for VE traversal [54]. However, its drawbacks are reduced spatial cognition, limited sense of presence and a low sense of self-motion [54].

Alternatively, there is Joystick-based navigation. The directional input on a joystick, usually integrated into the handheld controllers or wands, is translated to continuous movement. The mapping between the joystick's amplitude and the magnitude of the directional movement vector can be defined arbitrarily, but most often a full amplitude translates to speeds that are within the range of normal human walking or sprinting. This speed limitation is integral due to the well-known, but not yet well understood effect that continuous visual movements, which do not relate to the other human senses such as balance, inertia and proprioception, can cause severe cybersickness-related symptoms [186]. By slowing down the movements considerably, it is possible to mitigate some of the cybersickness-related symptoms [54]. This effect will be discussed in more detail in Section 2.7. Nevertheless, joystick-based navigation is an easy to understand mechanism that many users had prior exposure to, as joysticks are a standard input device for conventional gaming experiences.

Lastly, motion-based techniques can be employed. This type of locomotion translates the user's tracked posture, gestures or physical motion patters to continuous movements in the VE. This is by far the broadest category, and encompasses LUIs such as leaning [136], arm-swinging [205], stepper hardware [311], omni-directional treadmills [283] and walking-in-place [332]. Generally, these LUIs tend to synchronize proprioception with the user's experienced visual flow, and thus tend to be rated favorably with regard to cybersickness. However, to this day no singular motion-based technique found wide-spread adoption. This is likely due to their specific niche, requiring additional hardware or insufficient usability and perceived efficiency [56]. Because of this, teleportation and joystick-based navigation remain the de facto standards for MR locomotion despite their shortcomings.

All of the above described categories of locomotion have their merits and limitations. The major differences between them is the depiction of either continuous or noncontinuous movements, and their synchronization to other human perceptual senses [62]. Subsequently, Chapters 9 and 10 introduce novel LUIs that surmount the limitations for setups with limited available space. Additionally, a method to harness the advantages of both continuous and noncontinuous locomotion is discussed.

2.6.1 Related Work on Locomotion

In MR, the choice of locomotion has enormous impact on the user experience and interaction paradigms in general. Different techniques influence the interaction dimensions of immersion, flow, manageability, sense of effectiveness, and psychophysical discomfort (cybersickness) to varying degrees [38].

Room-scale-based locomotion includes real and redireted walking [184], while teleportation and joystick-based navigation represent the current default options for consumer VR headsets like the HTC Vive or Oculus Quest. Previous works have proposed solutions for the space limitations of room-scale MR, letting users explore a VE that is larger than their physical walking space, while retaining real world locomotion. A frequently mentioned approach is redirected walking, which is an umbrella term for methods that trick the walking user into the illusion of walking a straight line in the VE, while walking in circles in the real world [291]. This can be achieved for example through manipulating the rotation of the virtual camera [184] or by stimulating the user's leg muscles through electric impulses [18]. Steinicke et al. [291] quantified how much humans can unknowingly be redirected on physical paths that are different from the visually perceived paths. Their results show that users can be turned physically about 49 percent more or 20 percent less than the perceived virtual rotation, distances can be downscaled by 14 percent and upscaled by 26 percent, and users can be redirected on a circular arc with a radius greater than 22 m while they believe that they are walking straight. Suma et al. showed that instead of relying on redirected walking it is also possible to manipulate the VE by using physically impossible, overlapping spaces [298]. Their approach is however not universally applicable as the VE has to be designed with this concept in mind.

Funk et al. compared different approaches to teleportation locomotion in VR with regards to accuracy, task execution time and spatial orientation after the transition [107]. To this end, they proposed a targeting mechanism that lets the user define their orientation after the teleportation through a dial marker on the ground. The authors concluded that there is no significant difference in accuracy between straight, parabolic or curved targeting lines regarding accuracy, while reporting that the dial marker led to a higher execution time and less spatial orientation.

Several comparative studies of locomotion techniques regarding self-reported measures have been conducted to determine the most efficient and pleasant form of MR locomotion. Ragozin et al. performed an in-situ study to compare different types of locomotion regarding task execution time and cybersickness [247]. In their homes, participants had to play a VR game with either arm-swinging, joystick-based or walking-in-place locomotion. They reported that users finished the navigation task quickest with the arm-swing method, while walking-in-place was rated worst regarding the occurrence of cybersickness. Other studies on the other hand concluded that joystick-based navigation caused significantly more cybersickness-related symptoms compared to motion-based techniques such as leaning and pedalling [54, 104]. Nguyen-Vo et al. [218] compared multiple implementations of leaning techniques with real walking. They could not find any significant differences for task performance between upper-body leaning, full-body leaning and real walking. They concluded that leaning-based techniques can provide enough sensory information for supporting spatial updating, spatial awareness, and efficient locomotion in VR when synchronized with embodied translation cues.

Buttussi et al. [54] compared teleportation, joystick-based, and leaning-based navigation techniques in a VR travel task. They concluded that joysticks and leaning cause similar levels of cybersickness, while teleportation had no significant impact on the users' well-being. They could not find any significant differences for the self-reported sense of presence. Similarly, Xu et al. [337] examined distance estimation and spatial awareness in teleportation, joystick-based navigation and walking-in-place techniques and reported that no significant differences were found. Sarupuri et al. [264] further reported similar levels of induced cybersickness and usability between those techniques.

In a systematic literature review, Cherni et al. [62] found that over fifty percent of the reviewed techniques used body-centred self-motion cues. They concluded that these cues are a major factor in spatial orientation and awareness, while also reducing cybersickness [43, 255].

2.7 Cybersickness

Cybersickness is a fundamental issue of MR, which manifests predominantly in VR and VST headsets. Typically categorized as a form of visually induced sickness, it describes illnesses caused by observing computer-generated visual motion through immersive devices. The most commonly reported symptoms are general discomfort, headache, stomach awareness, nausea, sweating, fatigue, drowsiness, disorientation, and retching [174]. Because of this phenomenon, the time users can spend in immersive VEs through VR and VST devices is often limited [186]. Moreover, these symptoms are potentially dangerous as they can last for hours or even days after usage [287].

Stanney et al. [287] define cybersickness as a constellation of symptoms of discomfort and unease induced by VR exposure. They discussed the differences that can be found between cybersickness and simulator sickness, as these terms are often confused or falsely used interchangeably. To accurately define and separate these terms, a large number of self-reports were evaluated that contrasted symptoms caused by flight simulators and VR exposure. Three clusters of symptoms were identified: (i) oculomotor, (ii) nausea, and (iii) disorientation. In the flight simulator population, oculomotor symptoms were predominant, nausea symptoms less severe and disorientation symptoms the least pronounced. On the contrary, participants that were exposed to VR showed the opposite distribution of symptoms. In this group, the symptoms of disorientation were most strongly represented, while symptoms of nausea were less strong and oculomotor symptoms the least represented. Furthermore, symptoms of cybersickness were perceived on average three times as strong as the symptoms of simulator sickness [287].

Jerald [157] proposed *VR sickness* as an all encompassing term for motion, cyber- and simulator sicknesses caused by experiencing VR. According to Jerald, the specific cause of symptoms is not relevant for most VR users, and thus a general term suffices. In this work, we use the term *cybersickness* as defined by Stanney et al [287].

LaViola [185] discussed possible explanations for the occurrence of cybersickness in VEs. He investigated three theories for the emergence of motion sickness, and their applicability to cybersickness: (i) the postural instability theory, (ii) the poison theory and (iii) the sensory conflict theory.

Postural Instability Theory The postural instability theory is centered on the idea that one of the primary behavioral goals in humans is to maintain postural stability in the environment. In this case, postural stability is defined as the state in which uncontrolled movements of the perception and action systems are minimized [253]. Whenever the environment changes in an abrupt or significant way, in many cases, postural control will be lost especially if the control strategies are not available due to lack of experience.

Poison Theory The poison theory suggests that the ingestion of poison causes physiological effects involving the coordination of the visual, vestibular, and other sensory input systems [312]. The adverse stimulation found in some VEs can effect the visual and vestibular system in such a way that the body misreads the information and thinks it has ingested some type of toxic substance, thus causing symptoms which lead to an emetic response.

Sensory Conflict Theory The sensory conflict theory is the most widely accepted [81, 187, 312]. It is based on the premise that discrepancies between senses which provide motion and orientation cues cause a perceptual conflict within the body. For example, when the eyes register a visual flow through a VE, but the proprioception, also referred to as the sense of self-movement and body position, matches the experience of standing still. This is also true for VST HMDs, when the latency of the video feed causes mismatches between the user's vestibular system and the view in the headset; the head stops after the desired motion, but the delayed video stream suggests ongoing movement.

Furthermore, different amplifiers of cybersickness are presented in this thesis. These may be related to the display and the technology used for the VE, or errors in positional tracking, time delays between the user's movement and the movement in the VE, or flickering in the VE. Individual factors such as gender, age and diseases also affected the perception of cybersickness [185]. In this context, Bruck and Watters [50] conducted a study to find correlations between simulated motion and occurring cybersickness symptoms. They showed that the symptoms associated with cybersickness in a VE increase when the intensity of the simulated movement increases. To summarize, cybersickness elicits similar symptoms to motion sickness and simulator sickness, but has different triggers and a unique distribution of symptoms within the identified clusters.

as discussed in Section 2.6, cybersickness is a major drawback of continuous locomotion, whose intense effects lead to teleportation remaining as the most-often used technique despite its negative effect on the sense of presence and self-motion. In order to combat this drawback, several approaches have been proposed to make continuous locomotion a desirable alternative to teleportation. For instance, Fernandes and Feiner [101] presented a technique to prevent cybersickness by reducing the FOV during head movements in VR. The FOV was decreased by masking the peripheral vision, which is referred to as applying a vignette. They concluded that effects of cybersickness can be reduced and participants would be able to stay in the VE for longer periods of time. The disadvantage is the reduction of the FOV due to occlusion. Similarly, Nie et al. [219] showed that blurring the peripheral vision can mitigate cybersickness effects during head motions. Furthermore, the introduction of a novel LUIs often contains an evaluation of its effect on cybersickness [23]. According to Buttussi's literature review [54], many motion-based approaches receive positive ratings compared to joystick-based navigation, but are rated worse than teleportation in this regard.

As the general symptoms that can occur are the same between cybersickness and simulator sickness, albeit with a vastly different distribution, the Simulator Sickness Questionnaire (SSQ) [163] can be employed to measure cybersickness. The SSQ's usage was validated by Hirzle et al. [143], who performed a systematic literature review for general discomfort in VR applications. They concluded that additional items for eye strain and symptoms caused by the headsets' ergonomics should be included in a questionnaire that is specifically designed to be used in VR experiments. Nowadays, the SSQ consists of 16 items categorized in three subscales (oculomotor, nausea, and disorientation). Each item can be rated on a 5-point Likert scale ranging from 0 to 4. Weights are assigned to each of the categories and summed together to obtain a single score. Although the score is not intended to predict illness, it does provide a description of overall sickness scores for a given VE.

The ability to predict a user's susceptibility to cybersickness before usage has potential applications, however. As such, the following section presents a questionnaire that fulfills this purpose.

3. The Cybersickness Susceptibility Questionnaire

3.1 Motivation

As described in Section 2.7, cybersickness is one of the main limiting factors of VR adoption, as about 50-80 percent of all users are affected by symptoms of nausea, fatigue, headaches or vertigo [60, 170]. Just 20 minutes in a VE causes symptoms of illness in about 61 percent of users [252]. These symptoms often persist even after the VE is left and can last several hours [120]. This phenomenon, combined with the potential severity of symptoms, causes a high chance of affected users refusing future participation in VR studies or using VR systems in general [170]. This in turn may lead to a slowdown in development of immersive VR experiences and applications for training or health care [81].

In order to advance the development of VR applications, we propose the Cybersickness Susceptibility Questionnaire (CSSQ) that predicts a user's susceptibility to developing cybersickness-related symptoms. With this questionnaire, users can have a greater degree of confidence entering VR for the first time, and researchers performing studies in VR can select participants based on the study's purpose. Selecting especially susceptible participants might be desired for testing solutions that aim to reduce cybersickness-related symptoms, while selecting unaffected participants might be desirable for experiments where early, non-optimized prototypes of visualization, rendering or interaction techniques are tested. The items on this questionnaire were compiled from several previous works that have demonstrated the influence of various biological, chemical and psychological factors on the human senses and nervous system.

3.2 Cybersickness Susceptibility Questionnaire

With the CSSQ, a prediction about a person's tolerance to extended VR exposure can be made. The individual items on this questionnaire were compiled from previous related works that demonstrated an effect of the evaluated biological, chemical or psychological factor on the occurrence of cybersickness-related symptoms [159, 174, 185, 265, 273].

The Questionnaire consists of three categories, which are demographic data, health and fitness, and motion sickness triggers. First, demographic data is considered, consisting of age, height, gender and ethnicity [185]. Secondly, the CSSQ considers physical health. It is queried in the form of yes-no questions or five point Likert scales, referring to the frequency of each complaint's occurrence. These scales range from 0 to 4 with the labels "very rarely", "rarely", "occasionally", "frequently" and "very often". The frequency of headaches in the form of migraines and the frequency of stomach complaints were especially considered, as these can be an indication of a higher susceptibility of motion sickness and thus also symptoms of cybersickness [159, 273].

The yes-no questions refer to the recently experienced or current behaviour and well-being. The alcohol [265], medication [159] and drug consumption [273] of the last 24 hours are queried. As described above, this can have an influence on several neural systems of the body, altering the experienced severity of cybersickness-related symptoms. Chemical influence factors make predictions unreliable, as they can have both positive and negative effects on the sympathetic nervous system [159]. In addition, acute complaints and illnesses such as colds, influenza, ear infections, respiratory diseases, eye diseases and sleep deprivation are examined [159, 174].

For physical health, we also considered newly prescribed glasses, as the adoption of eye correction can have a direct influence on the vestibular and oculomotor system. Also, known disorders of the vestibular or oculomotor systems are queried in the form of medical conditions. We considered including items regarding hormonal fluctuations due to pregnancies or contraceptives as these were shown to influence susceptibility to cybersickness [174, 273], but ultimately removed them due to privacy concerns.

Lastly, we incorporated several distinct types of motion that are known to cause motion sickness in the real world. These are queried as Likert scales, denoting the frequency of complaints about motion sickness from individual types of motion. Motion sickness triggering events include driving as a passenger [81, 159], reading as a passenger [159, 182], steering small boats, traveling by train, traveling by train backwards, traveling by bus, traveling by bus backwards, flying in an airplane, sitting on a spinning swivel chairs, roller coaster rides, carousels and swings [159].

Following the SSQ, we employ 5-point Likert scales ranging from 0 to 4. Table 3.1 lists all items of the CSSQ, grouped by category and measuring type.

3.3 User Study

In this section the user study is described, which was conducted to find correlations between individual items of the CSSQ and the self-reported cybersickness through the SSQ [163]. As cybersickness and simulator sickness share general symptoms, the SSQ can be used to measure the induced cybersickness. For this purpose, a VE was created to recreate the motion sickness triggers queried in the CSSQ in the form of a roller coaster ride. The roller coaster was chosen due to its adaptability in simulating most of the triggering movement patterns through rotations around all available axis while driving at a constant or variable speed. Roller coasters were furthermore used in previous works to provoke cybersickness [82]. Henceforth, the different virtual movements that mimic a motion sickness trigger from the CSSQ will be referred to as *events*. To recognize cybersickness-related symptoms during individual events, the participants' skin conductance level was constantly measured during the entire experiment with a GSR test. The events were ordered in such a way that their intensity would increase over time, with 10 second breaks in between.

Based on these criteria, the experiment was designed as within-subject, and the following hypotheses were made:

H_1**:** There is a positive correlation between a high *CSSQ score* and a high *SSQ score*.
H_2**:** There is a difference between *SSQPre* and *SSQPost*.
H_3**:** There is a correlation between health claims and a high *SSQ score*.
H_4**:** Some motion sickness items will be a stronger indicator for a high *SSQ-Score* than others.
H_5**:** There is a difference between the *SSQ-Score* of female and male test subjects.
H_6**:** There is a correlation between demographic data and a high *SSQ score*.
H_7**:** There is a positive correlation between sense of presence and the *SSQ score*.

3.3.1 Participants and Apparatus

24 participants ($M = 24.79$, $SD = 4.863$, 13 female) took part in the experiment. The mean time per participant was about 30 minutes. The scene was rendered on an HTC Vive Pro.

Category	Question	Measure
Demographic	Age, Height, Gender, Ethnicity	Freetext
	How often do you use VR?	Likert Scale
Health and Fitness	How often do you suffer from a migraine headache?	Likert Scale
	How often do you suffer from stomach discomfort?	Likert Scale
	Did you consume alcohol in the last 24 hours?	Yes-No-Answer
	Did you consume drugs in the last 24 hours?	Yes-No-Answer
	Did you consume medication in the last 24 hours?	Yes-No-Answer
	Do you suffer from a cold or flu at the moment?	Yes-No-Answer
	Do you suffer from an ear infection at the moment?	Yes-No-Answer
	Do you suffer from a respiratory disease at the moment?	Yes-No-Answer
	Do you suffer from a lack of sleep at the moment?	Yes-No-Answer
	Do you suffer from an eye disease?	Yes-No-Answer
	Have you been prescribed new glasses recently?	Yes-No-Answer
	Do you suffer from a limitation of your vestibular system?	Yes-No-Answer
	Do you suffer from a limitation of your oculomotor system?	Yes-No-Answer
Motion Sickness	When I drive a car, I feel sick as a passenger.	Likert Scale
	Reading as a passenger makes me sick.	Likert Scale
	I feel sick when I am sailing.	Likert Scale
	I feel sick on small boats.	Likert Scale
	I feel sick when I go by train.	Likert Scale
	I feel sick when I go backwards by train.	Likert Scale
	I feel sick when I ride the bus or am a co-driver in a car.	Likert Scale
	I feel sick when I ride the bus sitting backwards.	Likert Scale
	I feel sick when I fly in an airplane.	Likert Scale
	I feel sick on rotating swivel chairs.	Likert Scale
	I feel sick when I ride a roller coaster.	Likert Scale
	I feel sick when I ride a carousel.	Likert Scale
	I feel sick when swinging on a swing.	Likert Scale

Table 3.1: Items of the CSSQ by category and measuring type.

3.3.2 Stimuli and Procedure

Before the experiment, both the CSSQ and SSQ were completed, followed by a short explanation about the imminent virtual roller coaster ride, and the collected data. The virtual roller coaster ride lasted 4 minutes and 55 seconds. During this time the GSR values were measured and recorded twice per second. After the ride was finished, the participant was asked to fill in the SSQ a second time. The Ipresence Group Questionnaire[147] was completed immediately afterwards.

3.4 Results

In this section the results of the experiment are summarized. All variables were tested for normal distribution and appropriate statistical tests were applied. The Spearmann correlation with a

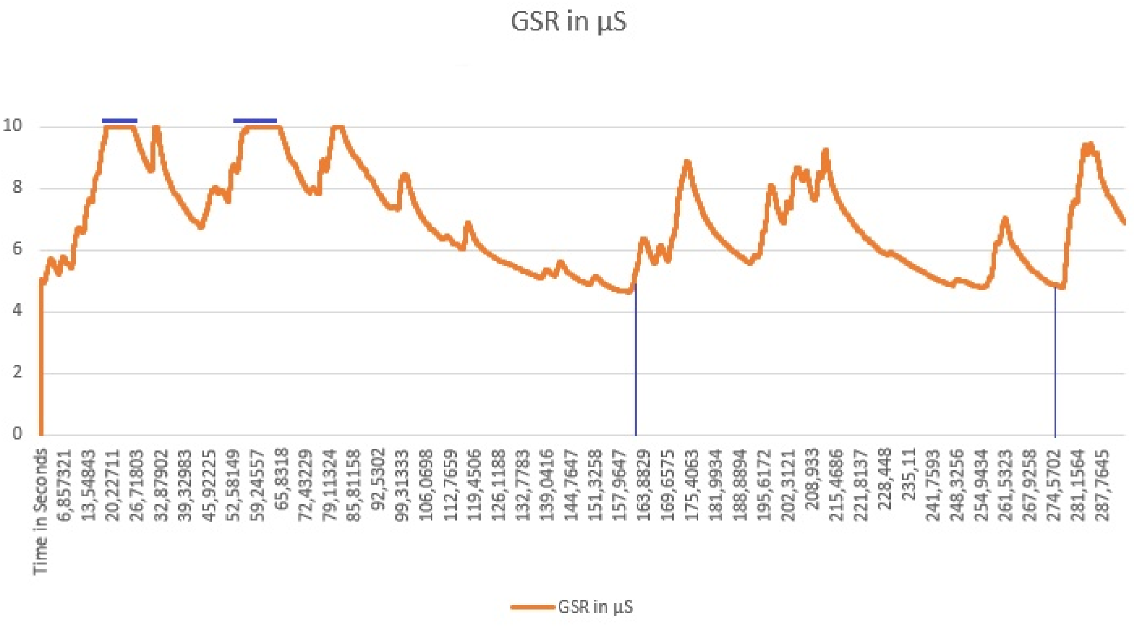


Figure 3.1: GSR values plotted over the course of a whole trial.

significance level of 5% was chosen to evaluate the correlations. To detect differences from a variable before and after the experiment, the Wilcoxon test was applied. Differences in the mean values of dependent variables were investigated using the Mann-Whitney U-test. Only significant results are reported.

For the difference between the *SSQPre* and the *SSQPost* the Wilcoxon test revealed significant differences between both the the total *SSQPre* and *SSQPost*, as well as its subscales. The *SSQPre* ($M = 13.71$, $SD = 12.51$) differed significantly from the *SSQPost* ($M = 28.83$, $SD = 18.79$, $p = 0.002$, $r(22) = 0.6375$), thus confirming hypothesis H_2 and confirming that cybersickness-related symptoms have occured. The following results could be found in detail:

A positive correlation between the *CSSQ-Score* ($M = 0.88$, $SD = 0.86$) and the *SSQ-Score* ($M = 15.11$, $SD = 20.14$, $p = 0.03$, $r(22) = 0.44$) was found, as was a strong positive correlation between the *frequency of stomach complaints* ($M = 0.79$, $SD = 0.721$) and the *SSQ-Score* ($M = 15.12$, $SD = 20.14$, $p = 0.001$, $r(22) = 0.67$). The analysis also showed that the triggers *riding a car* ($M = 0.71$, $SD = 0.99$, $r(22) = 0.49$), *riding a bus sitting backwards* ($M = 1.00$, $SD = 1.32$, $r(22) = =0.45$) and *swinging* ($M = 0.46$, $SD = 0.88$, $r(22) = 0.47$) had a positive correlation with the *SSQ-Score* ($M = 15.12$, $SD = 20.14$). Finally, a significant difference between the male participants' *SSQ-Post* score ($M = 4.42$, $SD = 16.37$) and the female participants' ($M = 24.17$, $SD = 18.98$, $p = 0.009$, $r(22) = 0.52$) was revealed.

We were interested in potential correlations between individual motion sickness triggers queried in the CSSQ and cybersickness occurrence during virtual emulation of these types of motion. Figure 3.1 depict example stress levels of an individual participant plotted over time measured by the GSR method. Here, the events *Passenger in a Driving Car* and *Boat Ride* measured the maximum value of 9,999 microsiemens, marked by the horizontal blue lines. Furthermore, several steep ascents occured, marked with vertical blue lines at the events *Flying* and *Swinging*. These indicate a strongly perceived cybersickness. The GSR data was smoothed using a low-pass Butterworth filter with the cutoff frequency at 1 Hz. However, no significant correlation between individual items of the CSSQ and the corresponding event in the VE were found in general. This indicates that the severity of cybersickness is indeed more closely related to the intensity of virtual motion in general, rather than specific types of virtual motions. This is conform with previous studies [50, 170] and shows that susceptibility to cybersickness affects the entire usage of VR, and not specific activities. Together with the SSQ results, this confirms that motion sickness of any kind is a strong indicator for general cybersickness susceptibility.

3.5 Discussion

The results revealed a positive correlation between the *CSSQ-Score* and the *SSQ-Score*, confirming a correlation between experienced motion sickness in real life and the perceived cybersickness. Hence, hypothesis H_1 was confirmed. Both the *SSQ-Post* and its categories *Nausea-Post*, *Oculomotor-Post* and *Disorientation-Post* had significantly higher mean values than the *SSQ-Pre* and its subcategories. This indicates that cybersickness has occurred, and real symptoms were reported, confirming hypothesis H_2. However, according to the recorded GSR values there is no direct link between susceptibility to a certain kind of motion sickness and similar virtual motions. From this we conclude that being susceptible to any kind of motion sickness indicates a general susceptibility to cybersickness, indifferent of the virtual activity.

The H_3 hypothesis about a correlation between health and cybersickness was also confirmed, as there was a strong positive correlation with the *frequency of stomach complaints* and a high *SSQ-Score*, indicating susceptibility to cybersickness. Further evidence of higher susceptibility may also be found in certain motion sickness trigger items from the CSSQ. Here, three moderately strong effects of correlations could be observed. Among these were the triggers *riding a car*, *riding a bus sitting backwards*, and *swinging*. This would mean that people who experience motion sickness in real life during these forms of motion are more likely to suffer from cybersickness than those who experience motion sickness from different triggers. These three triggers had stronger effects than others, confirming hypothesis H_4.

Hypothesis H_5 could be proven with a strong effect, as females reported significantly higher experienced cybersickness than males. This could also be seen strongly in the *SSQ-Score* mean values, with an SSQ-Score of 24 for females and about 4 for males. At the same time this result confirms hypothesis H_6, showing a connection between the demographic data of a person and the mean *SSQ-Score*. This result was already mentioned in previous works and was therefore to be expected [185]. There was no significant correlation between the sense of presence in the VE and perceived cybersickness in this study. Therefore H_7 was rejected. The CSSQ includes many more variables than those listed in the results. A large number of the yes-no questions were not considered further because they were answered exclusively with yes or no from all participants, making a statistical analysis futile. A larger sample size would be required to show a significant effect of specific factors. However, the chosen items have been individually shown by previous work to potentially influence susceptibility to cybersickness [159, 265, 273].

3.6 Conclusion

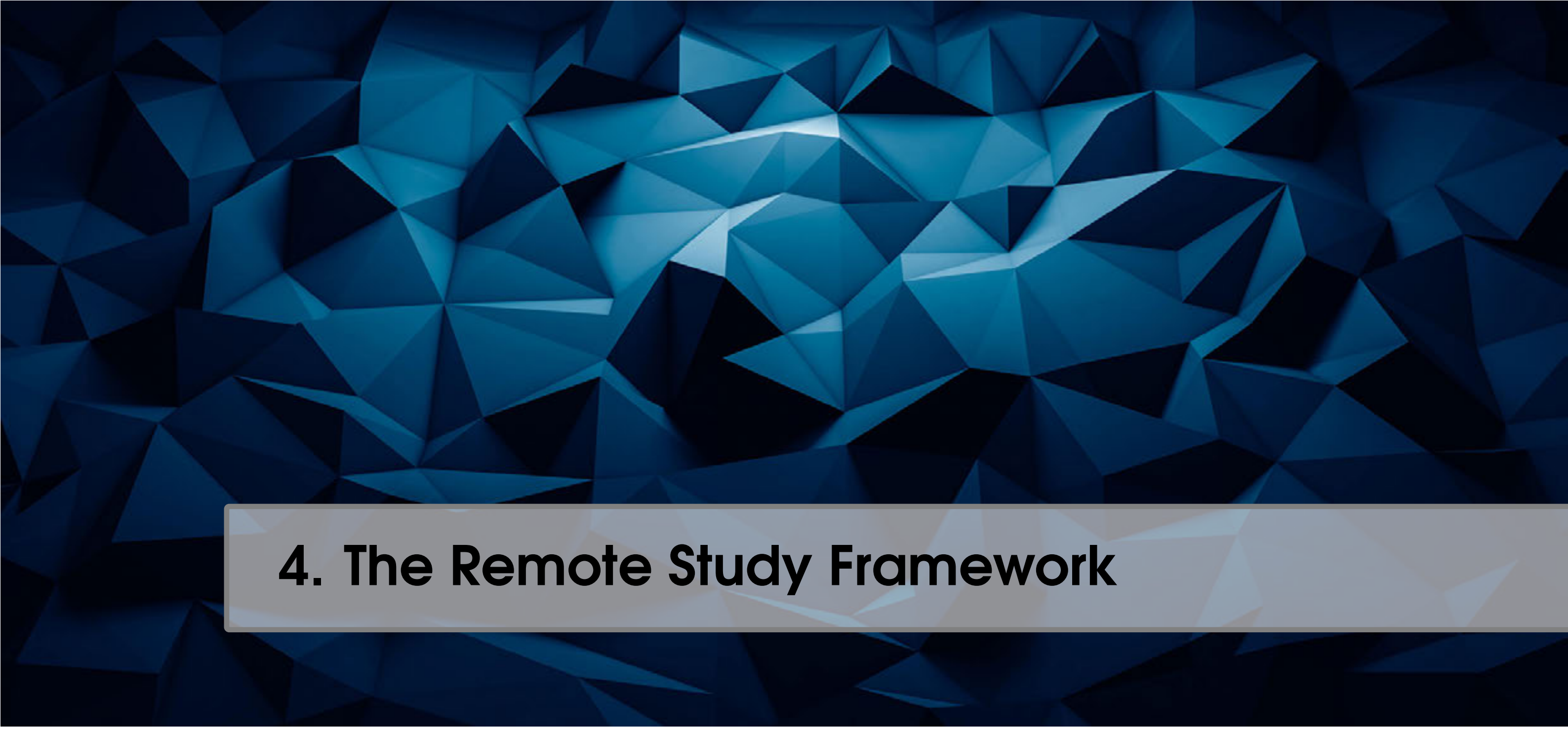
With our pilot study we confirmed which questions are most relevant for predicting cybersickness susceptibility. Especially the triggers *driving*, *riding a bus sitting backwards* and *swinging* correlated with a high *SSQ-Score*. *Female Gender* and *frequent stomach complaints* also showed a positive correlation with a high *SSQ-score*. These results suggest that people whose demographic data matches these characteristics or experience motion sickness from any of the above mentioned triggers are at a high risk of experiencing cybersickness.

From our results and the underlying previous work, a number of key factors can be isolated in regards to giving users a prediction for cybersickness susceptibility.

- Answering any yes-no item with "yes" indicates general susceptibility,
- Rating any Likert scale with at least a 1 indicates general susceptibility,
- Demographic data influenced the answers given on the CSSQ, and therefore certain age, gender and ethnic groups have a higher overall susceptibility,
- A higher total *CSSQ-Score* indicates stronger perceived cybersickness symptoms.

For future work we plan to iterate on the CSSQ with a weighting system for individual items based on our current findings, culminating in a scoring system that can give recommendations based on brackets of levels of risk. A second study with a larger pool of participants will be required to test and fine tune the scoring system, while also considering different VR scenarios.

The CSSQ is a useful generic tool for researchers performing MR user studies, for example to select participants based on the study's purpose. In the following section, a second generic tool is presented that allows researchers to perform MR user studies remotely.



4. The Remote Study Framework

4.1 The Remote Study Framework

Due to the holding health and hygiene regulations during the time of this work, it was partially necessary to perform user studies remotely. In contrast to in-person studies, this restriction brought several challenges. For instance, participants may have to be instructed before, during and after each trial or segment of an experiment. Additionally, an in-person study conductor can assist with configuring the hardware setup and application, as well as preparing all the necessary forms.

Furthermore, user studies usually require the usage of questionnaires. To this end, participants are typically asked to switch between an experimental computer and a dedicated questionnaire computer. This setup has a number of advantages. Experimental hardware like novel headsets or proprietary LUI solutions are prone to failure when switching applications, which is minimized by using a dedicated questionnaire computer. Furthermore, while the participant fills in the questionnaires, the next trial on the experimental computer can be configured by the conductor. Instructing each participant to set up the experiment on their personal computer or mobile XR device is difficult and tedious.

However, remote studies offer the possibility to invite a large number of participants, despite the aforementioned restrictions. Participants can be acquired through academic mailing lists, XR forums or learning platforms. Due to the rising popularity of XR, and especially mobile XR devices such as the Oculus Quest, the pool of potential participants constantly expands. Once participants have been gathered, the executable files and instructions can then be provided through file-sharing services like Dropbox or Google Drive. Additionally, hardware management services like ManageXR [152] offer the ability to install applications remotely on registered XR devices. In summary, providing the necessary files to participants requires little effort, but moderating the study process remotely remains challenging.

In order to overcome this challenge and accelerate the procedure of performing remote studies, a framework for the Unity engine was developed in the scope of this dissertation. The purpose of this framework is to specifically automate the following tasks:

(i) providing instructions and moderating the process during the experiment
(ii) answering common comprehension questions
(iii) switching between trials or experiment segments
(iv) randomizing the order of trials
(v) switching between the VE and questionnaires

This framework produces a singular executable, which has to be provided to the participant. As such, the file has to be made available to them with a message detailing how to install it on their personal computer or XR device, which can be done in the initial briefing. Every further instruction is then provided within that application. This application is built in the Unity engine, containing a singular scene. Within this scene, every Game Object required for the experiment is housed.

Providing Instructions In addition to the study dependent elements, this framework provides an adaptive message board. This is a 2D plane containing a header text box, a main text box, an image field and an indicator which button on the controller has to be pressed in order to advance to the next step. It is coupled to the XR user rig, constantly staying within the XR camera's FOV on the XZ-plane. This message board is used to give initial instructions before the experiment, and further instructions between experiment trials.

Frequently Asked Questions Additionally, a functionally identical message board can be displayed in the background of the scene. On this board, a reminder of the current task or answers to frequently asked questions can be displayed.

Switching Between Trials An experiment consists of many trials, and each trial of individual steps. For example, a trial may start with an instruction, followed by the task to move to a certain location within the VE, interacting with an object, and finally returning to the initial starting position. In this framework, each step of an experiment trial is represented as a singular *Study State*. The Study State defines which interactible objects are active, the message boards' properties and content, the XR rig's properties, and a condition for the Study State's completion. Study States are implemented as Scriptable Objects, a type of C# script that defines a data structure rather than a Game Object's component. Through a context menu, new Study States can be created within the Unity Editor, allowing the conductor to define each individual step of a trial. The above exemplary trial consists of four steps: instructions, reaching a position, performing an action, and again reaching a position. In this example, the first Study State defines that the message board is active and displays an instruction message, and that the Study State can be completed through a button press. Then, the following Study State hides the message board, enables movement through the VE, activates a location trigger, and is completed when the participant enters the location trigger. The third Study State is completed when the participant touches the object, and the fourth Study State is again completed when the participant enters a newly activated location trigger.

There are several choices for how a Study State can be completed. These completion conditions are implemented through an event-based architecture. Similar to Study States, events are represented as Scriptable Objects, making it easy for conductors to define custom events. Each Study State holds a reference to an event, and will be marked as completed once this event fires. Using the above example, an event is fired when a certain button is pressed, or when the XR rig enters a target destination. If this event matches the condition of the Study State, it is marked as completed.

In order to organize the experiment's individual steps, a central singleton manager is used, the *Study Manager*. The Study Manager holds the current Study State, and adjusts the properties of referenced Game Objects, such as the XR rig or the message boards, according to the settings held by the Study State. Most importantly, it holds the ordered list of all Study States. If an event fulfills the current Study State's completion condition, the Study Manager progresses to the next Study State, again adjusting the properties of the scene and its content accordingly.

Randomizing Trial Order The order of trials is typically randomized in experiments to maximize reliability and validity of statistical estimates of an independent variable's effect. This necessitates that Study States which belong to the same trial are grouped together, such that the order of trials can be randomized while the steps within each trial remain consistent. As such, a new Scriptable Object is introduced, called the *Study Sequence*, which represents a list of Study States. A Study Sequence can thus be used to model an encapsulated trial or a series of instructions on the message board. Similarly, the *Study Block* is introduced as a list of Study Sequences, representing segments of the overall experiment that should have a fixed order, such as introduction, tutorial, task, and ending. The order of Study Blocks and Study Sequences can be altered through a simple drag-and-drop metaphor, giving the conductor an addition useful tool to quickly test individual segments of the study, which can be difficult and time consuming with a strict order of segments. Through Study Blocks, the order of Study Sequences and thus trials can be randomized. However, it is insufficient to simply randomize the order of Study Sequences within a Block, as this obfuscates the relation to the external questionnaire results. To achieve this randomization, there must be several alternative Study Blocks that represent the same task, each with a unique and identifiable order of Study Sequences. When the order of trials is known, the questionnaire results can be allocated to individual trials. The conductor manually prepares a number of Study Blocks for the same task, each with a different order of the same Study Sequences in accordance with the Latin Square principle. The resulting study procedure can be thought of as a flow-chart between Study Blocks (cf. Figure 4.1). For example, the first Study Block is the introduction, followed directly by the Study Block for the tutorial. Then, the randomized task Study Block is loaded. Upon entering this Study Block, one of many alternative internal Study Blocks is chosen based on the participant's group ID. This ID is randomly assigned at the beginning of the Experiment. Finally, the experiment progresses to the ending Study Block.

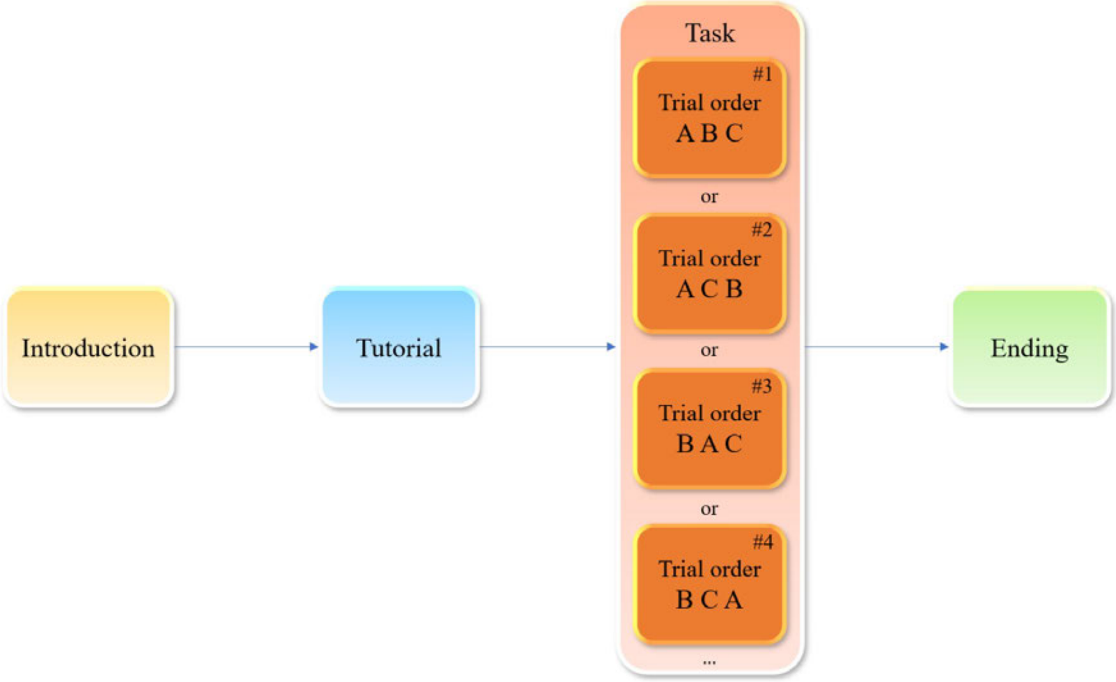


Figure 4.1: Illustration of the remote study framework as flow-chart. Colored elements represent Study Blocks, while the letters beneath the "Trial order" denote a Study Sequence. Each Sequence again is an ordered list of Study States.

Questionnaire Integration To circumvent the need for constantly switching between computers, a view to the questionnaires was directly integrated into the VE. This was implemented through the Vuplex WebView extension [326], a web browser integration for Unity. On the message board, this view can be enabled instead of displaying text. The initial website address can be preconfigured, for example to an anonymous survey on LimeSurvey [134]. On this survey, participants can click elements through a virtual laserpointer. The WebView is self-contained, signifying that the content of the shown website cannot trigger callbacks in the VE directly. However, WebView can detect if the title of the website changes. If the survey is configured in such a way that the website title reflects the current sub-questionnaire's name, it is possible to infer the stage of progression. Thus, if a button is pressed on this website that brings the participant to the next questionnaire page, an event is fired, and the Study Manager checks if the new website title is equal to the completion condition of the current Study State. If this is the case, then the WebView can be hidden and the next Study State in the Study Sequence is activated. This approach allows for an arbitrary number of individual questionnaire segments that correspond to a given trial. As the relation between group ID and alternative Study Block is deterministic, the conductor can reconstruct which questionnaire results belong to which trial and therefore the levels of the independent variables. This framework was used to conduct the user study in Chapter 10. [1]

---

[1] Due to licensing restrictions of the Vuplex WebView extension, the availability to this framework's code is currently limited to the Human-Computer Interaction work group of the Universität Hamburg.

III

Hybrid Collaborations

5. VR Invite

Before visualization and interaction techniques within shared MR spaces are discussed, we will address the integration of different types of devices in the context of hybrid user groups. While collaborations involving VR and OST HMDs have been demonstrated numerous times, e.g., in Microsoft's holoportation [236] and social platforms such as Meta's Horizon Worlds [195], two types of devices require further research in terms of multi-user MR integration, i.e., mobile devices and VST HMDs. This part is dedicated to integrating them with VR and OST AR devices in order to broaden the accessibility of shared MR spaces.

The user group of co-located people is not necessarily homogeneous and may employ a variety of technologies and devices to access shared MR spaces. As such, they are not necessarily occupied exclusively by HMD users. Shared MR spaces can further be used for visualizations and interactions where one or multiple bystanders either observe an immersed HMD user or simply desire individual viewpoints within a VE and on its content. Conveying this information to observers and non-immersive users is challenging, as is providing direct participation. Thus, the following RQ will be addressed in this chapter:

- RQ_1: *How can bystanders and users of mobile devices be integrated into shared MR experiences, and how should they be represented?*

5.1 Motivation

Current VR setups seldom are a collaborative or social experience for a group of co-located people. This is due to the requirement of VR being rendered on a single HMD, with a limited ability to include bystanders. Typically, only passive participation is offered by either mirroring the video feed or rendering the scene from a third person perspective to an external stationary monitor. To provide multi-user integration certain applications support the use of a number of HMDs connected via local area network (LAN) or internet. However, those are tailored to a specific use case and require substantial amounts of technical equipment and expertise to use in a local environment.

In single HMD setups exploration of the VE and social exchange with the HMD user are difficult for bystanders, as there is a need for input devices that are directly connected to the rendering computer, which in turn need to be provided in tandem with the VR setup. Such devices need to be tightly integrated and require an explicit implementation of camera control or other forms of interaction. However, the ubiquitous availability of smartphones allows for integrating bystanders into a social VR experience without the need for proprietary input devices with the

Figure 5.1: The VR Invite application allows a bystander to observe and interact with the VE from any perspective.

"bring your own device" metaphor.

In this chapter we introduce a project-independent smartphone application called *VR Invite*, which connects to a self-contained package for the Unity engine on a host computer or standalone VR headset via local wireless network. It allows bystanders to hold a viewport to the VE and observe the VR from any natural angle or position, as depicted in Figure 5.1. The package additionally supports transfer of touch inputs from the smartphone to the host computer. This enables an easy implementation of bystander interactivity for experiences that go beyond passive or verbal participation. Furthermore, we added visualizations of the additional viewports to increase the social presence of the HMD user. These visualizations provide a sense of spatial relation and participation between all users.

VR Invite can be used to extend existing projects with either active or passive participation capabilities, or build applications specifically tailored for an asymmetrical multi-user experience. To test the technical soundness of this library's first prototype, we conducted a user study in the scope of rehabilitation of older adults in a retirement home. Here, we measured the effect of the ability to freely move a handheld VR viewport on both an HMD user and a bystander, primarily regarding sense of presence, social presence and usability.

To summarize, the contributions of this chapter are:

- Development of VR Invite, a project-independent smartphone VR viewer application, and
- a user study to compare bystander integration techniques for the support role in an asymmetrical rehabilitation scenario.

5.2 Background

This chapter builds upon three areas of research: CVEs, asymmetric MR collaborations, and incorporation of bystanders into MR setups.

5.2.1 Collaborative Virtual Environments

In 1993 Carlsson et al. published DIVE, one of the first distributed interactive VR systems [57]. They focused on multi-user interactions in VEs and networking solutions for synchronization of virtual objects. CVEs now have a widespread use in MR applications, including rehabilitation, education, training, gaming and artistry [148, 160, 238, 313]. For example, Tsoupikova et al. used VR CVEs for rehabilitation in patients that suffered from a stroke [313]. They implemented motoric exercises for up to four patients with a focus on the social component of therapy. Kallioniemi et al. demonstrated how VR CVEs can be used to a help learning a foreign language [162]. In their CityCompass VR application, two users move through a virtual city and alternate between the roles of tourist and guide. To reach a common goal, they have to communicate in a foreign language.

Regarding co-located collaboration, Billinghurst et al. [29, 30] presented systems that let users perform a variety of interactions and visualizations in AR that are closer to face-to-face dialogue than screen-based collaborations. They predicted that AR properties such as seamless interaction between real world and VEs as well as superimposed spatial cues will enable science fiction-like teleconferencing when integrated in hybrid interfaces with other collaborative technologies. Jones et al. proposed RoomAlive [160], following the concepts of Billinghurst et al. with an implementation of projection-based co-located collaboration. They explored the transformation of a living room into an interactive playing environment for multiple users. The diverse use of CVEs is also evident in areas such as security training [238], medicine [59] or project planning in architecture [148].

Furthermore, numerous studies have investigated user experiences in CVEs along the above described main criteria [20, 58, 244]. For instace, Podkosova and Kaufmann [244] studied co-presence and motion patterns of VR CVE users. In particular, they compared the movement patterns of multiple users in a shared physical tracking space with those of users in individual tracking spaces. Their results indicate a stronger awareness of other users in shared tracking spaces, potentially due to the desire of collision avoidance. The effect is mitigated, however, if participants are far apart.

Technical advances in the field of web-based telepresence in CVEs have been reported for aspects such as network performance, user synchronization, and system security [129, 191]. Here, Gunkel et al. proposed a social MR setup based on WebRTC [129]. Their framework for 360 degree video transmission showed promising results with a high level of engagement and interaction between users utilizing current web technologies. Lee et al. presented a similar approach of transmitting the display of emotions through a 360 degree camera, but used a cloud service as infrastructure instead [191].

5.2.2 Asymmetric Mixed Reality Collaborations

VR Invite allows a number of bystanders to observe and interact with a VE, using smartphones rather than tracked hand controllers. Hence, their form of interaction is asymmetrical to the HMD wearing user. Several prior works have researched asymmetric interactions in VR or MR [70, 83, 90, 146, 148, 230]. For instance, Oliveira et al. presented a distributed asymmetric CVE for training in industrial scenarios, using screen-based GUIs to guide an HMD wearing user [232]. Their goal was to teach a trainee to operate and repair faulty hardware through remote avatars.

Oda et al. presented an asymmetric case study between a remote user and a local HMD wearing user [230]. They compared the use of traditional 2D interfaces and VR headsets as medium for the remote user regarding their ability to explain a certain task. The results indicate that demonstrating a task with a VR headset was easier to understand than annotations through 2D interfaces.

Lindley et al. compared gamepad-based input to tracked body movements for asymmetric avatar interactions. They found that natural body movements elicited a higher social interaction when compared to predefined animations [193].

5.2.3 Bystanders in Mixed Reality

Gugenheimer et al. proposed two approaches regarding the incorporation of bystanders in a VR setting. Their FaceDisplay [128] is an extension for conventional HMDs, equipping the headset with three external touch-sensitive displays and a depth camera. This way bystanders can see the VE from the outside and trigger actions by touching the displays on the head. This of course comes with the problem that an active HMD user moves their head frequently, which makes observing the displays and interacting with them difficult. An exploratory user study showed that the FaceDisplay led to a high degree of dominance and responsibility of the bystander over the HMD user.

With ShareVR [127] Gugenheimer et al. investigated an asymmetric gaming scenario, with one HMD user and one bystander. The bystander is equipped with a Vive Wand controller that has a tethered screen attached, acting as a second viewport. The scene is rendered from a second point of view on the VR computer and transmitted via standard HDMI cable. Also, a projector shows the VE from a top-down perspective on the floor. The setup showed an improvement in entertainment value, sense of presence and social interaction compared to the common TV-gamepad combination. The Master of Shapes commercial solution used a similar setup, but replaced the Vive Wand with a standalone Vive Tracker for positional tracking of a display [231].

Finally, Owlchemy Labs showed a prototype solution that uses a Smartphone for positional tracking and display of the viewport [181]. As long as the front of the headset is in the smartphone's view, their relative position can be calculated. The viewport is then rendered on the VR computer and transmitted via wireless LAN.

5.3 Implementation and Setup

Similar to the approaches of Gugenheimer et al. and Owlchemy Labs we designed a mobile viewport solution for desktop and standalone VR experiences. Like Owlchemy Labs we implemented positional tracking on a smartphone and transmit the data to the VR host computer, which then renders the image and streams it back via wireless LAN. Contrary to the aforementioned solutions, we used Google's ARCore library [119] to define a world anchor point, which could be any kind of picture or easily recognized object. For simplicity's sake we used a QR code printed on a piece of paper, detected by ARCore's "Plane Detection", "Augmented Images" and "Anchors" algorithms. The anchor can be placed anywhere in the physical room and is used to calculate the relative pose of a smartphone to the virtual origin point. Once the anchor is established by pointing the smartphone camera at it, ARCore builds and constantly refines a model of the physical room. The world anchor position can be defined in the engine's editor or by using the current position of an arbitrary tracked input device like a Vive Wand at runtime. This approach allows the smartphone user to look in any direction while we retain knowledge about the relative position and orientation to the VR scene's origin point, and transitively the VR player's position.

VR Invite was designed to be self-contained package for the Unity engine. When imported into an existing VR solution, the networking server doesn't need to be configured. At runtime, the server waits for smartphone application clients to connect to it, and then assigns them an in-engine camera object. The camera images are compressed to JPGs, serialized and sent via networking protocol to the client. Our current proof-of-concept prototype uses a simple TCP/IP implementation, which segments the image bytes into several chunks that can easily be reassembled on the client due to TCP/IP's guarantee of package order. Image resolution, compression and frame rate are parameters that can be adjusted depending on the number of clients. Figure 5.2 illustrates the networking scheme used in the current version of VR Invite.

The client within our smartphone application connects directly to a VR host computer by a given IP address and searches for a predefined anchor image. Upon identification of the world anchor point, the application streams its calculated relative pose to the anchor to the connected server. The camera's pose within the VR scene is now defined as the Vector from origin to world anchor position plus the vector from world anchor position to the smartphone, as determined by the application client.

This approach makes the combination of application and package project-independent, as the client only sends positional tracking data and receives a video stream in return, regardless of the scene's content or interactivity. It also allows a theoretically unlimited number of concurrent smartphone viewers, albeit at the cost of linearly increasing computational requirements for the host computer. The application also transmits touch inputs to the server, allowing optional implementation of project dependent interactivity. For example, touch positions can be interpreted as ray casts from the camera to highlight certain objects or to trigger events tied to virtual buttons or other interactive elements. To increase the VR player's social presence, an avatar should be displayed for each connected client. Obvious approaches are either displaying the image rendered for the viewport as floating plane, or using a 3D model positioned where the viewport is rendered from. For the following study, we chose to display a 3D model of a smartphone imitating the pose of the real device to relay a sense of spatial relation. An alternative method of displaying a video

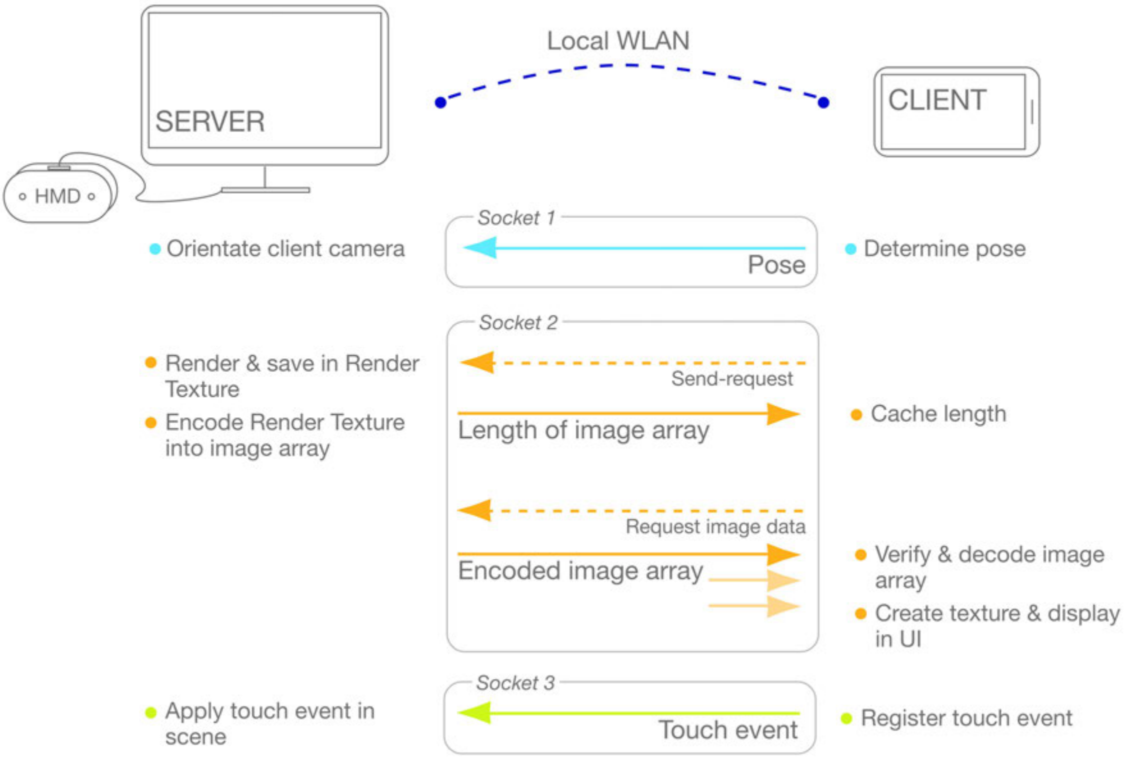


Figure 5.2: The networking scheme of VR Invite.

mirror from the smartphone's viewpoint will be discussed and evaluated in a separate study in Chapter 7.

The source code of both the VR Invite Unity package and smartphone application is available on GitHub [115].

5.4 User Study

In the context of rehabilitation for older adults in retirement homes, we conducted a user study where a VR player had to play multiple rounds of a memory game, while a bystander supported them. The VR exergame *Memory* of the EXGAVINE project [257] was chosen as a case study. This interdisciplinary project is concerned with the development and evaluation of medically and therapeutically effective VR movement games for the treatment of patients with neurological diseases. While VR Invite can be incorporated into any VR scenario, we focused on use cases which usually have one or multiple bystanders, as opposed to single user or explicit multi user scenarios. The use case of the rehabilitation project is a scenario that not only fulfills this requirement, but could potentially greatly benefit from active bystander integration. Here, bystanders are currently not integrated at all, or only through a passive monitor setup. We expected an increase in sense of co-presence for the older VR users, and thus a positive influence on their enjoyment and engagement with the rehabilitation program. For these reasons, we decided to use this application as a test bed for our case study. Their *Memory* game is meant to be played by older adults with verbal support from family members or nursing staff for training physical and mental capabilities. The term "exergame" is a portmanteau of "exercise" and "gaming" and describes fitness driven game designs. Our goal was to extend the exergame with VR Invite as an uncomplicated and intuitive form of interactivity to increase engagement between players and bystanders. Using the smartphone like a camera view finder was meant to be a concept that is easy to grasp without any knowledge of interactive video games or VR applications. Bobeth et al. [36] showed that older adults have a very positive attitude towards gesture-based interactions, supporting the choice of this VR exergame as a representative case study.

The focus of the study therefore was to determine if VR Invite is a suitable general purpose tool to observe and interact with a VR scene. To this end we compared VR Invite to a TV-gamepad-combination as input method for the support role regarding sense of presence, social presence, workload and usability, each with purely verbal or active assistance capabilities. In the chosen exergame the player is in a virtual park and has to solve a memory game with eight pairs of tiles (cf. Figure 5.3). Selecting a memory tile is done by throwing a virtual ball onto it, which simultaneously trains logical thinking and physical movement. We tested 2×2 combinations of independent variables: *Smartphone* versus *TV*, and *active* versus *passive* interactivity. Thus, each participant had to perform four trials. There was no focus on task performance due to the nature of this exergame's training and gradual self-improvement intent.

Passive TV represents the most common local VR scenario, where one or many bystanders can observe the real HMD user and have a TV screen or monitor showing the viewpoint of the VR user. *Active TV* adds input and a 3rd person perspective to the prior setup. Here, we gave the bystander a gamepad that could control the position and rotation of the TV screens's camera. Pressing any button while a memory tile is under a center crosshair visually highlights the tile in the VE.

The *Passive Smartphone* condition behaves the same as the *Passive TV*. It displays the VE from the HMD user's perspective on the smartphone screen, without the ability to move the camera or interact with the scene. Finally, *Active Smartphone* represents VR Invite. The smartphone can be used as a standalone viewport, which can be freely moved. Touching a memory tile on the screen visually highlights it in the VE, as depicted in Figure 5.3. The avatar of the HMD user always consisted of a simple head with black HMD and a representation of the controllers. Active viewports were represented with a smartphone 3D model in the VE.

For each condition the workload was measured by the NASA Task Load Index (NASA-TLX) [137], the usability by the System Usability Scale (SUS) [47], the presence by the SUSP [280] and the social presence by the NMSPI [32, 33]. The NASA-TLX questionnaire consists of six 20-point Likert scales, and is used to subjectively assess physical and mental workload. The SUS likewise is a ten-item attitude Likert scale, which is used to subjectively assess usability of systems and interfaces. The SUSP again uses Likert scales, with 6 items revolving around the sense of being in a VE and the extent to which the VE becomes the subjective dominant reality. Lastly, the NMSPI is a tool to assess the social and co-presence, as well as psycho-behavioral interactions between multiple study participants. We used the NMSPI version 1.2, consisting of 34 7-point Likert scales. In addition to the questionnaires, we measured the distance covered by the bystander in the VE as well as the duration of visual contact during active conditions as an indication of active participation and social interaction. Here, visual contact was determined by checking if the avatars are within the other participant's camera frustum.

Based on the above described criteria, the experiment was designed as within-subject, and the following hypotheses were formed:

H_1**:** Active conditions are rated higher in sense of presence by the bystander.
H_2**:** Active conditions are rated higher in social presence by the bystander.
H_3**:** Active conditions are rated higher in usability by the bystander.
H_4**:** The Smartphone causes a higher workload than the TV for the bystander.
H_5**:** There is no difference in sense of presence between all conditions for the HMD user.
H_6**:** Active conditions are rated higher in social presence by the HMD user.
H_7**:** There is no difference in usability between all conditions for the HMD user.
H_8**:** There is no difference in workload between all conditions for the HMD user.
H_9**:** The moved distance with active smartphones is greater than with gamepads.
H_{10}**:** The visual contact duration with active smartphones is greater than with gamepads.

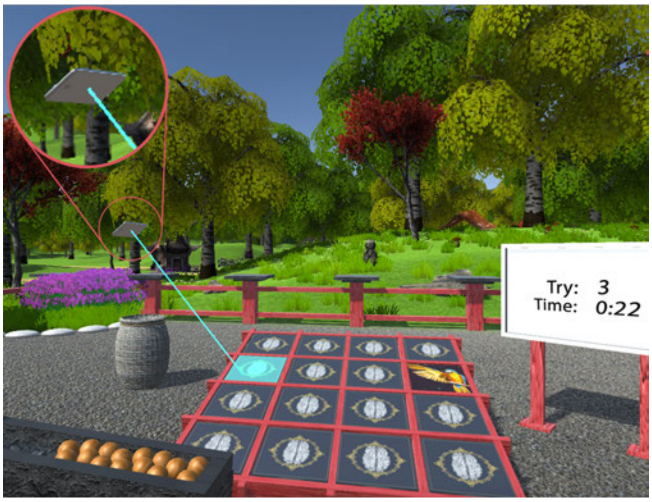


Figure 5.3: The VE of the VR memory game. A bystander with a smartphone highlights a tile.

5.4.1 Participants and Apparatus

26 participants (9 female, 1 diverse) took part in the experiment in pairs of two. In total, the study lasted about 90 minutes. Due to current health and hygiene regulations, the study had to be performed predominantly with students rather than the intended demographic of older adults, their relatives and nursing staff (cf. section 'Limitations'). The age range was between 19 - 62 years with an average age of 26.81 years ($SD = 10.08$, $M = 24.5$). All participants had prior experience with stereoscopic displays through VR headsets or 3D cinema, while 57,7% had prior experience with studies in VR. On a scale of 0 to 4, participants reported a mean 3D gaming experience of 2.54 ($SD = 1.10$) and their mean gaming time was 11.17 hours ($SD = 15.99$) per week. An HTC Vive Pro with Wand controllers was used for the VR player and depending on the condition a Google Pixel 3XL or XBox One controller for the bystander. A computer equipped with Windows 10, an Intel Core i7-4930K, an NVIDIA GeForce RTX 2080 Ti and 16 GB of RAM was used to render both the VR and the VR Invite viewports.

5.4.2 Stimuli and Procedure

After giving their informed consent and filling in a demographic questionnaire, participants were introduced to the memory game as well as their assigned device. They were instructed to cooperatively solve the memory challenge. As described above, each trial consisted of a full game of memory, where matching images needed to be turned over consecutively. To select a tile, the VR user had to throw a virtual ball at it. Not hitting two tiles with the same image consecutively resets the last two tiles. The trial ended when all tiles were flipped over. While the VR user selects the tiles, the bystander could assist them purely verbally during passive conditions and additionally by highlighting a single tile during active conditions. The trials were arranged via latin square, each taking circa 5 minutes. Following each trial, both participants filled out the set of questionnaires. Once all conditions were completed, the pairs switched roles and repeated the experiment.

5.4.3 Results

In this section the results of the statistical analysis are presented. When the Shapiro-Wilk test showed normal distribution of the samples, a repeated-measure ANOVA and post-hoc paired t tests were used to test for differences between conditions. Otherwise, the Friedman test and Wilcoxon Signed Rank test were used. A 5% significance level was assumed, and only significant results are reported.

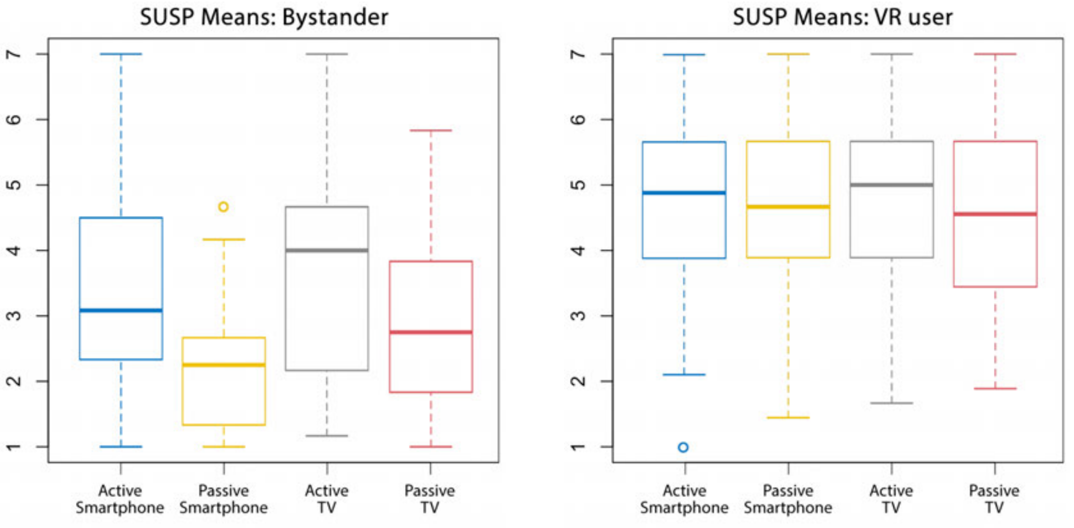


Figure 5.4: Mean presence scores for Bystander (left) and VR user (right). Higher is better.

5.4.3.1 Presence

For the sense of presence the SUSP score and the arithmetic mean of the SUSP results were considered. The wilcoxon test shows higher presence values for *Active TV* (score = 1.2) than the passive conditions, *Passive TV* (score = 0.3, p = .005) and *Passive Smartphone* (score = 0.2, p = .007). This is also evident in the arithmetic mean by paired t tests: *Active TV* ($M = 3.6$, $SD = 1.63$) is rated significantly better than *Passive TV* ($M = 2.8$, $SD = 1.33$, p = .003) and *Passive Smartphone* ($M = 2.3$, $SD = 1.04$, $p < .0001$). *Active Smartphone* ($M = 3.4$, $SD = 1.59$) is also higher than *Passive TV* (p = .007) and *Passive Smartphone* (p = 0003). In passive conditions, *Passive TV* performs significantly better than *Passive Smartphone* (p = .019). For the VR user role, no significant differences in sense of presence between the conditions could be observed either in the SUSP score or in the arithmetic mean. Figure 5.4 depicts the mean SUSP scores for both bystander and VR user.

5.4.3.2 Social Presence

We found significant differences in the social presence of the bystander. The result of *Passive TV* ($M = 3.2$, $SD = 0.78$) is significantly higher than that of *Active Smartphone* ($M = 2.8$, $SD = 0.82$, p = .0004) and *Active TV* ($M = 2.7$, $SD = 0.87$, p = .0001). *Passive Smartphone* ($M = 3.4$, $SD = 0.89$) shows a significantly higher result than *Active Smartphone* (p = .0002) and *Active TV* (p = .0001). For the VR user role, *Passive TV* ($M = 3.3$, $SD = 0.90$) is higher than *Active TV* ($M = 2.7$, $SD = 1.02$, p = .002) and *Active Smartphone* ($M = 2.9$, $SD = 0.75$, p = .022). *Passive Smartphone* ($M = 3.2$, $SD = 0.79$) was rated significantly higher than *Active TV* (p = .007). Figure 5.5 depicts the NMSPI scores for both bystander and VR user.

5.4.3.3 Visual Contact

We observed that some participants deliberately positioned themselves so that they always had the other participant in their viewport's frustum, while others focused entirely on the memory tiles. Because of this, the deviations are of considerable size. On average, bystanders looked at the VR users 3.3 times with an *Active Smartphone* ($SD = 3.58$, $min = 0.0$, $max = 14.0$), and 11.3 times ($SD = 7.04$, $min = 1.0$, $max = 28.0$) with an *Active TV*. The average total time looking at the VR user was 11.0 seconds with an *Active Smartphone* ($SD = 13.34$, median=7.5, $min = 0.0$, $max = 59.0$) and 82.7 seconds with the *Active TV* condition ($SD = 77.01$, median=58.5, $min = 2.0$, $max = 367.0$). Wilcoxon tests showed that both *time* ($p < .0001$) and *number* of eye contacts

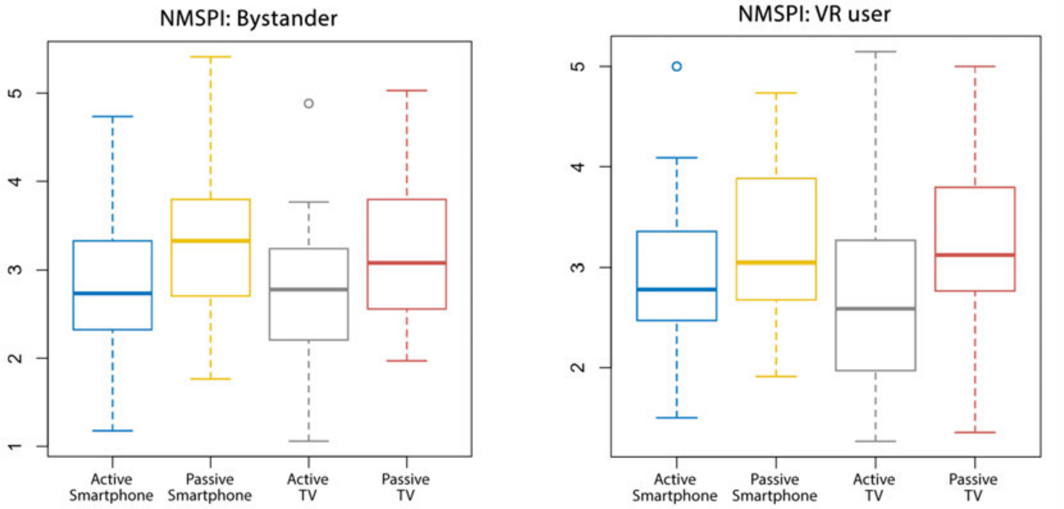


Figure 5.5: Social presence scores for Bystander (left) and VR user (right). Higher is better.

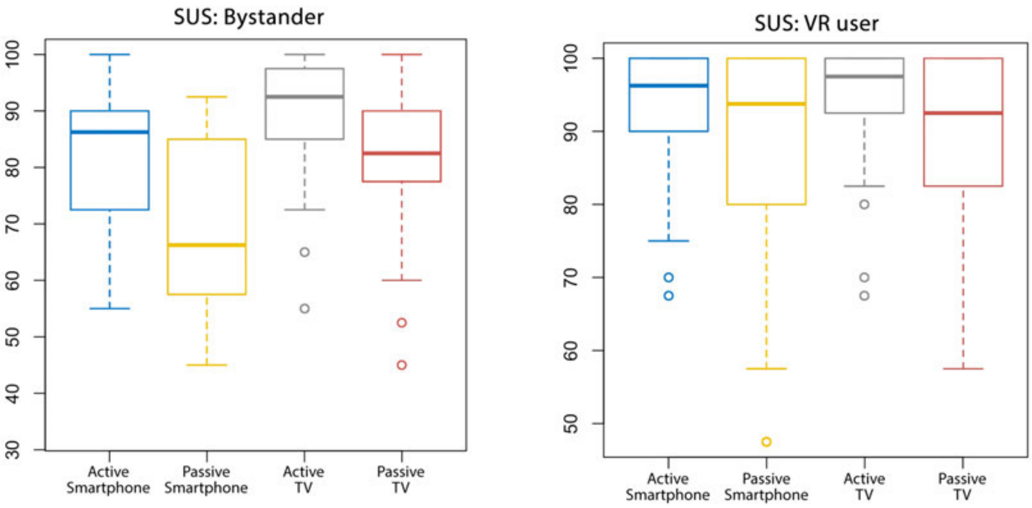


Figure 5.6: Usability scores for Bystander (left) and VR user (right). Higher is better.

($p < .0001$) are significantly higher for *TV* overall than for the *Smartphone*. On the other hand, the VR user looked at the Bystander an average of 11.5 times in the *Active Smartphone* condition ($SD = 10.09$, $min = 2.0$, $max = 39.0$) and 8.0 times when using the *Active TV* ($SD = 6.12$, $min = 0.0$, $max = 23.0$). The average total time looking at the bystander was 65.1 s during the *Active Smartphone* condition($SD = 63.38$, $min = 7.0$, $max = 309.0$) and 45.6 s when using the *Active TV* ($SD = 64.70$, $min = 0.0$, $max = 320.0$).

5.4.3.4 Usability

On the SUS, bystanders rated the usability of *Active TV* ($M = 89.5$, $SD = 11.66$) significantly better than all other conditions: *Passive TV* ($M = 81.3$, $SD = 13.75$, $p = .009$), *Active Smartphone* ($M = 81.0$, $SD = 14.95$, $p = .02$) and *Passive Smartphone* ($M = 68.9$, $SD = 15.05$, $p < .0001$). The usability value in the *Active Smartphone* is significantly higher than in the *Passive Smartphone* ($p = .0002$). The result of *Passive Smartphone* is significantly lower than that of *Passive TV* ($p = .003$). No significant differences in usability were found for the VR user role. Figure 5.6 illustrates the SUS score distribution for both bystander and VR user.

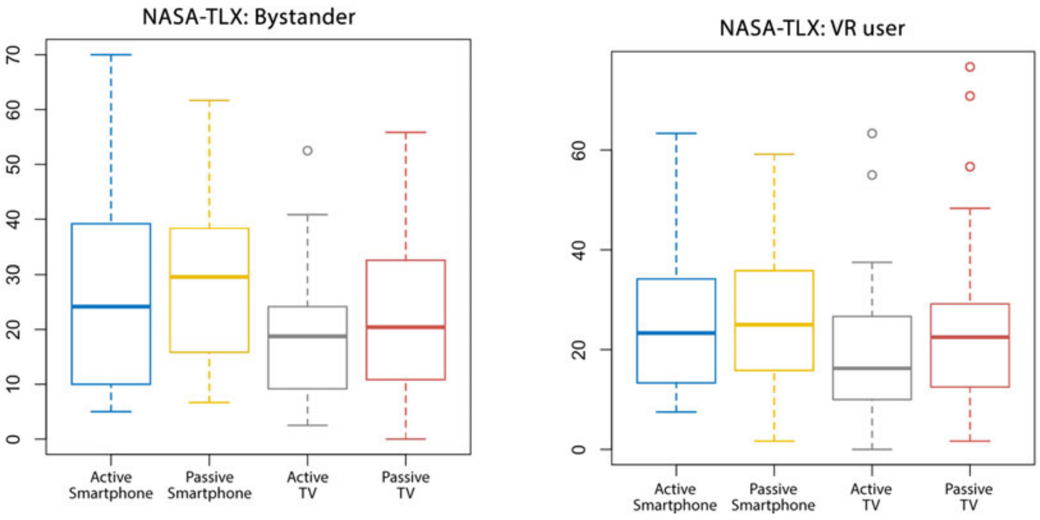


Figure 5.7: Workload scores for Bystander (left) and VR user (right). Lower is better.

5.4.3.5 Workload

We found a significantly higher workload for the bystander in the condition *Passive Smartphone* ($M = 29.5$, $SD = 14.50$) than in *Active TV* ($M = 19.2$, $SD = 11.56$, $p = .0006$) and *Passive TV* ($M = 21.4$, $SD = 14.23$, $p = .008$). *Active Smartphone* ($M = 25.9$, $SD = 16.07$) is significantly higher than *Active TV* ($p = .02$). There was no significant difference between the workload results for the VR user. Figure 5.7 show the NASA-TLX scores for both bystander and VR user.

5.4.3.6 Bystander travel

For the calculation of bystander travel two participant pairs had to be removed from the evaluation. Due short losses of ARCore's tracking, the measurements showed huge momentary spikes in covered distance. All other trials showed no tracking data inconsistencies.

A significantly longer distance was covered in the *Active TV* condition than in the *Active Smartphone* condition ($p < .0001$). The average distance in *Active Smartphone* is 15.9 m ($SD = 12.46$, $max = 64.1$) with a minimum value of 6.0 m. In *Active TV* this value is 92.1 m ($SD = 75.43$, $max = 368.0$) with a minimum value of 8.2 m. Only one participant used the provided tools for movement excessively, traveling 64.1 m for *Active Smartphone* and 368.0 m for*Active TV*. When excluding this outlier, the maximum recorded values become 35.5 m for *Active Smartphone* and 177.1 m for *Active TV*. This however has no impact on the significance of the difference between the conditions.

5.5 Limitations

As described above, relatives and nursing staff were supposed to use the VR Invite mobile application to interact with elderly participants, who are immersed in a VR rehabilitation training environment. However, due to current health and hygiene regulations, it was only partially possible to perform the study with representative participants and only single bystanders. We therefore had to predominantly rely on university students for both roles of bystander and VR user, which will undoubtedly have skewed our collected data in regards to gaming experience and familiarity with input devices like gamepad controllers. Yet, there is an overlap of the tested and targeted demographic for the bystander role, which would have included family members and nursing staff. Nevertheless, the collected data shows strong significances which cannot solely be attributed to the selection of test subjects. VR Invite itself is a general purpose VR cooperation tool, which can be used in a variety of scenarios. Therefore, we believe that the general trends we found are representative for a wide range of use cases, but not necessarily for the intended elderly VR users of the EXGAVINE project.

The version of VR Invite that was used during the experiment is an in-development prototype and is not yet ready to be released. While the tracking was consistent throughout the study, we received feedback that the frame rate on the smartphone was not as smooth as the participants had liked it to be. We attribute this to the naive serialization and network transportation implementation, which we chose to use for our proof of concept prototype. While the tested solution was not optimized, we deemed it good enough for initial testing and thus conducted the experiment. The performance of the image transfer also had no impact on the rendering for the VR headset, as the transfer is handled in a separate thread. To make VR Invite production ready, the networking solution will have to be exchanged for a more sophisticated image compression and transfer stack. This in turn could possibly have a positive impact on the user ratings for usability and possibly sense of presence as well. We don't however expect the reported data to drastically shift with improved frame rates, as the results are quite unambiguous.

5.6 Discussion

Sense of presence The results of the study highlight the connection between active participation in a collaborative experience and an increase in the self-reported sense of presence. When looking at the arithmetic mean, both *Active Smartphone* and *Active TV* received favorable ratings when compared to their passive counterparts. As expected, the ability to independently move and interact with the VE is a major factor in feeling involved in a CVE, confirming hypothesis H_1 (Active conditions are rated higher in sense of presence by the bystander). However, the overall average sense of presence was at a relatively low value of 3.025 out of 7. This can be explained by the fact that the bystanders perceived the CE only via screens and always had the real environment in their field of vision. As mentioned in the limitations section, technical improvements of VR Invite could further shift the bystanders sense of presence in the active conditions favor. We found no significant difference between the conditions for the VR user's sense of presence, which confirms hypothesis H_5 (There is no difference in sense of presence between all conditions for the HMD user). The displayed camera model was perceived equally as present when it was in a fixed location as to it following the bystanders real position. This is most likely due to the fact that bystanders had a tendency to remain standing in the same position for the majority of the trials. They reported an initial wow-effect and tried to move the viewport around the tracking space. Once they found a position that was out of the VR user's range from which they could observe the memory tiles, they remained stationary. Because of that, there was no significant difference between the camera model's movement from the VR user's perspective.

Social Presence Contrary to the hypotheses H_2 (Active conditions are rated higher in social presence by the bystander) and H_6 (Active conditions are rated higher in social presence by the HMD user), the sense of social presence was rated significantly higher by both bystander and VR user in the passive conditions. This might seem counter intuitive at first, but can be explained by two factors. One, the bystander's attention is primarily focused on the VR user in the real world. Instead of looking at a smartphone or TV screen, and thus an avatar, bystanders tended to spend a longer time observing the real human being. Two, when not able to directly interact with the VE, bystanders had to go through the VR user as an intermediary by verbally communicating with them. In the active conditions the pointing interaction was used frequently and reported as joyful. However, the pairs spoke noticeably less with each other. This leads us to believe that verbal communication is a more important factor for sense of social presence than visual communication through cues and individual agency. Thus, autonomy in bystander integration appears to negatively correlate with social presence. Prior studies have found that current VR meetings do not reach the same sense of social presence as a real in-person meeting, which can be applied here as well [144].

Visual Contact Regarding visual contact, we found that the *Active Smartphone* was used significantly more often to interact with the memory tiles, while the *Active TV* condition showed more visual contacts with the VR user, disconfirming hypothesis H_{10} (The visual contact duration with active smartphones is greater than with gamepads). The *Active Smartphone* incited the bystander to behave autonomous and solve the memory game on their own, and in turn giving more visual cues to the VR user than the *Active TV*. In a sense, *Active Smartphone* users were more engaged with the experience, but less so on a social level. On the other hand, the VR user looked at the Bystander more often and for longer periods of time during the *Active Smartphone* condition, confirming they are not within their reach to accidentally hit them with the Vive Wands (11.5 times versus 8.0 times and 65.1 seconds versus 45.6 seconds). Thus, VR users were actively aware of the bystander's location.

We also made the following noteworthy observations regarding visual contact and collected demographic data. Male VR users held visual contact with the *Smartphone* significantly longer than female users. This behaviour was inverted during the *TV* conditions. Overall, while the bystanders used a *Smartphone* as opposed to a controller, they looked at the VR user less often, but the VR user looked at them more often. Vice versa for trials where the bystander used a controller. There was a negative correlation between hours played per week and duration of visual contacts during the *Active Smartphone* condition. A positive correlation was found between age and duration of visual contacts during the *Active Smartphone* condition.

Usability Bystanders rated the usability higher in the active conditions. *Active TV* achieved the highest rating, followed by *Active Smartphone*. This confirms hypothesis H_3 (Active conditions are rated higher in usability by the bystander). From participant comments we deduced that control over the viewport and the interaction made the game easier and more enjoyable, which aligns with the findings of Gugenheimer et al. [127]. The mean values for *Active TV*, *Active Smartphone* and *Passive TV* are above 80 points, a result close to or exceeding the "Excellent" usability category according to the SUS evaluation guidelines of Bangor et al. [24]. *Passive Smartphone* was rated lowest with 68.9 points, which Bangor et al. classify as borderline between "OK" and "Good" usability on the adjective rating scale.

As mentioned above, the technical shortcomings of our implementation could have had an impact on the usability rating. We don't however expect changes to the networking stack to overcome the gap in ratings that presented itself. Not having to actively move within the real space or holding up a smartphone was rated significantly more usable than the opportunity for natural viewport manipulation and freedom of movement. The convenience of sitting down outweighed the increase in precision and naturalness of movement.

We found no significant difference for the reported usability of the VR user between the conditions, confirming hypothesis H_7 (There is no difference in usability between all conditions for the HMD user). The visual cues did neither increase or decrease the usability of the entire setup or the memory game as a whole. Looking at the mean values, the active conditions at over 90 points are slightly higher than the passive ones at an average of 88.55 points. All values indicate a satisfactory usability. In the qualitative feedback, especially the visual cues and the avatars were found to be helpful in cooperating with the bystander. Overall this did however not alter the way the VR user interacts with the VE.

Workload As expected, we also could not find a significant difference for the VR user's workload. While additional visual guidance from the bystanders was perceived as helpful, it also had to be mentally processed and combined with verbal communication and their own memory of the tiles. These two factors effectively canceled each other out, while there was no impact on the physical aspect of the game. This confirms hypothesis H_8 (There is no difference in workload between all conditions for the HMD user). On the Bystanders' side, the workload in the *Smartphone* conditions is reported higher than for the *TV* conditions. This was to be expected; holding up a phone screen is more strenuous than sitting down with a controller. Similarly, actively walking through the real world with a screen in hand less comfortable than sitting down and using the joystick movement. This confirms hypothesis H_4 (The Smartphone causes a higher workload than the TV for the bystander.). Interestingly, bystanders reported a higher workload for holding the *Passive Smartphone* to their face than actively participating with an *Active Smartphone*. It is possible that the smaller screen size compared to the TV resulted in a higher cognitive and physical challenge. However, the average workload values do not exceed 29.5 points in any condition and are thus in a similar range to the values of the VR users.

Travel Distance A significantly longer distance was covered in the *Active TV* condition than with an *Active Smartphone*, with a factor of 5.8 (92.1 m versus 15.9 m). We expected participants to make use of the freedom of movement, which they did not. This disconfirms hypothesis H_9 (The moved distance with active smartphones is greater than with gamepads). Only one participant used the provided tools for movement excessively, traveling 64.1 m with the *Active Smartphone* and 368.0 m with the *Active TV*. Bystanders tried to avoid the movement radius of the VR user, and were content with a position that let them comfortably see the entire memory board on their screen.

We found several correlations between the traveled distance and demographic data. Participants that were not familiar with gamepad controls generally moved less than those experienced with gamepads, and rated VR Invite more positively overall. There was a negative correlation between *Active TV* and age. Older participants moved significantly less. There was a positive correlation between *Active TV* and hours of playtime per week. More hours of playtime led to a significant increase in travel.

In short, there is an influence of 3D gaming experience on virtual scene exploration. There was no such effect for the *Active Smartphone*. This indicates that VR Invite is an intuitive tool that can be used independently of experience with virtual scenes. The familiarity with smartphones in general and the metaphor of the portable window led to a quick adoption of the technique.

Convenience Over Agency After the study participants were asked to indicate which input method they preferred overall. Here, *Active TV* was chosen 18 times and *Active Smartphone* 8 times. Participants had an initial "wow-effect" with freedom of motion through VR Invite, but ultimately valued the comfort of sitting down with a controller in hand over room-scale movement and user agency. Screen size and having to hold up the phone over long periods of time were quoted as reasons for preferring the TV condition. It appears that there has to be a balance between convenience and interactivity, where avoidable movement is seen as a cost. The payoff or the incentive for movement seemingly has to be disproportionately big to outweigh the loss of convenience when there is an alternative form of interaction. To summarize, convenience predominates interactivity and agency if the payoff is not disproportionately big.

5.7 Conclusion

We proposed and evaluated a prototype of a project-independent smartphone viewer app, which enables bystanders to explore and interact with PC or standalone VR applications with unexpected and interesting results. In the utilized memory VR exergame, the bystanders in particular reported a higher sense of presence and higher usability of the active conditions compared to the passive ones. However, the self-reported social presence was significantly lower during active than in passive conditions. Bystanders showed a higher independence and agency when using VR Invite, focusing on the task at hand rather than the VR user. However, agency appeared to negatively correlate with social presence in the explored bystander scenario. This indicates that VR Invite is best suited for implementations with active participation rather than pure observation, where a TV screen was preferred thanks to the convenience it provided. Overall, participants preferred the *TV* over VR Invite for single bystander scenarios, quoting convenience as main driving factor. This demonstrates that while user agency and interactivity are welcome and improve the experience over passive participation, convenience cannot be understated as a central factor of interaction paradigms.

VR Invite was rated to have satisfactory usability, and showed promising results that highlight the potential of individual viewports in use cases with multiple bystanders, where a single TV screen is not sufficient for multi user interactivity. The active component of the *TV* condition can also not be applied to multiple users, as the screen has either to be split or only one bystander can have control. Contrary, impromptu social sessions can make use of VR Invite's ability to quickly join an experience, providing multiple viewing angles. TV and VR Invite can of course be combined, with a neutral perspective on the TV and individual viewports for each bystander for personal agency and interactivity. Regarding agency, Gugenheimer et al. reported that mobile systems could significantly increase the enjoyment of a collaborative game [127]. Their findings indicate that VR Invite would be best suited for deliberately designed multi-user applications, making use of the provided touch interactivity.

Our next goal is the optimization of VR Invite with regards to image serialization and networking stack to make it more pleasant and versatile to use. This in turn will enable us to perform further studies with multiple bystanders to investigate social presence in groups of actively supporting users and one or multiple VR users. Furthermore, we plan to extend VR Invite to support AR and MR. In XR mode, the host computer or standalone XR device should only render the virtual objects of interest on a transparent background instead of the entire VE. The smartphone passes through its built-in camera feed to its display, and layers the video stream received from the rendering host on top. Implementing and testing these capabilities remains as a target for future work.

6. Camera Time Warp

As discussed in previous chapters, shared MR spaces can be accessed through a variety of devices from the entire reality-virtuality continuum. Contemporary mobile devices offer VST AR, albeit in a handheld form factor. This principle can likewise be applied to the VST HMD, which allows for stereoscopic rendering of virtual objects in the real-world environment, e.g., the avatars of other users. Compared to real humans within one's FOV, virtual human representations are easy to identify in VST HMDs, even when using high resolution human 3D scans. Their position is not stable within the visible real world reference frame due to the camera-to-photon latency, which is a detriment to both the technical immersion and psychological sense of presence. Hence, the following RQ will be addressed:

- *RQ_2: Can the registration error in VST HMDs be decreased and will this have a positive effect on the emergence of cybersickness-related symptoms?*

6.1 Motivation

VST HMDs have benefits over OST HMDs such as a wider FOV and the ability to render fully opaque virtual objects to cover the whole reality-virtuality continuum [209, 210, 217, 235, 288]. As AR technology further matures, users will increasingly spend more time in AR or MR environments. However, when users undergo longer sessions in VST AR, there is an increased risk for cybersickness because of the inherent latency of the video feed [180, 217]. Depending on the camera, real world images arrive with a delay of around 30 - 50 ms in the 3D engine. Then, the frame needs to be rendered (e. g. 11 ms at 90 Hz) and transferred to the HMD's screen (10 - 20 ms) [80]. Therefore, the video feed has a camera-to-photon latency between 50 and 80 ms in total. This delay and the resulting registration error has a considerable risk to cause cybersickness-related symptoms including eye strain, headaches and nausea as well as degraded visuo-motor interaction performance [22, 188, 204, 341]. In contrast, virtual objects only require rendering and transfer to the screen. This discrepancy becomes critical during head movements, since the delay of the camera images will not match the delay of the virtual objects. The virtual objects, in essence, become reference points, thus emphasizing the latency of the video feed in the background. This phenomenon is akin to display latency, which, according to Kruijff et al. [180], is one of the fundamental perceptual issues of AR. Although this problem has been previously identified by Nelson et al. [217] and Yokokohji et. al. [341], to our knowledge, no solution has sufficiently addressed the three main issues: compensating the camera-to-photon latency, stabilizing the images during mismatches of camera frame rate and HMD refresh rate, and masking artifacts during fast head movements [80, 234]. To address these challenges, we propose Camera Time Warp (CamWarp), a technique to mitigate the effects of latency on cybersickness. CamWarp's core idea is to induce compensatory latency to the HMD's tracking system to match the latency of the video stream.

To summarize, the contributions of this chapter are:

- Introduction of the novel CamWarp technique to mitigate the effects of latency on cybersickness-related symptoms in VST AR,
- A user study to confirm the effectiveness of the CamWarp technique regarding subjective levels of discomfort,
- A user study to explore the transfer of the CamWarp technique to scenarios which require efficient hand-eye coordination.

6.2 Background

As discussed in Section 2.7, LaViola [188] examined the applicability of the sensory conflict theory [312] to the emergence of cybersickness in VR. This conflict also applies when the latency of the video feed causes mismatches between the user's vestibular system and the view in the headset; the head stops after the desired motion, but the delayed video stream suggests ongoing movement. Nelson et al. conducted an experiment in which seven participants had to perform a head tracking task with and without added delay to the images [217]. Their results indicated that time spent in AR was the deciding factor for cybersickness symptoms, more than changes in camera delay.

Van Waveren et al. described a technique called *Asynchronous Time Warp* [322]. It applies a 2D translation to the final rendered image that incorporates the latest head movements of the user that have been measured after the rendering of a frame finished. When ran asynchronous to the refresh rate of the HMD, the perceived frame rate was increased by essentially interpolating frames. Since this is a general solution for HMDs, it has no effect on the latency discrepancy of rendered objects and the AR video stream.

Kim et al. [166] proposed *DotWarp*, a technique that relies on an additional high frame rate IR camera. By blending the stereo camera images with those of the IR camera, they achieved an increase in perceived frame rate and reduced the perceived registration error. Nevertheless the IR camera is still affected by latency.

Yokokohji et. al [341] implemented a predictive approach where the imminent movement of the user is estimated from IMUs on the HMD. They shifted the virtual objects based on this information. Nevertheless predictive algorithms are not as consistent and reliable as non-predictive approaches. Differences in camera frame rate and HMD refresh rate were not considered.

Bajura & Neumann [22] proposed a technique to reduce registration errors in VST AR setups by inducing additional latency to specific parts of the setup. Their premise was that at the time, camera technology had a lower latency than tracking technology. They buffered the video stream's images to match the latency of both technologies, resulting in a reduction of registration errors. However, this comes with several drawbacks, such as a decrease in dexterity and hand-eye coordination. Currently, the latency of tracking systems has decreased to levels below that of stereo cameras. This means that Bajura's technique cannot be applied to modern VST AR hardware.

6.3 Technique Description

CamWarp likewise is a reprojection technique for VST AR that reduces the registration error between AR video streams and virtual objects by matching the camera and tracking latency. We compensate the camera-to-photon latency of the physical cameras by inducing a delay to the tracking system, as opposed to buffering the video stream. This is the opposite of what Bajura [22] proposed. Our experiments in sections 6.4 and 6.5 demonstrate that CamWarp comes with all of Bajura's technique's merits but none of the drawbacks.

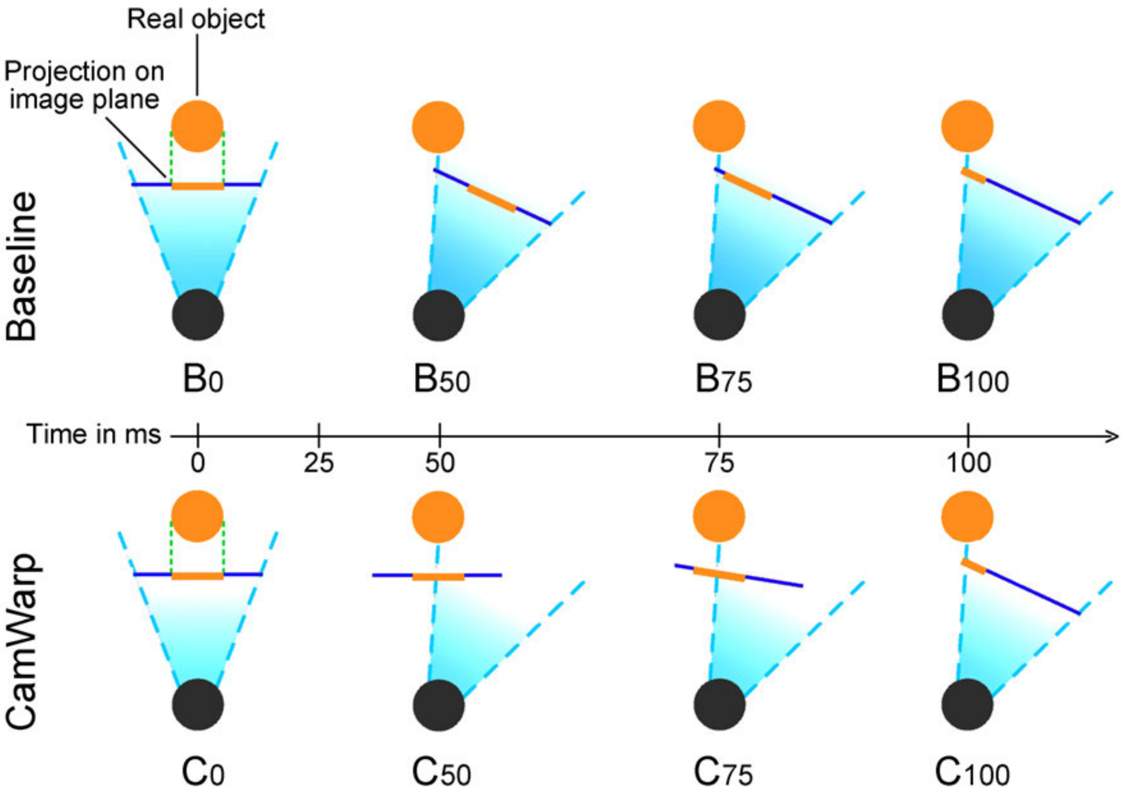


Figure 6.1: Top row: The baseline procedure in which the projection of a real object on the IP (blue line) gets off the line of sight from the user's head (**black circle**) to the real object during head movements. Bottom row: CamWarp, in which the IP is positioned where it has been recorded 50 ms before, ensuring that the real object and its projection remain on the same line.

6.3.1 Formalization of the Problem

Latency in the context of VST AR is defined as the delay between recording and display of visual information. Assuming a camera-to-photon latency of 50 ms and an HMD resolution of 1000 x 1000 pixels over a 100° FOV, we can calculate that the effective displacement of recorded objects for a head rotation speed of 60°/s results in an error of 3° in 50 ms. Thus, the displacement of recorded objects is 30 pixels, which is wider than the thumb held at arm's length.

Figure 6.1 illustrates this behaviour for a static scene that contains a real object within the user's view frustum. The object's projection on the image plane (IP) leaves the line of sight from the user's head to the real object during head movements due to latency. It will require twice the camera-to-photon latency to return onto the line. The result is a perceived movement of the object followed by back snapping to its real position. This is a discrepancy between expected and experienced visual information, which leads to discomfort and in turn can cause cybersickness [188].

6.3.2 Camera Time Warp

To solve this effect of apparently moving and back snapping objects, we display the information where it has been recorded. This is achieved by inducing a 50 ms latency to the movement of the IP. During the 50 ms of rotation between C_0 and C_{50} in Figure 6.1 the IP remains static. From the user's perspective, the real and the projected object still align. At C_{75} the IP has followed half of the user's movement from C_0 and C_{50} and is positioned according to data from C_{25}. Meanwhile, the projected object on the IP shifted half way from the center to its end position. In C_{100} both the IP and the object's projection have caught up with the user's movement.

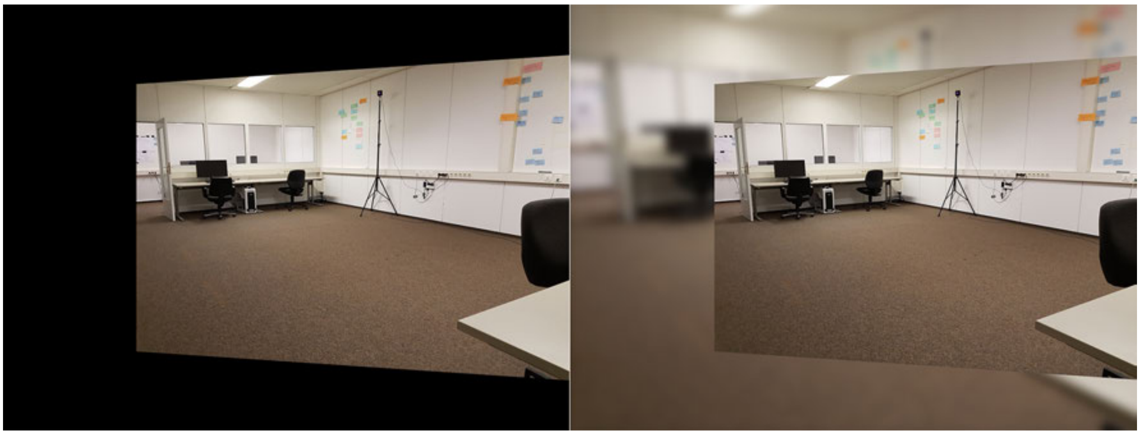


Figure 6.2: A frame without and with the blurred background. The FOV is wider than usual to illustrate the effect.

We also consider the update rate or *frames per second* (FPS) value of the stereo camera. If the FPS are lower than the refresh rate of the HMD's display, the IP would be falsely moved in between camera frame cycles [80]. We only update the pose of the IP with delayed tracking data when a new frame is received from the camera driver.

6.3.3 Implementation

To integrate CamWarp in an AR scene, first, we need to buffer the pose of the IP for each eye. The buffer length is defined by the latency from the camera to the 3D engine and the HMD's refresh rate. In this example, 30 ms and 90Hz. We divide the latency with the time it takes to render one frame ($\frac{1}{90} = 11$ ms) to obtain the number of frames that need to be buffered. In this example, we have to buffer $\frac{30ms}{11ms} = 2.7$, rounded to 3 frames. One possible implementation of this buffer is a queue of up to 3 elements [234]. Second, not every rendered frame should move the IP. The buffer needs to be read only if the last update from the camera driver happened $\frac{1}{FPS}$ seconds or more ago [80]. By using a queue for the pose of the IP we effectively delay the input from the tracking of the HMD, and subsequently match the latency of tracking and video feed.

During fast head motions, black borders representing pixels without video information are displayed in the periphery due to the delayed streaking of the IP. To mitigate this effect, different masking techniques can be applied. We chose to display a copy of the IP centered in the view frustum, and blur it. This approach was inspired by the established practice used to display videos in landscape mode that were shot in portrait mode to mask the black edges on the sides. Studies in VR also showed that blurring the peripheral vision can mitigate cybersickness effects during head motions [219]. The computational overhead is minimal as we simply use a lower resolution version of our current video frame, as depicted in Figure 6.2.

Figure 6.3 illustrates the processing of a head motion over several frames. By buffering the image plane motion we synchronize its processing with that of the video stream, giving us the desired optical effect that is shown in Figure 6.1. Old video data is now displayed at the position in 3D space where it has been recorded 3 frames before, resulting in a stabilization of the user's vision. This compensates the stereo camera to engine latency. The remaining latency of transporting the rendered image from the engine to the headset is then handled by the headsets driver with established techniques such as Asynchronous Time Warp (ATW). ATW is an algorithm for HMD's that increases the perceived frame rate through interpolation by incorporating tracking data that was not yet available during frame buffering [322]. As a general solution for VR headsets it addresses a different source of cybersickness and thus complements CamWarp. We utilized them in conjunction for both our experiments.

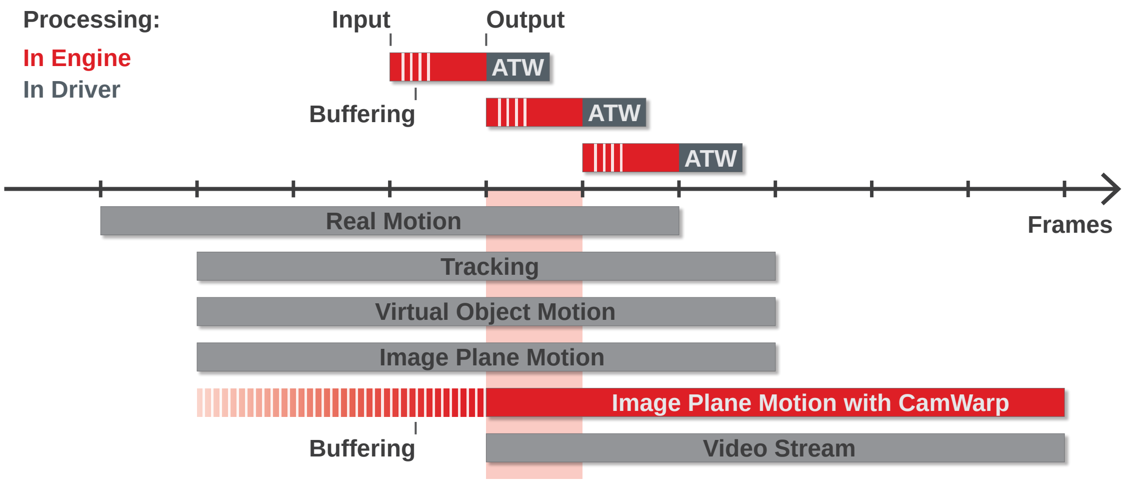


Figure 6.3: Timeline of the processing of a motion. The top shows the processing of single frames, while the bottom illustrates the latency discrepancy of several components.

6.4 Cybersickness Experiment

We conducted an experiment to confirm the effectiveness of our approach. We used ATW as baseline and compared it to ATW plus CamWarp with regards to subjective discomfort, similar to Fernandes and Feiner [101]. In the experiment, participants had to move their head in a Fitts' Law/ISO 9241-9-inspired pattern [102] at different speeds, while wearing an HMD with a stereo camera extension.

Our hypothesis was that CamWarp will have a stronger effect on low FPS values, as we expected an increased registration error without latency compensation when individual frames are displayed for a long time. Verifying this hypothesis required low baseline FPS, the maximum FPS that match the refresh rate of the HMD, and equidistant conditions. Of the camera driver's 15, 30, 45, 60, 75 and 90 fps settings we chose to use 30, 60 and 90 to meet the above mentioned criteria. We used the highest available resolution at 90 FPS, 480p per eye, across all FPS conditions for consistency.

Another hypothesis was that discomfort would increase with higher head motion speeds due to an emphasized latency. In a short pilot study we asked five participants to mount the HMD and move their head at slow, medium and fast speeds. From the measurements of the HMD's rotation we derived a speed of 40 °/s as baseline and multiples of it with 80 and 120 °/s for medium and fast movements.

Based on the above described criteria, the study was designed as within-subject, and the following hypotheses were formed:

H_1: CamWarp has a stronger effect on low FPS values.
H_2: Discomfort increases with higher head motion speeds.

6.4.1 Participants and Apparatus

24 participants (19 and 69 years old - $M = 34.39$, $SD = 13.02$) took part in the experiment. The mean time per subject was 20 minutes. We used the Unity3D engine to render the VST AR on an Oculus Rift CV1 at its maximum available resolution of 1080×1200 pixels per eye over a horizontal FOV of 110° and the Ovrvision [237] stereo camera with its default intrinsic calibration and a horizontal FOV of 115° for the video feed. Figure 6.4 depicts the experiment room and superimposed virtual spheres that were used to guide the participants' view. A simple user interface was included through that participants could state their level of discomfort via a horizontal slider after each trial. They used an Xbox One Controller for input.

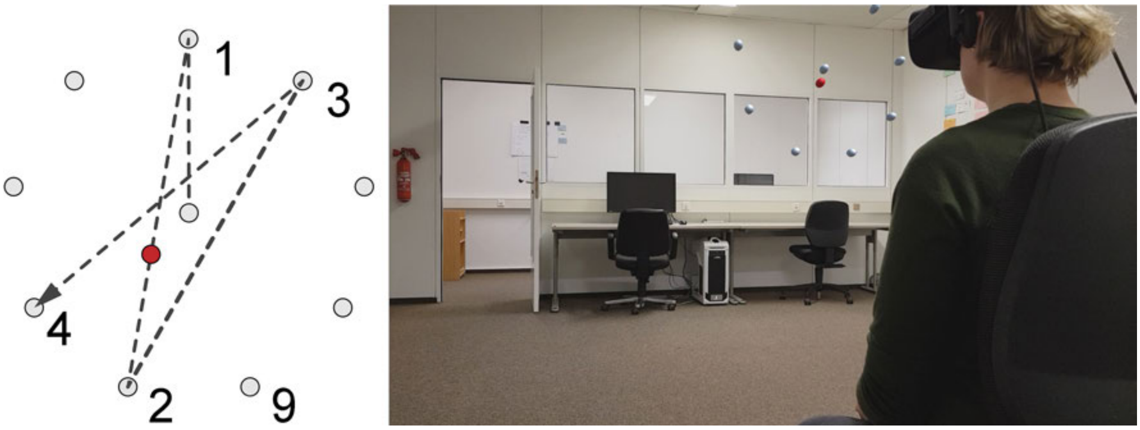


Figure 6.4: Our variation of MacKenzie's pattern [198] that guided the participant's view. One red and a ring of white spheres are superimposed one meter in front of the user's head.

6.4.2 Stimuli and Procedure

The experiment started with a demographic questionnaire and the SSQ [163]. The Participants were seated in a uniformly lit room and received an explanation of the term *discomfort*, i.e. the aggregation of anything that would make them want to leave the setup [101]. This includes cybersickness symptoms like nausea, disorientation and similar entries of the SSQ. We chose to base our experiment setup on the study of Fernandes and Feiner [101] because of its practicality. An increase to any of the SSQ's aspects has the same effect: it shortens the time users can expose themselves to VST AR. Therefore, an aggregated score is sufficient for this experiment.

Following briefing, the origin of the scene in Unity3D was reset to position the virtual spheres in front of the participant. To ensure that the data of all participants is comparable, their head movement was restricted by giving them the task to gaze at a virtual red sphere which moved in a Fitts' law inspired pattern [198] at a speed that produced the desired angular head movement. This also ensured an even distribution of yaw and pitch rotations. As illustrated in Figure 6.4, a ring of 9 white spheres was placed at a set 1 meter distance away from the participant. The red sphere starts at the center of the ring and then proceeds to move in a zig-zag pattern [198]. The first white sphere was randomly chosen in each trial. The participants were told to keep the red sphere as close to the center of their FOV as possible while paying attention to the video feed. The diameter of the ring was defined by the distance to the user (1 m) and the maximum desired head turning angle of 45° in any direction from the center. This large size was chosen to encourage users to rotate their whole head. This forced head rotation is crucial, since a stationary head does not trigger the cybersickness symptoms because of a static IP.

The red sphere effectively travels on a plane where the white spheres lie. That plane is oriented orthogonal to the virtual camera's view direction. A constant speed for the red sphere on that plane would therefore lead to an inconsistent rotation speed for the head and camera-rig. The head moves slower the greater the distance to the sphere. To counteract this behaviour, we instead moved the red sphere with a technique that relies on ray casting and projection. In the origin position of the virtual camera (after calibration) we located a virtual projector, that turns at a constant speed of 40, 80, or 120 °/s and sends a ray cast in front of itself. The red sphere is positioned at the point in space where the ray cast intersects the plane of the white spheres. We tracked the average deviation angle from the participants center of view and the target to determine if the instructions were followed. No participant showed a variation in any trial that would indicate an interruption, so no data had to be erased. The order of conditions was randomly assigned.

After each trial, a panel was displayed that asked the participant to input their subjective level of discomfort caused by the trial, on a scale of 1 (no problems) to 10 (severe discomfort) [101,

Discomfort	30 FPS	60 FPS	90 FPS
CamWarp	3.58 (SD=1.99)	1.69 (SD= .69)	1.61 (SD= .58)
Baseline	6.58 (SD=2.27)	2.99 (SD=1.98)	2.38 (SD=1.65)
Discomfort	**40 °/s**	**80 °/s**	**120 °/s**
CamWarp	2.06 (SD= .18)	2.15 (SD= .19)	2.18 (SD= .21)
Baseline	3.98 (SD= .22)	4.11 (SD= .27)	4.32 (SD= .34)

Table 6.1: Means of the dependent variable **Discomfort** for the combination of **Technique** and **FPS / Speed**. Lower is better.

Effect	DFn	DFd	F	p	$p < .05$	η^2
Technique	1	23	48.74	4.08 e-07	*	3.88 e-01
FPS	1	23	136.34	3.80 e-11	*	5.87 e-01
Speed	1	23	1.23	2.79 e-01		8.22 e-03
Technique x FPS	1	23	32.56	8.24 e-06	*	1.57 e-01
Technique x Speed	1	23	3.97	5.83 e-02		7.29 e-03
FPS x Speed	1	23	9.59	5.08 e-03	*	3.26 e-02
Technique x FPS x Speed	1	23	3.24	8.49 e-02		7.22 e-03

Table 6.2: Results of the ANOVA tests at the 5% significance level.

251]. The original question reads *On a scale of 0 to 10, 0 being how you felt coming in, 10 is that you want to stop, where are you now?* [101, 251]. This value is described as subjective discomfort and is used in the analysis by Fernandes [101] and Rebenitsch [251] to derive conclusions about cybersickness-related symptoms. We excluded 0 from our scale because of reports from initial test runs. They stated that at least one input on the controller's stick should be made before enabling the submission to prevent false entries. The experiment had to be a within-subject design because of the subjective nature of discomfort. By testing all configurations, participants could rate their experiences relative to each other. After successfully completing all 18 trials, the participants were asked to answer the SSQ a second time.

6.4.3 Results

The results of the experiment are shown in Table 6.1 and Figure 6.5 . We considered 3 conditions for FPS, 3 conditions for head movement speed and 2 conditions for the technique, for a total of 3×3×2 = 18 combinations. We ran a repeated measures ANOVA test at the 5% significance level for all variables to measure the effects of FPS, speed and technique. Table 6.2 lists all the calculation details.

6.4.4 Discussion

The results of the study demonstrate the effectiveness of CamWarp. It received better discomfort ratings compared to the baseline in all tested configurations (cf. Table 6.1). The mean for CamWarp 60 FPS is lower than the one of baseline 90 FPS, which suggests that it is possible to trade FPS for an increase in camera resolution to maximize comfort. Apparently, there is a point at which increasing FPS does less for comfort than increasing camera resolution, so these parameters need to be balanced carefully to optimize the experience. To quantify these potential benefits, an experiment with varying camera resolutions is required.

As expected, FPS has the greatest effect on comfort, confirming hypothesis H_1 (CamWarp has a stronger effect on low FPS values). Figure 6.6 also shows that the algorithm has a strong interaction

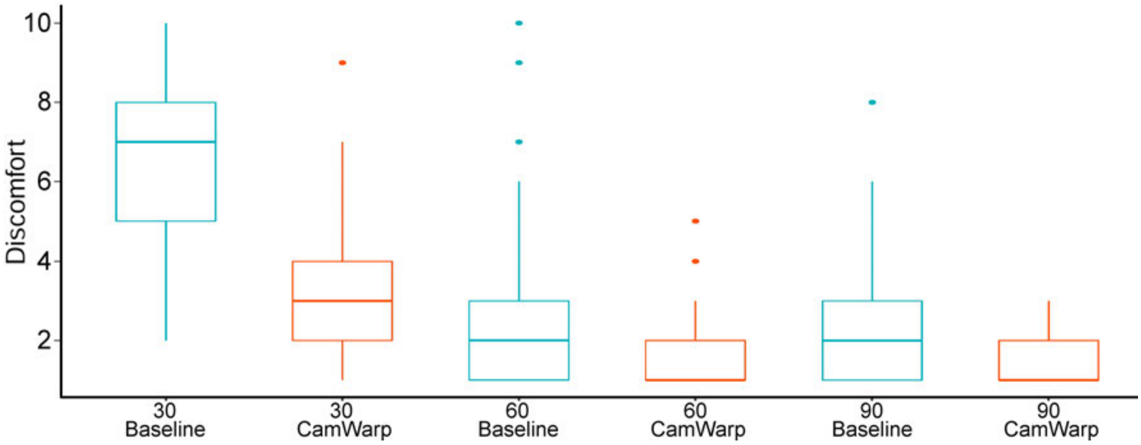


Figure 6.5: The distribution of the participants' polled **Discomfort** for the conditions **FPS** and **Technique**. Lower is better.

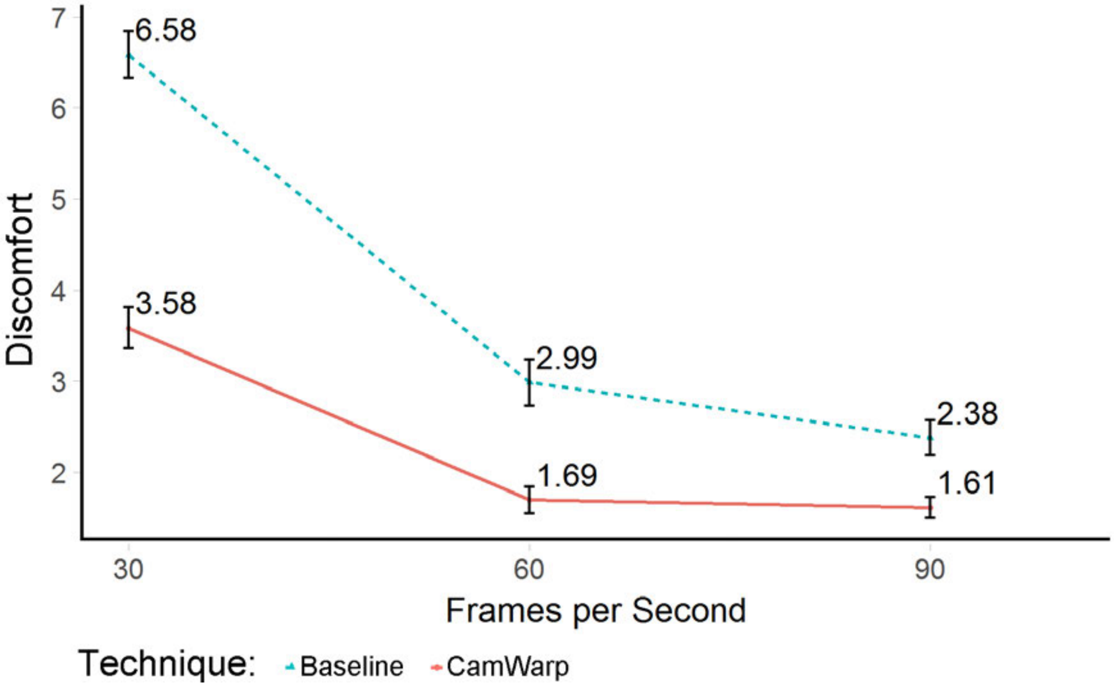


Figure 6.6: Plot corresponding to the mean of the dependent variable **Discomfort** for **FPS** and **Technique**. Lower is better.

effect with FPS, depicting a greater improvement in lower FPS settings, especially 30 FPS. This also matches our expectations. Meanwhile, the speed variable could not produce significant effects on the comfort level. Although, multiple participants reported that higher speeds made them feel uneasy, more so when the FPS were low. This is in accord with the significance of the interaction effect between FPS and head movement speed and indicates that head movement speed has a stronger influence on comfort if the FPS setting causes discomfort by itself, having essentially a super-additive effect. This is true for both baseline and CamWarp, as there is no interaction effect between technique and head movement speed. This matches the findings of studies that investigated the effect of head movement frequencies on cybersickness in AR [52], and thus partially confirming hypothesis H_2 (Discomfort increases with higher head motion speeds).

Contrary to what Nelson et al. [217] found in their study, we could not find a significant influence of time spent in AR on comfort, as the means for each trial slot (1 to 18) did not show a trend. This indicates that trials were not influenced by their precursors. This might be due to the strong variance in the discomfort that was experienced over the course of the experiment, as opposed to a linear accumulation.

Studies in VR [101] that reported a significant effect of their technique on discomfort concluded that their technique reduces cybersickness effects. Since our study showed a similar effect, this subsequently implies that our approach likewise decreases cybersickness effects overall. To determine if a lower aggregated discomfort score is sufficient to reflect such an overall decrease like Fernandes [101] and Rebenitsch [251] claimed, we asked participants to answer an SSQ [259] before and after the experiment. From a change in actual symptoms we can derive that participants correctly reported on experienced cybersickness effects during the experiment contrary to mere displeasure. We analyzed the data with a non-parametric Wilcoxon Signed-Rank test at the 5% significance level. Although the overall mean for all the questions in the second SSQ ($M = .177$, $SD = .171$) was almost 60% higher than the mean for the first ($M = .104$, $SD = .129$), we found no statistically significant increases. We attribute this to the circumstance that cybersickness is very subjective and participants were each susceptible to a subset of effects, which means that we need to look at individual entries of the SSQ instead. Applying the same test to the general discomfort entry resulted in a significant effect ($p < .001$) from before ($M = .0$, $SD = .0$) to after the experiment ($M = .375$, $SD = .564$). Similar effects can be found for fatigue (from $M = .208$, $SD = .415$ to $M = .416$, $SD = .583$), eyestrain (from $M = .25$, $SD = .442$ to $M = .75$, $SD = .737$), nausea (from $M = .0$, $SD = .0$ to $M = .125$, $SD = .564$) and blurred vision (from $M = .0$, $SD = .0$ to $M = 0.333$, $SD = 0.381$). This result suggests that participants reported lower cybersickness-related symptoms overall when CamWarp was enabled. This in turn increases the likelihood that users can have longer sessions in AR.

6.5 User Performance Experiment

The discomfort experiment showed the effectiveness of CamWarp for head movements in scenes with a static real-world environment. We therefore asked the question: Does our approach have similar beneficial effects in a real-world interaction scenario? In this section, we describe an experiment we conducted to confirm our hypothesis that this would indeed be the case. Like the previous study, we took the standard technique as a baseline and compared it to CamWarp. This time we compared them in regards to speed and accuracy of the execution of hand-eye coordination tasks in AR.

The participants had to move physical objects from one shelf compartment to another through a projected 3D-tunnel (Figure 6.7). The experiment is inspired by the steering law, which predicts movement speed in a tunnel on a 2D-plane. It reflects scenarios like a mechanic receiving virtual support for fixing an engine, a surgeon getting instructions for a cut or a warehouse worker sorting the inventory.

Our hypothesis was that CamWarp would have positive effects on user performance, so we opted to use the same camera configurations as in the first experiment (30, 60 and 90 FPS at 480p per eye). This includes a stronger impact on low camera FPS values and a significant effect on tasks that incorporate frequent head movement. Real-world applications typically do not restrict the speed at which the user's head moves, so there was no enforced head rotation in this experiment. Through the experiments design participants were encouraged to perform sidesteps and therefore move the camera with a mix of rotational and translational motions.

Based on the above described criteria, the study was designed as within-subject, and the following hypotheses were formed:

H_3: CamWarp has a positive effect on task accuracy.
H_4: CamWarp has a positive effect on task completion time.
H_5: CamWarp's effect is stronger on low FPS values.

Figure 6.7: The cardboard shelf and rendered 3D-tunnel. The color and animation were picked to associate flowing water, promoting quick identification of start and end points of the tunnel.

6.5.1 Participants and Apparatus

23 participants (20 and 38 years old - $M = 27.33$, $SD = 4.49$) took part in the experiment. The mean time per subject was 15 minutes. To track the objects in 3D space, we exchanged the HMD for the HTC Vive and its room-scale tracking, and fitted it with the Ovrvision. This headset features the same resolution of 1080×1200 pixels per eye over a 110° horizontal FOV. Figure 6.7 depicts the setup and the superimposed virtual tunnel that was used to guide the participants movement. The participants stood in front of a shelf that consisted of 9 compartments arranged in a 3 by 3 layout, each compartment measuring 35 by 45 centimeters.

6.5.2 Stimuli and Procedure

The experiment started with a demographic questionnaire and an SSQ [163]. The Vive controllers were put in the top left and top right compartments of the shelf to calibrate the pose of the shelf in the tracking space. This enabled occlusion of the tunnels through a co-located transparent model of the shelf. Following calibration, participants were tasked to move one of the controllers from the compartment it was located to an unoccupied one. Their movement was guided by a 3D tunnel that consisted of a white center line and a blue half-transparent cylindrical wall with a flowing animation (cf. Figure 6.7). It started at the center of the occupied compartment and extended 10 centimeters towards the user. From there, it continued in a straight line to a point 10 centimeters in front of the destination compartment, and finished at its center. Participants were tasked to pick up the indicated controller and move it to its destination as quickly as possible, while keeping the center of the controller's ring as close to the center line as they can. As these goals are conflicting, participants were asked to maintain a balance between the two. As soon as the controller reached its destination, the active controller and its destination were picked at random for the next trial.

Each combination of conditions consisted of 10 trials. We tracked three parameters during each. For *accuracy*, we measured the average distance from the center of the controllers ring to the closest point on the white center line. For *speed* we had to track the time between the active controller being picked up from its compartment and being laid down at its destination. Because the distance from one shelf compartment to another is not uniform, we chose to track the average

Accuracy	30 FPS	60 FPS	90 FPS
CamWarp	2.51 (SD=1.40)	2.22 (SD=1.36)	2.07 (SD=1.59)
Baseline	3.93 (SD=2.55)	3.46 (SD=1.77)	2.90 (SD=1.64)
Movement Velocity	**30 FPS**	**60 FPS**	**90 FPS**
CamWarp	.93 (SD= .30)	.98 (SD= .33)	1.13 (SD= .39)
Baseline	.85 (SD= .36)	.87 (SD= .39)	.88 (SD= .52)
Search Velocity	**30 FPS**	**60 FPS**	**90 FPS**
CamWarp	1.16 (SD= .24)	.96 (SD= .18)	1.39 (SD= .28)
Baseline	.55 (SD= .19)	.68 (SD= .27)	.62 (SD= .23)

Table 6.3: Means of the dependent variables **Accuracy** (lower is better), **Movement Velocity** (higher is better) and **Search Velocity** (higher is better) for the combination of **Technique** and **FPS**.

Accuracy	DFn	DFd	F	p	$p < .05$	η^2
Technique	1	22	27.74	2.77 e-05	*	3.91 e-01
FPS	1	22	13.82	1.19 e-03	*	1.46 e-01
Technique x FPS	1	22	2.88	1.04 e-01		2.79 e-02
Movement Velocity	**DFn**	**DFd**	**F**	**p**	**$p < .05$**	**η^2**
Technique	1	22	16.70	4.89 e-04	*	2.60 e-01
FPS	1	22	8.16	9.19 e-03	*	1.20 e-01
Technique x FPS	1	22	9.75	4.96 e-03	*	6.93 e-02
Search Velocity	**DFn**	**DFd**	**F**	**p**	**$p < .05$**	**η^2**
Technique	1	22	1.51	2.32 e-01		1.98 e-02
FPS	1	22	0.13	7.15 e-01		2.16 e-03
Technique x FPS	1	22	0.04	8.41 e-01		6.62 e-04

Table 6.4: Results of the ANOVA tests at the 5% significance level.

velocity of the controller as a function of the length of the current tunnel and the time from picking up to laying down the active controller to get comparable data on the speed. This will henceforth be called *movement velocity*. Similarly we tracked the average velocity of the participants hand from laying down the controller to picking up the next one as the average *search velocity*. The experiment was a within-subjects design to account for the individual differences in speed and accuracy. After successfully completing all 60 trials, participants were asked to answer a second SSQ.

6.5.3 Results

The results of the experiment are shown in Table 6.3. Due to the erasure of the rotation speed, we consider 2 conditions for the technique and 3 conditions for FPS for a total of $2 \times 3 = 6$ combinations. We ran three separate repeated measures ANOVA tests at the 5% significance level for all variables for accuracy, movement velocity and search velocity. Table 6.4 lists all the calculation details.

6.5.4 Discussion

The results of the user performance study confirm our hypotheses stemming from the cybersickness experiment. In all tested configurations CamWarp had a positive impact on the accuracy and the movement velocity of the controller (cf. Table 6.3), thus confirming hypothesis H_3 (CamWarp has a positive effect on task accuracy). We did not find an interaction effect between technique and FPS, meaning that CamWarp has a similar rate of improvement on all tested FPS values ($F_{1,22} = 2.88, p = .104$). Hypothesis H_5 (CamWarp's effect is stronger on low FPS values) was thus rejected. The means of the search velocity were favorable overall, but did not amount to a statistically significant effect. We assume that this circumstance is due to the low number of individual controllers. The VR system we used for our experiment only allows for tracking two hand-held wands. This potentially limits the study by introducing memory effects. Participants had the next controller to be grasped either directly in front, or could perhaps remember where they put it last. This is reflected in the FPS value having no significant effect on this parameter either. Hence, our assumption that CamWarp reduces the time it takes participants to find the next controller could not be validated, and hypothesis H_4 (CamWarp has a positive effect on task completion time) was rejected. A higher number of controllers could potentially have resulted in more head movements and searching, thus making the benefits of CamWarp more salient.

We found strong evidence for CamWarp's effectiveness in the movement velocity that represents the participants' ability to interact in MR setups, while searching appears to be restricted mainly to cognitive processes. Based on the effect sizes of the ANOVA tests, the technique has an even stronger effect on performance than FPS. This indicates that when participants could freely choose how fast they wanted to turn or move their head, the reduced latency of the CamWarp technique encouraged them to do so more quickly, resulting in faster movements overall. The interaction effect between FPS and technique suggests that high FPS have a positive effect on the effectiveness of CamWarp in regards to movement velocity. Like the super-additive effect of low FPS and no latency compensation on cybersickness, we see a super-additive effect of high FPS and latency compensation on movement velocity. The standard deviation is of considerable size in both measured velocities and the accuracy. This is due to the free choice of balancing these factors. Participants answered an SSQ before and after, with identical results to the first experiment.

6.6 Conclusion

We presented a novel and easy to implement technique to reduce cybersickness symptoms caused by latency of the video stream and improve user performance in hand-eye coordination tasks in VST AR. We contributed the implementation as well as evidence from a user study suggesting that this approach can be used to significantly increase comfort in VST AR, enabling longer sessions in such setups. Our results suggest that it is possible to lower the FPS settings of the stereo camera in favor of higher resolution, without decreasing the comfort for certain configurations. At the highest available FPS configuration, on average participants moved objects 28.6% faster while being 22.1% closer to the perfect trajectory.

Despite increases in camera sensor resolution and frame rates, a certain amount of latency from recording to display will most likely persist. With this latency being a constant in HMDs, compensating its effects remains a necessity and solutions like CamWarp will be relevant even when the hardware of VST HMDs advances. I.e., even 4K video per eye at 360FPS would still benefit because of the camera-to-photon latency. In comparison to the referenced techniques, we showed that CamWarp has a positive effect on hand-eye coordination and registration errors without the need for additional hardware or predictive algorithms. It masks black border artifacts during head movements while compensating latency and frame rate discrepancies. This technique can also be applied to setups with multiple stereo cameras, for example in foveated VST AR where a secondary stereo camera provides high resolution images for the FOV's center. Through CamWarp, the difference in latency between multiple cameras can be compensated to minimize cybersickness.

CamWarp can be applied to a wide range of setups that rely on camera input and positional tracking. We are interested in testing it on hardware that is responsible for the safety of passengers in vehicles. Confirming its positive effect on hand-eye coordination during optical flows originating from the center of the FOV in VST AR helmets for motorcyclists and pilots as well as a reduction in registration errors in camera-based 3D geometry reconstruction in autonomous cars remains as a target for future work.

VR Invite and CamWarp demonstrably increase the accessibility of shared MR spaces for mobile devices and VST HMDs. The hybrid user group aspect of heterogeneous device setups could thus be addressed successfully. In the next part we will discuss visualization and interaction techniques within shared MR spaces.

IV

Shared Experiences

7. Conveying Perspective

As discussed in Chapter 1, shared MR spaces offer additional information layers for human-to-human communication. In this context, passive information visualization mechanisms support the users' ability to forecast imminent actions of others. Spatial awareness of other users in shared virtual spaces is a necessity which is reflected in the usability ISO norm 9241-11. Unexpected behavior should be avoided as it leads to aversion, which also applies to the avatars of others. In this part we will discuss visualization techniques with regard to shared spatial awareness as well as the sense of spatial and co-presence. This includes rendering techniques for points of interest, avatar animations and visualization of locomotion.

Interactions with other people heavily rely on non-verbal communication. This is only possible to a lesser extent in shared MR spaces due to technical limitations, such as avatar fidelity and accuracy of facial expression tracking, which may affect the expressiveness of body language. However, a shared understanding of spatial relations is important in many collaborative or competitive scenarios [14]. This likewise applies to co-located users of the *VR Invite* smartphone application. In this chapter, we investigate the following RQ:

- *RQ*$_3$*: How should the individual perspective of a user on different points of interest be communicated, independently of available avatar types and tracking capabilities, and what are the implications for an observer's spatial awareness?*

7.1 Motivation

According to the original media richness theory, the richest communication medium is a face-to-face dialogue [74]. It is superior to video conferencing due to a number of factors, including depth perception and media naturalness [173]. However, collaborative MR systems have the potential to deliver an even higher effectiveness in communication, when the perceived naturalness of the medium will eventually reach that of a face-to-face conversation. In an indistinguishable MR, virtual face-to-face communication can be enriched by computer aided tools for brainstorming, design or other specific tasks[84]. One method of improving virtual face-to-face communication is to use visualization cues to convey what each user in a collaborative environment sees, freeing up other media channels like verbal or gestural communication.

There are several implementations of visualization cues, ranging from line-of-sight indicators to full displays of a user's actual image inside the headset (video mirroring). To find the most appropriate type of passive visualization cue for shared MR spaces, we implemented three distinct

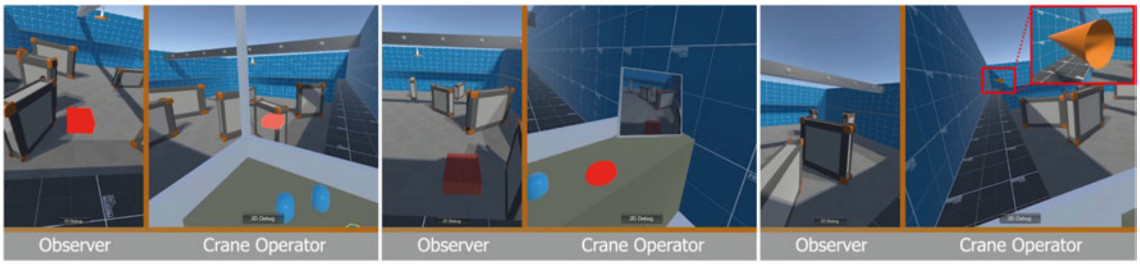


Figure 7.1: The implemented visualization cues. Left: *Object Highlight*. Center: *Video Mirror*. Right: *View Cone*.

universally applicable methods and compared them in a controlled study regarding task completion time, precision and error rate. The implemented visualizations are *Object Highlighting*, *View Cones*, and *Video Mirroring* [86, 89, 310].

Each of these perspective conveying methods is a direct or indirect interaction form between user & user or user & environment. There are many possible perceivable interaction forms as Manninen [202] defines them: "Interaction forms are perceivable actions that act as manifestations of the user-user and user-environment interaction. These forms are used to convey the actions of the user to oneself and to others. They enable awareness of actions by offering mutually perceivable visualizations and auralizations within the environment. In addition, the sense of presence is increased due to the communication, co-ordination and collaboration aspects these forms bring forward."

7.2 Background

According to Endsley [92], situational awareness is the ability to understand a complex situation and predict its future states in order to make decisions. A high level of situational awareness supports decision making, enabling an actor to identify one or more correct courses of action. In collaborative environments, being aware of other users' status and activities is crucial to coordinate activity and avoid interferences [93].

As collaborating team members develop shared mental models [113], their situational awareness grows [92, 93]. This in turn allows them to coordinate more efficiently [91]. Stemming from this situational awareness, the team's reliance on verbal communication is gradually reduced [200, 323], shifting the mode of operation to an implicit coordination [94, 96, 308]. This in turn lets team members utilize the verbal communication channel for exchanging additional information, or shift their attention to other cues from their working environment. While it is relatively easy to maintain social and positional awareness in co-located collaborative environments, remote collaboration proved to require a greater mental effort in regards to these types of awareness [131]. Thus, groupware research has focused on interface techniques that facilitate communication and increase group awareness and awareness cues [130].

In this context, Wuertz et al. [336] developed a framework to classify awareness cues in multiplayer games based on an analysis of games, game mechanics and game interfaces. They describe the information each type of cue provides, how it is communicated, and how awareness cue design choices can impact the user experience. Additionally, prior works on Workspace Awareness [130] and Gamespace Awareness [307] identified four types of information that team members use when collaborating, which also applies to MR CVEs. The *Who* - Presence and identity. The *What* - Status, task and social structure. The *Where* - Location, positioning, and gaze. And finally the *How* - Communicating the way important events occur. By emphasizing the gaze and enriching its expressiveness through highlighting or other rendering techniques, this type of awareness cue can be used to improve the efficiency of collaboration [86, 202, 310].

Regarding the visualization of awareness cues, Dodds and Ruddle [86] presented a suite of techniques that aim to improve the quality and quantity of communication in large-scale CVEs. Their social awareness increasing method, called 'group graph', which is an explicit hierarchical grouping system, visualizes who is currently talking to whom. Additionally, they addressed the difficulty of users to understand what other users are currently looking at, by rendering small viewports as planes in 3D space that displayed their respective user's point of view. The suite of techniques was evaluated using an urban planning scenario, highlighting that multiple viewports increased the maximum distance of effective communication between participants.

Wössner et al. [335] proposed a collaborative volume rendering application, that mirrors the point of view from one CVE user to another in a master/slave configuration. They compared the tight coupling of complete master control over the point of view to a more flexible approach, where the slave could change their rendered orientation, but not position. They concluded that the loose coupling approach was rated most comfortable and that user enjoyment was significantly higher in this scenario. Sonnenwald et al. [282] found that users saw potential benefits in both independent and shared perspectives, depending on the task and form of collaboration at hand.

Lee et al. [190] compared three different view-sharing techniques for MR scenarios of local users receiving support from an external expert. The real world viewport of the local user is transferred either as a flat 2D plane, a 360° projection on a sphere or a 3D reconstruction of the real environment. They concluded that transferring 3D reconstructions of the real world allowed the expert to give the highest degree of freedom in perspective, resulting in shorter task completion time and higher social presence.

Trapp et al. [310] described multiple approaches to highlighting objects of interest and potential fields of application. They divided the applicability of object highlighting into the categories *User Selection Preview*, *Visualization of Computational Results* and *Navigation Aid*. Here, *User Selection Preview* refers to rendering a single selected object, which is in the user's focus, in a different style. *Visualization of Computational Results* refers to highlighting, grouping and clustering results from database queries or other computational selection processes. Lastly, *Navigation Aid* refers to highlighting points of interest, routes, as well as navigation way points in order to guide the user's attention. Trapp et al. further categorized three rendering techniques for highlighting. These are *Style-variance*, *Outlining* and *Glyph-based* techniques.

Knowledge about the user's visual attention can also be used to facilitate mediated human-to-human interaction in MR environments. Duchowski et al. [89] demonstrated this by applying the eye movements of a user onto a virtual avatar, and additionally showed advantages of a visible line of sight for the communication of points of interest. Gamer and Hecht [109] further highlighted the importance of understanding other users' view cones. They investigated the influence of several factors such as observer distance and head orientation on user interactions and their view cones. They concluded that the processing of gaze cues plays an important role in social interactions, and mutual gaze in particular is relevant for natural as well as video-mediated communications.

7.3 Conveying Perspective

Based on prior work described above, we selected three distinct techniques of visual awareness cues and gaze visualization to determine which type is most appropriate for multi-user MR collaboration scenarios. We chose to differentiate them with Manninens [202] categorization of interaction forms between users and environment to have distinct approaches of visualization.

To represent a user-environment visualization cue, *Object Highlighting* was chosen. The information receiving user, henceforth *receiver*, does not have to look at the information sending user, henceforth *sender*. Instead, the additional information is embedded into the environment. The object of interest is being highlighted with a salient color that is visible through other objects and the environment, and therefore cannot be obstructed. To highlight an object, the sender has to simply have it in the center of their FOV. This technique is based on Trapp et al.'s user selection

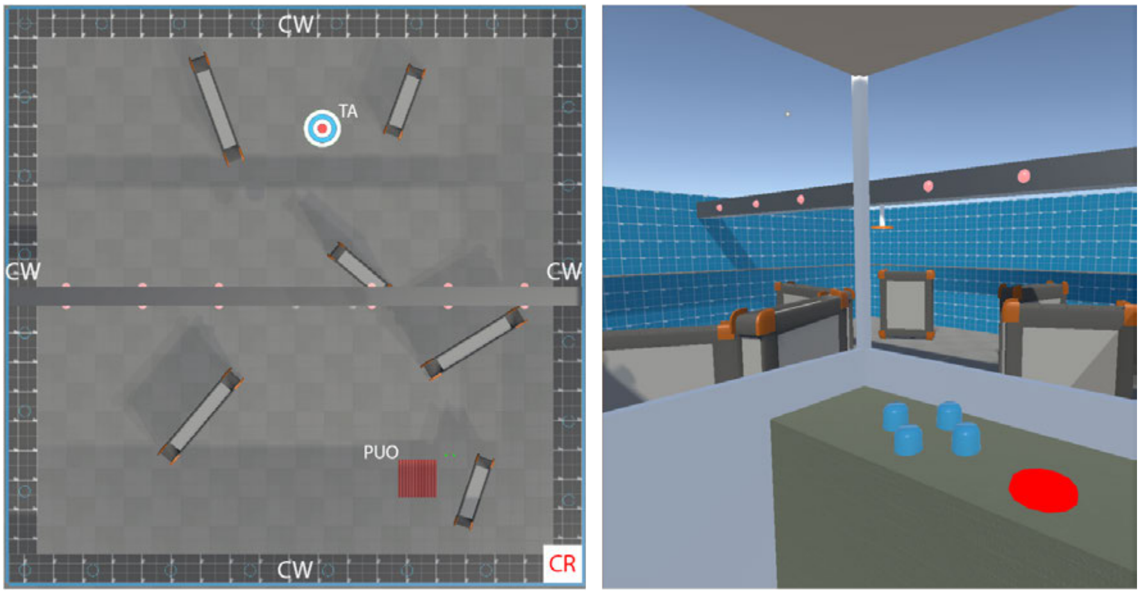


Figure 7.2: A view to the VE for the experiment.

preview, rendered with a style-variance approach [310]. The result is depicted in Figure 7.1, left. This technique is equivalent to the tile highlighting used in the *VR Invite* experiment described in Chapter 5.

As a representation of an enhanced user-user communication, the *View Cone* method was implemented. It is a simple pyramid or cone 3D model attached to the avatar head of the sender, making it easier to distinguish where a user is looking at from a distance. It is similar to a visible line-of-sight [89], but includes a representation of the boundaries of a user's FOV. Based on Duchowski et al.'s visible line of sight [89] and Gamer's evaluation of view cones [109], we combined both aspects into a visualization of a user's actual view cone. An example cone is depicted in Figure 7.1, right. This representation is similar in concept to displaying a 3D model of a smartphone inside the VE when using *VR Invite*.

A blending of user-user and user-environment enhancement can be achieved with the *Video Mirror* technique. It is a playback of the actual images displayed in a user's HMD, visible on a plane in the virtual space. It is an adaptation of Dodds and Ruddle's viewport visualization [86]. We originally planned to display the virtual image plane attached to the sender's avatar head. However, initial tests showed that this approach was only useful when sender and receiver were relatively close to each other, when the image plane was rendered large enough to read intuitively. In our experiment's test setup the users are potentially far apart, so the video plane was added as a fixed environment element in the vicinity of the receiver. To implement the video mirror, it is sufficient to place a virtual camera in the receiver's scene that is synchronized in pose with the sender's head. The result is depicted in Figure 7.1, center. As discussed in Chapter 5, this type of visualization can also be positioned such that the *Video Mirror*'s pose reflects the tracked smartphone's position in the real world when using *VR Invite*.

7.4 User Study

In this section we describe the user study that was conducted to evaluate and compare the three distinct visualization mechanisms. Our hypothesis was that each method would be better suited for varying tasks and scales of VEs. For example, we expected the *View Cone* to be less effective in large scale CVEs, while the *Video Mirror* should deliver a more constant and scale independent information flow. Therefore, we chose to design a within-subject user study based on a simple collaboration task with one assisting and one executive user with three different sizes of environment. In this experiment, the assistant is tasked with providing additional information to the executive by looking at relevant points of interest. The executive is tasked with interpreting this information, which would otherwise be unavailable, and interact with the indicated objects. More specifically, we built a virtual warehouse scenario in that the assistant indicates boxes that need to be relocated to a target area. The assistant can do so by moving through the VE via teleportation and thus provide different points of view. The executive is tasked with operating a virtual crane that moves on rails at the warehouse's ceiling, controlled by a simple virtual directional pad and two buttons. A second point of view is necessary to solve the executive's task as the target boxes and the target area are obstructed by the environment and other objects. This setup represents a selection task for environments with visual occlusions. As selection is one of the important canonical tasks in the field of 3D interaction and is a necessity in most applications, we chose it over other forms of interaction to be used to compare the effectiveness of the presented visualization techniques.

To isolate the effect of the enriched visual communication medium, we separated it from other information channels like gestures and verbal communication. As described in the motivation, the purpose of visual awareness cues is to leave the other communication channels unoccupied, as to enrich the MR collaboration medium with more concurrent information flow. As a result, verbal communication between assistant and executive was prohibited during the experiment. This ensures that the interaction between them relied solely on the particular visualization method. For this reason, we decided against testing the methods against pure verbal communication and compared them with each other regarding task completion $Time$, the *Precision* in placing the target box onto the target zone, and the $Error\ Rate$ as number of attempts to grab the target box.

Likewise, only the *View Cone* method employed a visible avatar, consisting entirely of the cone 3D model. As such, the executive has to exclusively rely on triangulation in this condition, and on the superimposed visualization techniques in the other conditions. Because of this, there is a clear separation between the different visualization methods and their effects on the dependent variables.

For the environment setup, we chose to have three differently sized warehouses (10x10m, 15x15m and 20x20m) to simulate different working conditions with other users remaining in viewing and cooperation distance. As an aid to assessing distances between participants, boxes and the crane gripper, the floor and walls consisted of 1x1m tiles with visual guidelines (cf. Figure 7.2). Furthermore, the assistant could use teleportation to freely move on a catwalk along the sides of the warehouse to find optimal vantage points. To ensure a consistent flow of information from assistant to executive between all participants, we chose to always have an experiment conductor take the roll of assistant and behave similarly during each trial.

Our hypothesis was that the *Video Mirror* technique would be superior in $Error\ Rate$ for all environment sizes, as it provides a directly rendered view on the target, albeit from another perspective, independent from the distance from assistant to executive. We further expected *Object Highlighting* to have the lowest *Precision*, as this technique renders the target object and target area in front of occlusions and therefore potentially disturbs the natural depth perception. The *View Cone* was expected to perform the worst in terms of $Error\ Rate$, as the executive has to triangulate where the assistant is looking at before placing the crane over the box. The larger the warehouse, the higher the $Error\ Rate$. Looking at another user's face to determine the gaze direction is also a real world process, indicating a higher medium naturalness. As the gripper moves fairly quickly on the Z-axis, grabbing attempts can be made in quick succession. Therefore, we did not expect a significant time difference between the techniques. Therefore, $Time$ was not explicitly announced as a critical factor but was recorded nevertheless to see potential differences between the means of methods.

Hence, we assumed the following hypotheses:

H_1**:** *Video Mirror* has the lowest $Error\ Rate$ for all environment sizes.
H_2**:** *Video Mirror* has the highest *Precision* for all environment sizes.
H_3**:** *Object Highlighting* has the lowest *Precision* for all environment sizes.

H_4**:** *View Cone*'s *Error Rate* increases with environment size.
H_5**:** *View Cone* is perceived as the most natural feeling method. It has the lowest degree of subjective loneliness.
H_6**:** Environment size affects *Time* stronger than the methods.

7.4.1 Participants and Apparatus

24 participants ($M = 24.66$, $SD = 3.31$, 8 female) took part in the experiment. The mean time per participant was about 45 minutes. Participants varied widely in 3D video game experience, time and frequency of play per week. Unity3D was used to render the scene on two HTC Vive Pros. Figure 7.2 depicts the VE used for the experiment. Networking was implemented via Mirror, synchronizing the visualization cues, the participants' position and orientation, as well as the crane and room setup.

7.4.2 Stimuli and Procedure

The experiment started with a demographic and SSQ [163], followed by a briefing about the nature of the task and an explanation of the different perspective conveying methods. The introduction to the task provided information about how to steer the crane with the virtual buttons and how to identify the box that needs to be picked up. The participant was then instructed to use the crane to pick up the indicated box and put it down on the target area as precisely as possible. In a trial phase the crane controls could be practiced before the experiment started. During the experiment, participants were confined to the virtual control room, while the assistant, who was always an experiment conductor, could move freely along the catwalks to silently provide visual cues. The assistant periodically changed their position to provide different points of view during each trial. Picking up the box required the gripper of the crane to be within a one meter radius around the center of the box (1 m^3), otherwise a red lamp in the control room was lit to indicate a failed attempt. The trial was considered complete when the correct box was picked up and then put down anywhere in the warehouse.

We used a different warehouse layout during every trial to reduce learning effects between the methods for trials with the same environment size. The order of methods was randomized, as was order of environment sizes within each method group.

7.4.3 Measures

Three distinct measures were recording during each trial:

- *Error Rate*: the number of grabbing attempts made,
- *Precision*: the distance from the target box to the center of the target area at the end of a trial,
- *Time*: the time from start to completion of a trial.

In addition to the demographic data and these measures, participants were asked to fill in an SSQ after every method trial group. After each trial, participants rated their subjective degree of loneliness on a 5-point Likert scale ranging from 1 to 5. After completing all trials, participants stated which of the three perspective conveying methods they preferred.

7.5 Results

In this section, we summarize the results from the experiment. The results were normally distributed according to a Shapiro-Wilk test at the 5% significance level. We analyzed the results with a repeated measures ANOVA test and a Post-hoc paired t-test for significant results. Whenever the Mauchly's test indicated a violation of the assumption of sphericity, the degrees of freedom were corrected using the Greenhouse-Geisser method. The full statistics can be found in the appendix, Section 13.

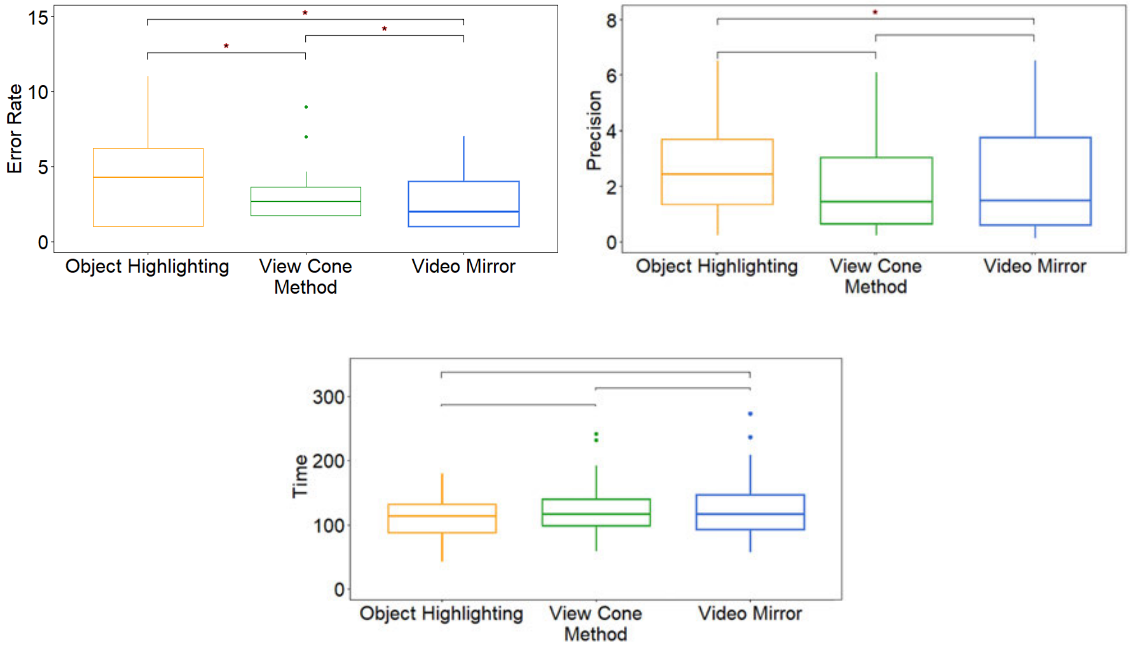


Figure 7.3: *Error Rate* (top left), *Precision* (top right), and *Time* (bottom) for *Methods*. Lower is better in all plots.

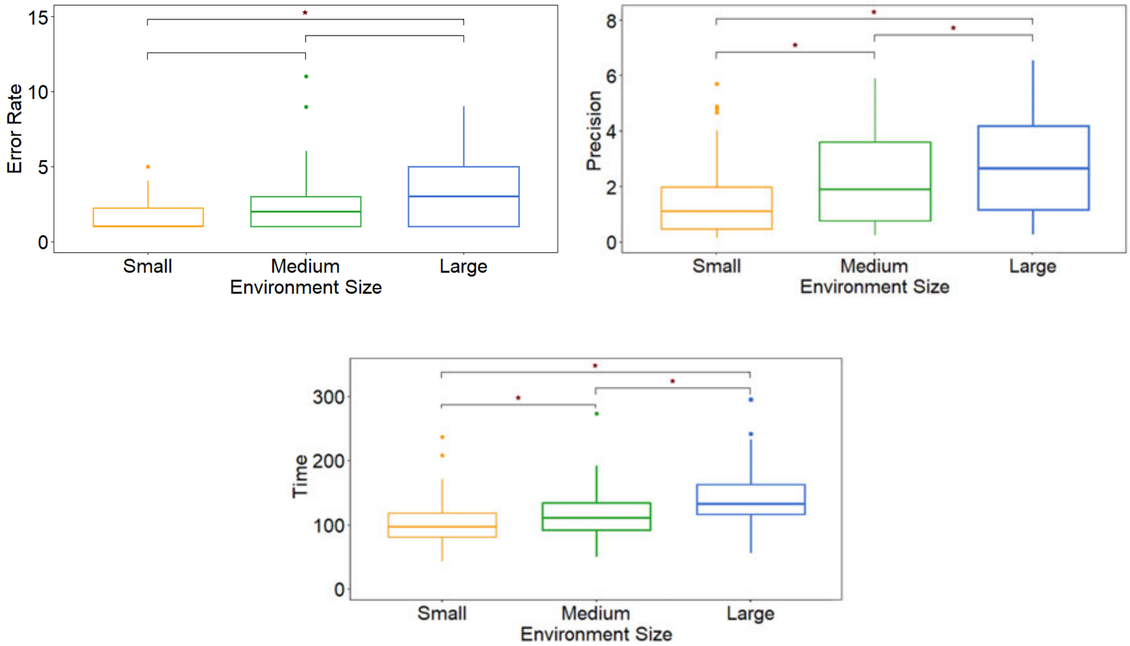


Figure 7.4: *Error Rate* (top left), *Precision* (top right), and *Time* (bottom) for *Environment Size*. Lower is better in all plots.

Error Rate

The results for the *Error Rate* are illustrated in Figure 7.3 (left) for the *Methods* and in Figure 7.4 (left) for the *Environment Size*. We found significant differences between the *Methods* and the

Environment Sizes. The *Error Rate* was significantly lower for *Video Mirror* ($M = 1.944$ attempts, $SD = 1.686$) than *View Cone* ($M = 2.639$ attempts, $SD = 2.387$) and *Object Highlighting* ($M = 4.806$ attempts, $SD = 6.856$). All the resulting details can be found on Table 7.1. Also, *Large* environments ($M = 4.26$ attempts, $SD = 4.88$) resulted in a significantly larger *Error Rate* than *Small* ones ($M = 2.21$ attempts, $SD = 3.96$, $p = 0.006$).

Precision

The results for the *Precision* are illustrated in Figure 7.3 (center) for the *Methods* and in Figure 7.4 (center) for the *Environment Size*. We found significant differences between the *Methods* and the *Environment Sizes*. The mean *Precision* was significantly higher for *Video Mirror* ($M = 1.837$ meters, $SD = 1.437$) in comparison with *Object Highlighting* ($M = 2.769$ meters, $SD = 2.064$), revealing that *Video Mirror* was the method with the highest recorded *Precision*. *Object Highlighting* showed the lowest *Precision*, with *View Cone* ($M = 2.374$ meters, $SD = 2.489$) in between. All the resulting details can be found on Table 7.2. Also, *Large* environments ($M = 3.15$ meters, $SD = 2.71$) resulted in a significantly lower *Precision* than *Small* ($M = 1.54$ meters, $SD = 1.37$, $p < 0.001$) and *Medium* ($M = 2.28$ meters, $SD = 1.56$, $p = 0.021$) levels. The difference between *Small* and *Medium* was significant as well (p = 0.003).

Task Completion Time

The results for the task completion *Time* are illustrated in Figure 7.3 (right) for the *Methods* and in Figure 7.4 (right) for the *Environment Size*. We found significant differences between the *Environment Sizes*. As expected, the mean task completion *Time* was significantly higher for *Large* VEs ($M = 151.07$ seconds, $SD = 60.67$) in comparison with *Small* ($M = 104.67$ seconds, $SD = 41.85$, $p < 0.001$) and *Medium* ($M = 123.46$ seconds, $SD = 53.29$, $p = 0.004$) levels. The difference between *Small* and *Medium* was significant as well ($p = 0.02$). There was no significant effect between the methods. All the resulting details can be found on Table 7.3.

Questionnaires

Differences from the pre- and post-SSQ gathered for each method were compared with a Wilcoxon signed-rank test. No *Method* caused a significant amount of cybersickness.

For the subjective degree of loneliness, a Mann-Whitney U test revealed a significant difference between the *View Cone* method and *Object Highlighting* (z = -2.55, p = 0.011).

After the experiment was finished, participants stated their preferred perspective conveying method. Here the *Video Mirror* method scored highest (50%), followed by *View Cone* (37.5%) and *Object Highlighting* (12.5%) (cf. Figure 7.5).

7.6 Discussion and Limitations

The results of the study highlight that methods which can be categorized as user-user interactions as opposed to user-environment interactions were not only rated favorably by participants, they also achieved a higher *Precision* and a lower *Error Rate*. Information provided by the assistant was easier to interpret for the executive when there was an incentive to look at the assistant.

Object Highlighting, representing a user-environment interaction, was rated the least favorite method. As a reason participants cited the perceived loneliness and the impracticality in large setups due to having difficulty with distance estimation. *Object Highlighting* likewise scored the highest *Error Rate*, signaling difficulty of the participants to adapt after a failed attempt. This might be due to the assistant not being able to provide a different perspective after a failed attempt, as the highlighting looks identical regardless of the assistant's point of view. A rendering technique that indicates the assistant's point of view with a color grading or similar effect might have a positive effect on this limitation. Furthermore, rendering the selected object over obstructing opaque geometry is an unnatural visualization, which can work well on 2D renderings [310], but evidently less so in stereoscopic 3D. To achieve a higher naturalness factor, the obstructing geometry could be rendered as translucent to imitate the real world experience of looking through tinted glass. Participants also had great difficulty setting the box into the target area with the *Object Highlighting* technique, as the assistant could not provide additional cues for the relative positioning of crane gripper to target area. When the box was picked up with the crane gripper, the box is in line of sight to the executive regardless of highlighting. This confirms hypothesis H_3.

With the *View Cone* method, the executive could triangulate the position of the target area by looking at the assistant and interpreting the gaze cue. This user-user interaction has a high naturalness factor, as it is possible to do so in the real world as well. Participants reported the lowest perceived loneliness for this method, signaling that this passive user-user interaction had a positive effect on the collaborative social experience, confirming hypothesis H_5. The exaggerated *View Cone* metaphor was successfully interpreted at even great distances, as there was no interaction effect with the *Environment Size*. Thus hypothesis H_4 was rejected. The *View Cone* similarly showed improved results over *Object Highlighting* when placing the gripper above the obstructed target area, with a superior mean *Precision* value.

The *Video Mirror* technique achieved both the highest *Precision* and the lowest $Error\ Rate$, confirming hypotheses H_1 and H_2. Like the *View Cone*, it could provide multiple points of view, enabling the executive to triangulate the position of the target box and target area. The additional rendering from a second perspective with a non-obstructed view on the working area transmitted the greatest amount of information to the executive. By constantly adjusting the point of view, the assistant could provide valuable information about the relative position of points of interest. An aspect that has not been explored is the potential difference between an assistant providing a second dynamic perspective and a static camera attached to the crane gripper. We hypothesize that the advantage of the assistant solution is the ability to reposition the point of view independently from the crane, as a fixed camera might not always give an unobstructed view on the points of interest. It is also a form of user-user interaction, while the fixed camera is a strict user-environment interaction.

Contrary to our hypotheses, there was no interaction effect between *Method* and *Environment Size* for *Precision* and $Error\ Rate$. We expected the *Video Mirror* technique to be the only consistent technique over any *Environment Size* level as the rendered image is displayed in a fixed position. However, all techniques scaled with the environment size similarly, visible in the significant effect of *Environment Size* on both dependent variables. We attribute this to the uniformly scaling difficulty of precisely maneuvering the crane gripper with medium and large *Environment Sizes*, while the methods were a minor factor for this scaling.

The *Time* scaled with *Environment Size* as well, but no significant differences between the methods were found, confirming hypothesis H_6. Multiple grabbing attempts could be made in quick succession, and did not prolong a trial significantly.

The implemented visualization techniques were chosen as representatives of the in the motivation described interaction forms between user & user, user & environment, as well as a hybrid approach. During and after the experiment, we received feedback suggesting that the environmental aspects of the visualizations were too specific to the task and setup and would be difficult to translate effectively to other applications. While we believe that this does not have an impact on the integrity of our study, we agree that certain changes need to be made in order to enable usage of the implemented and compared techniques in other contexts. We therefore recommend to make adjustments to each of the visualization techniques as follows.

For the *View Cone*, the shape of the visible cone should match the rendered view frustum of its user, making it a pyramid design. The size and transparency of the model can optionally be scaled with distance and angle to the observer. Also, adding a visible central line of sight indicator

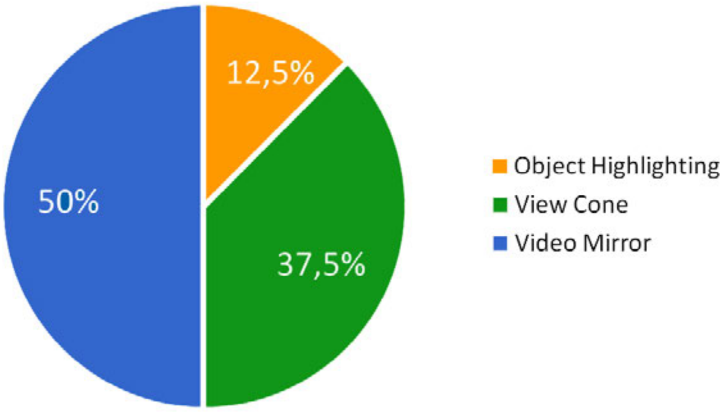


Figure 7.5: Results of querying the preferred method.

increases the readability of the visualization in medium to long distances. This might potentially have a positive effect on the sense of presence and readability, but needs to be confirmed in a separate study.

Our implementation of *Object Highlighting* only works effectively on objects of a suitable size and distance, as the entire object is altered with a style-variance shader. A better approach would be to lightly alter the appearance of the entire object, and use an indicator on the point where the user's line of sight intersects with the object's surface. This could be realized as a simple gradient of light, similar to a laser pointer. By always rendering the indicator in front of the object through manipulation of the z-buffer, the point of view becomes visible regardless of obstruction through geometry. This conveys not only selection, but also the relative position of observer and projector. Here, a combination with the aforementioned visible line of sight could be greatly beneficial. This also enables the usage of this technique to highlight environmental elements as opposed to predefined interactive objects.

The *Video Mirror* technique was implemented as a hybrid approach with a fixed position in the environment. It could prove superior to attach the viewport render as a world space UI element in the form of a floating head-up display (HUD). This might improve the sense of presence and interaction between users, as the user-to-user interaction has been received as presence enhancing compared to user-environment visualizations. A direct comparison would however be required to quantify these potential benefits. A non-stationary display is however independent of the environment setup.

Every tested gaze visualization technique proved to be effective and useful in certain scenarios and constellations. Depending on the task and environment setup at hand, we make the following recommendations for the choice of visualization techniques.

- If there are many sources of occlusion in the environment, the *Video Mirror* with the aforementioned HUD integration should be used. It demonstrated to be the most effective in our heavily occluded warehouse setup.
- If a selection between multiple objects has to be indicated, *Object Highlighting* can provide a quick distinction between similar or close objects.
- If the relative position between multiple users is of importance, it can be communicated via *View Cones*.
- If multiple factors apply, a combination of techniques can be used.

Depending on the context, individual techniques can be enabled or disabled. For example, the *Video Mirror* can be toggled manually as a HUD element when the user sees fit, or be active while the avatar of the other user is occluded. *View Cones* can similarly be enabled while the avatar is not occluded, and be scaled or faded depending on the distance to the observer. finally, the intensity of *Object Highlighting* can be derived from the distance between projector and the object and size of the looked at object.

7.7 Conclusion

Three distinct methods for conveying perspective in multi-user MR collaborations were selected and compared against each other in a real world inspired cooperation task. The *Video Mirror* technique proved to be the best performing technique in regards to *Precision* and *Error Rate*, as well as being the preferred visualization method. However, the View Cone was perceived to have the highest social presence. As evident from our user study, Object Highlighting could not establish itself as a convincing perspective conveying tool. While it is likely that highlighting objects which are in both the assistant's and executive's FOV improves visualization of user selection, it is not a feasible solution for passively conveying perspective and information that would otherwise be unavailable. This method makes verbal communication support necessary, resulting in a blocked communication channel. These findings suggest that users of *VR Invite* should be represented as either physical representations of the handheld device, or in form of a *Video Mirror*.

As an extension of the three distinct methods, it would be a conceivable approach to switch between multiple visualization methods automatically, depending on the context and distance to a point of interest. For example, objects within both users' FOV can be indicated with *Object Highlighting*, while communication over greater virtual distances can be supported by the *Video Mirror* technique. The size or visibility of a *View Cone* can be scaled linearly with the distance between users. The goal here is to achieve the best possible balance of naturalness and information exchange for any situation, without blocking other communication channels. Implementing a suitable framework for automatically switching between visualization cues and verifying a positive effect on social presence and task performance remains a target for future work.

In the next chapter, we will explore the effects of avatar systems and the depiction of locomotion on both the user's and the observer's perception within shared MR spaces.

8. Co-Presence

The ability to induce a high sense of spatial and co-presence is a major benefit of shared MR spaces over traditional non-immersive online conferences. Previous studies reported that the appearance of an avatar has an effect on the sense of co-presence of observers [58, 244], but a yet mostly unexplored aspect of human representation is the effect of locomotion visualization on observers with regard to spatial awareness. In this chapter, we discuss the hybrid user group aspect of differences in avatar representation and visualization of locomotion. In particular, the following RQ will be answered:

- *RQ_4: Does the depiction of continuous locomotion induce a stronger sense of spatial and co-presence and spatial awareness than abrupt teleportation movements?*

8.1 Motivation

As discussed in Chapter 1, shared MR spaces have several applications, for instance in the *new work* paradigm. Over the last decade, distributed work groups and remote collaborations have surged in popularity, making CVEs become increasingly relevant. In many cases, virtual meetings are the only viable solution for synchronous group communication over long distances [290]. In CVEs, various forms of technology are impacting how we communicate, learn, work, and make decisions [301]. Depending on the chosen technology and platform, however, many essential elements of human-to-human communication and interaction can be obscured. In particular, eye contact, spatial perception, and body language are only possible to a limited extent; sometimes even not at all. The popularity of MR headsets, however, has led to a degree of experimentation with MR group meetings in shared MR spaces. With devices such as the Oculus Quest 2, MR CVEs have become an affordable and accessible means to address some of the aforementioned limitations through services like Mozilla Hubs or Spatial.io. Yet, the potential of social interaction during MR meetings is still not fully explored, with many challenges remaining to be solved. While several aspects of self-movement such as presence, cybersickness and efficiency have been covered [104], insights into social aspects such as co-presence and perceived fairness in a competitive setting are lacking to date, which is unfortunate given the importance of these aspects on the user experience [254]. For example, locomotion is a limiting factor in current setups, as many developers have to rely on hand-based navigation through joysticks or teleportation. From an observer's point of view, locomotion techniques can only be distinguished as continuous or non-continuous movements, with optional animations. Only the visual and auditory representation of the locomotion can be perceived by the observer, which is bound to the moving user's avatar. Likewise, the appearance of a participant's avatar can be presented in a number of ways, from abstract to photo-realistic. Yet, it is not definitely settled if these design choices have a significant effect on the interaction between multiple users. In this context we contribute a study to determine the effects of locomotion techniques and the appearance of avatars on mainly the sense of co-presence and perceived fairness in a competitive multi-user VR game. Required related work in the areas of co-presence, CVEs, as well as avatar representation and locomotion techniques has been discussed in Chapters 2 and 5.

8.2 User Study

Co-presence is a vital aspect of shared MR spaces [343]. It can be impacted by a number of factors, including the appearance of other users' avatars and the visualization of their movement. These interactions have, however, to our knowledge, not yet been fully explored. To this end, we conducted an experiment to determine the effects of these two factors on mainly the sense of presence, co-presence, cybersickness, perceived fairness and game enjoyment in a two-user competitive VR game.

The VE for our study had to be deliberately designed with regards to encouraging interaction and visual contact between the participants. The other user's avatar had to be the main focus of the participant's task at hand, while also encouraging frequent movement. We decided to implement a game of a virtual two-player competitive snowball fight, with the goal to hit the opponent's avatar with snow balls, while dodging theirs in return. The VE consisted of a separated area for each user (cf. Fig. 8.1).

Figure 8.1: VE with scoreboard and abstract snowman avatar.

8.2.1 Independent Variables

In this experiment two different avatar appearances and two types of locomotion were used, for a total of four combinations. This design choice was made partially to keep the duration of the overall experiment short enough to not cause aborts because of the occurrence of cybersickness [82]. Based on previous works, we expected a rather small effect of the two independent variables on co-presence [20, 58]. A within-subject study design was thus chosen to cope with a relatively low number of available participants due to current health and hygiene regulations. For the locomotion types we chose teleportation and joystick-based navigation as representatives of non-continuous and continuous movement respectively, which currently are the most prevalent techniques in consumer VR solutions. Like in most previous studies, the teleportation was abrupt and not animated [45]. As a result, it also causes relatively little cybersickness to the teleporting user when compared to continuous movement [50, 104]. However, it also corresponds least to real or natural movement. The joystick-based locomotion technique was continuous and looked convincingly natural when paired with walking animations.

For the avatar appearance we chose one realistic and one abstract visualization (cf. Fig. 8.2). We chose to couple movement animations to the avatar appearance. The realistic avatar used predefined walking animations, while the abstract one was not animated. Participants also could not see their own avatar.

8.2.2 Dependent Variables

To assess users' self-reported sense of presence, co-presence, cybersickness, and game enjoyment, we used a questionnaire mainly based on the work of Podkosova et al. [244]. A compressed questionnaire was used to keep the overall duration of the experiment at a reasonable level, as it had to be answered eight times. Table 8.2.2 shows all questions included in the questionnaire. It contains items from both the IPQ [268] and the SUSP [280]. Presence and embodiment are not the main focus of the study; however, they may have an influence on co-presence and are therefore included. To assess embodiment, questions regarding body ownership and agency were used based on a modified version of Gonzalez-Franco et al.'s work [118]. Co-presence was measured using items of the the NMSPI [32, 33].

To measure potential side effects of cybersickness, we used a compressed version of the SSQ [163]. Here, we included elements that lead to very high dropout rates, namely nausea, vertigo and general discomfort [23]. Questions were answered via 4-point Likert scale, with 0 indicating no symptoms and 3 indicating severe symptoms.

In addition to these measures, we were also interested in implications for the sense of fairness during the game. We assumed that the ability of the teleportation to instantly cover long distances would be perceived as less fair due to its unpredictability, ending games with an overall lower received hit count. Accordingly, it stands to reason that the locomotion technique has an effect on the perceived level of fairness, and in turn the self-reported enjoyment of the game.

Besides subjective measures using Likert scales, some objective data was collected. This included distance traveled, number of snowballs thrown, and the number of received hits.

Based on the above mentioned criteria, we formulated the following hypotheses:

H_1**:** Appearance (avatars and animations) has an effect on the perceived level of co-presence.
H_2**:** The locomotion technique used has an effect on the perceived level of co-presence.
H_3**:** Fairness and game enjoyment are positively correlated.
H_4**:** Number of received hits is affected by locomotion technique but not by avatars.
H_5**:** Fairness is affected by locomotion technique but not by avatars.
H_6**:** Having both participants use the same locomotion technique increases game enjoyment.
H_7**:** Lower cybersickness leads to greater game enjoyment.

8.2.3 Participants and Apparatus

18 participants (9 female) took part in the experiment in pairs of two. Of these 18 data sets one had to be discarded due to technical errors during the experimental recording. However, the data set of his counterpart could still be analyzed. In total, the study lasted about 75 minutes. The age range was between 19 - 29 years ($M = 21.94$, $SD = 2.54$). Men on average rated their experience with VR ($M = 4.13$, $SD = 1.55$) higher than female participants ($M = 3.22$, $SD = 1.92$) on a 7-point Likert

Figure 8.2: The compared abstract snowman avatar [314] (left) and realistic avatar [5] (right).

Category	Item
Presence Q1	How aware of real events occurring in the room around you were you during the game?
Presence Q2	How involved were you in playing the game?
Presence Q3	Rate your sense of being there in the VE
Presence Q4	When you think about the VE, was it rather like something that you saw or somewhere that you visited?
Body Ownership Q1	I felt as if the virtual body was my own body
Agency Q1	It felt like I could control the virtual body like if it was my own
Agency Q2	I felt as if the virtual body was moving by itself
Co-presence Q1	To what extent did you have the feeling of the other player being together with you in the VE?
Co-presence Q2	Rate how closely the sense of being together with the other player resembles the sense of being with others in the real world
Mutual Awareness Q1	I was aware of the other player as a person
Attention Allocation Q1	I paid close attention to the other player
Attention Allocation Q2	The other player paid close attention to me
Behavioral Engagement Q1	My actions depended on the the other players actions
Behavioral Engagement Q2	The other players actions depended on my actions
Game Enjoyment Q1	How much did you enjoy playing the game?
Motivation Q1	How hard did you try to win the game?
Fairness Q1	It felt like I had the same chance of winning as the other player
Cybersickness Q1	General discomfort
Cybersickness Q2	Nausea
Cybersickness Q3	Stomach awareness
Cybersickness Q4	Headache
Cybersickness Q5	Vertigo

Table 8.1: The compressed questionnaire, based on Podkosova et al. [244].

scale. Similarly, men rated their experience with 3D computer games ($M = 6.38$, $SD = 0.74$) higher than female participants ($M = 2.4$, $SD = 1.67$). Two HTC Vive Pros were used as HMDs for the experiment, with two separate rooms for the tracking spaces. The VE was implemented in Unity3D, using the Unity XR Interaction Toolkit [317] as the basis for teleportation and interaction with objects. The Vive Wands were used for teleportation through a point-and-click metaphor, while the joystick movement used the Wand's trackpad. Networking was implemented through Photon Unity Networking (PUN2) [97].

8.2.4 Stimuli and Procedure

After giving their informed consent and filling in a demographic questionnaire, participants were introduced to the HMD devices and received an explanation of the competitive VR game. Both locomotion types could then be tested in a short tutorial section. Four two-minute runs of the virtual snowball game had to be played, each with a combination of avatar appearance and locomotion type in random order by latin square for one user, while the other used a fixed joystick locomotion and realistic avatar. Only the fixed participant's measurements were considered.

During the trial, participants had to compete by picking up a snowball from the ground, and successfully throwing it at their opponent. We decided that snowballs could be picked up in the immediate vicinity of the avatar (1 meter radius) without having to look and bend down to maximize the time a user is looking at their opponent, and trying to predict their next movement. To score points, users thus had to keep their attention focused on the opposing user's avatar in order to aim their throw.

After the throw, a new snowball was spawned within 2 meters of the participant on their side of the playing field, maintaining a constant total of 10 randomly placed snowballs per side. The required movement after a throw was therefore reasonably consistent between trials.

Each trial was followed by the questionnaire described in Sec. 8.2.1). After a full run of four trials and a small break, roles were switched and the experiment was repeated, resulting in a total of eight trials per pair of participants.

8.2.5 Results

We used an ordinal logistic regression model to analyse correlations between questionnaire items and the dependent variables. The recorded values from within our VE were normal distributed according to the Shapiro-Wilk test ($p = 0.05$) and could therefore be analyzed with a two-way repeated-measure ANOVA, followed by Bonferroni corrected post-hoc paired t-tests. A 5% significance level was assumed, and only significant results are reported.

8.2.5.1 Presence

Regarding sense of presence, Presence Q3 does not correlate with any of the other measurements, Presence Q4 correlates only with Presence Q2 ($p = 0.002$; $r = 0.383$), and Presence Q1 is negatively correlated with Presence Q2 ($p < 0.001$; $r = -0.497$).

8.2.5.2 Co-Presence

Co-presence, attentional allocation, and behavioral engagement, showed high intercorrelations, respectively ($p < 0.001$; $r > 0.4$). Agency showed a moderate correlation with co-presence ($p = 0.026$; $r = 0.273$). The ordinal logistic regression showed a significant effect of locomotion on co-presence. Here, continuous movement ($M = 4.59$, $SD = 1.37$) was rated significantly higher than teleportation ($M = 3.82$, $SD = 1.36$, $p = 0.031$). However, there was no significant effect for avatar appearance on co-presence ($p = 0.136$). Abstract avatars ($M = 4.09$, $SD = 1.46$) received similar ratings to realistic avatars ($M = 4.32$, $SD = 1.37$). We found several significant correlations between demographic data and co-presence, which are shown in Table 8.2.

8.2.5.3 Avatars

The variable Avatar correlated only with mutual awareness ($p = 0.022$; $r = 0.28$). According to our data it seems that the avatar has a small effect on the number of received hits, but this difference is not significant (Realistic $M = 4.29$, $SD = 2.80$; Abstract $M = 3.82$, $SD = 3.0$; F(1, 16) = 1.262; $p = 0.278$). The *distance traveled* did not vary significantly for either avatar (Realistic $M = 140.35$m, $SD = 30.51$m; Abstract $M = 137.35$m, $SD = 30.29$m; F(1, 16) = 0.826; $p = 0.377$).

8.2.5.4 Locomotion

The locomotion technique variable correlated with co-presence ($p = 0.021$; $r = 0.281$) and fairness ($p < 0.001$; $r = 0.514$). Users were hit significantly more often while using joystick locomotion ($M = 4.71$, $SD = 3.26$) compared to teleportation ($M = 3.41$, $SD = 2.35$; F(1, 16) = 6.505; $p = 0.021$). The *distance traveled* did not vary significantly for either locomotion technique (Joystick $M = 137.12$, $SD = 30.62$; Teleportation $M = 140.58$, $SD = 30.16$; F(1, 16) = 0.615; $p = 0.444$).

Locomotion technique and avatars had no significant effect on any of the self-reported user experience measures. Moreover, there was no interaction effect between avatar and locomotion technique in any metric.

8.2.5.5 Fairness

The locomotion type had a significant effect on the subjectively perceived level of fairness ($p < 0.001$), while the avatar variable was not significant ($p = 0.496$). Joystick $M = 4.94$, $SD = 1.71$; Teleportation $M = 2.97$, $SD = 1.54$; Abstract $M = 4.09$, $SD = 1.95$; Realistic $M = 3.82$, $SD = 1.85$. Nagelkerke's pseudo R^2 was 0.26 and the teleportation's predictor was -2.031. Thus, the model has an acceptable fit and, based on the negative sign, it can be seen that teleportation was perceived as less fair than the continuous joystick movement.

8.2.5.6 Game Enjoyment

Game enjoyment correlated significantly positively with fairness ($p = 0.019$; $r = 0.285$) and motivation, while correlating negatively with cybersickness, score, age, gender, left-handedness, computer game experience, and VR experience (cf. Table 8.3). No significant effect of locomotion technique or avatars on game enjoyment was found (Omnibus test: $p = 0.205$; locomotion: $p = 0.085$; avatar: $p = 0.653$).

8.2.5.7 Cybersickness

The total cybersickness score showed a high correlation with game enjoyment ($p < 0.001$, Negelkerke's pseudo-R^2 = 0.579). In addition to the total score, we tested individual items of the SSQ in regards to their influence on game enjoyment. We found significant positive correlations between game enjoyment with no ($p < 0.001$) or minor symptoms ($p = 0.027$) of stomach awareness. Similar results were found for no ($p < 0.001$) or minor symptoms ($p = 0.002$) of headaches. Severe nausea symptoms significantly reduced game enjoyment ($p < 0.001$), whereas no or mild nausea showed no significant difference. Other items showed no significant effect.

		Gender	Handedness	Gaming exp.	VR exp.	Enjoyment	Fairness	Motivation	Locomotion
Co-presence 1	Corr. coefficient	-.483	-.248	.419	.404	.321	.298	.325	.281
	Significance	.000	.043	.000	.001	.008	.014	.007	.021
Co-presence 2	Corr. coefficient	-.160	.014	.195	.398	.476	.305	.402	.256
	Significance	.195	.913	.114	.001	.000	.012	.001	.036

Table 8.2: Selection of significant correlations between co-presence and questionnaire items (by Spearman).

		Age	Gender	Handedness	Gaming Exp.	VR exp.	Fairness	Motivation	Cybersickness
Enjoyment	Corr. coefficient	.389	-.264	.317	.247	.370	.285	.642	-.439
	Significance	.001	.031	.009	.044	.002	.019	.000	.000

Table 8.3: Selection of significant correlations between Game Enjoyment and questionnaire items (by Spearman).

8.3 Discussion

Our results show several significant effects of avatar appearance and type of locomotion on the qualitative and quantitative data. Previous works found a significant effect of avatar appearance on co-presence in CVEs [58], as well as significant effects of partner perception and proxemics on co-presence in VR [244]. We found no such effect in our VR CVE, and thus hypothesis H_1 (Appearance has an effect on the perceived level of co-presence) has to be rejected. However, our study differs from others in regards to the nature of the task the participants had to perform. While most previous studies had participants perform a cooperative task, we conducted an experiment in a competitive setting. Previous work has shown that human emotions can strongly vary depending on the context of interaction between a group of people [197]. Here, Loch et al. determined that humans are hardwired to act social or selfish given specific group dynamics through evolutionary psychology. The converse argument could now be made that a given specific group dynamic, i.e. cooperative or competitive settings, the context influences human emotions, and more specifically co-presence. This leads us to the conclusion that a distinction for the effect of avatar appearance has to be made based on the context of the multi-user VR application. This is further supported by our finding that the avatars had a weak but significant effect on mutual awareness, which depends on co-presence according to Biocca et al. [33]. It is possible that our humanoid avatar produced an Uncanny Valley effect [213], however both types showed similar results. Weak effects of avatar appearance might potentially be revealed through more elaborate animations, using motion capture or IK [124]. Alternatively, a real-time 3D scan of the user could displayed in the VE, to potentially increase the sense of embodiment and thus co-presence. As participants could only see their partner's avatar and not their own avatar, it is not surprising that the sense of presence was not influenced by avatar appearance. Enabling this feature, paired with avatar customization as many commercially available solutions like Spatial.io provide, has been shown to increase the sense of presence and body ownership [329]. We found no correlation between presence and co-presence; this change however might have an effect on presence strong enough to reveal a correlation that was undetectable in our current setup.

Locomotion had a significant effect on the perceived level of co-presence, thus confirming hypothesis H_2 (The locomotion technique used has an effect on the perceived level of co-presence). Here, continuous movement was rated significantly higher than non-continuous movement (omnibus test, $p = 0.031$). Although the locomotion technique significantly influenced fairness, it did not affect game enjoyment, even though fairness and game enjoyment were strongly positively correlated. This correlation however confirmed hypothesis H_3 (fairness and game enjoyment are positively correlated). Teleportation was perceived as less fair, but did not significantly reduce game enjoyment. Thus, hypothesis H_6 (having both participants use the same locomotion technique increases game enjoyment) must be rejected. However, it is possible that fairness only had a small effect on game enjoyment due to the chosen study design and that in a context where fairness is more relevant (e.g. rankings and prizes for the winner), there could be a more pronounced effect.

Overall, participants reported that humanoid avatars were easier to read and continuous movements easier to predict. This corresponds to our recorded data, as participants were hit significantly more often while using the continuous joystick locomotion. This result is not surprising, since teleportation can cover long distances in an instant and the resulting movements are generally more difficult to predict. The other user's avatar, on the other hand, had no effect on score and subjectively perceived fairness, as expected. Thus, hypotheses H_4 (number of received hits is affected by locomotion technique but not by avatars) and H_5 (fairness is affected by locomotion technique but not by avatars) are confirmed.

Female participants overall reported feeling less co-presence. This correlation may at least partially be due to the fact that the humanoid avatar always had a male appearance. For this reason, we are interested in conducting a similar study with gendered avatars. With our equal numbers of male and female participants, we also found a correlation between gender and game enjoyment. This could be attributed to the fact that female subjects generally had less experience with VR and computer games, while these two variables also correlate positively with gaming enjoyment (VR experience: $p < 0.001$; $r =$ -0.490; Game experience: $p < 0.001$; $r =$ -0.727). This correlation can likely be attributed to the fact that experienced users were significantly more likely to land hits ($p = 0.014$; $r = 0.3$) and score was positively correlated with game enjoyment ($p < 0.001$; $r = 0.426$).

Three of the five cybersickness aspects also had an influence on game enjoyment. Hypothesis H_7 (lower cybersickness leads to greater game enjoyment) was thus confirmed at least with regard to the symptoms nausea, stomach awareness and headache. These results should be viewed with caution, however, as the vast majority of participants experienced no symptoms or only mild symptoms in the experiment. The three subjects who had moderate or severe symptoms greatly influenced the calculations. Thus, it is quite possible that if the experiment is repeated with more subjects, other or additional symptoms will become significant.

In this study, both participants were in different rooms and thus verbal communication was not possible. Since speaking is an important part of interpersonal communication, adding different forms of communication to the factors tested in our study might reveal interesting interaction effects. Furthermore, since we observed a number of weak but significant interactions, a larger sample size could potentially lead to further findings and implications [260].

8.3.1 Limitations

As our number of participants is not a multiple of the total combination of factors in our latin square arrangement of trials, the study is not fully counterbalanced, which means that there is a possibility of order effects. However, there was only one additional participant in one latin square entry, leading us to believe that this had a minimal impact.

As described in the Section 8.2, we used a compressed version of several standard questionnaires similar to Podkosova et al. [244], focusing on items relevant to our study. We deemed this necessary to keep the overall duration of the experiment at a reasonable level, as it had to be answered eight times. The questionnaire items were evaluated separately as opposed to groups of items, like some of the source questionnaires suggest. This could potentially limit the validity of results as we are not inquiring the full range of items. Therefore, we cannot give a total score and make direct comparisons to other works based on these values. However, individual comparisons of questionnaire items can still provide valuable indications and lead to conclusions.

8.4 Conclusion

In this chapter, we conducted a study to explore the effects of avatar appearance and visualization of locomotion in a competitive multi-user VR scenario. In summary, the avatar appearance generally had little influence on the observing participants, except for a weak positive effect of realistic avatars on mutual awareness. The avatar appearance thus offers creative freedom, at least in competitive game environments, and causes no significant effects in terms of game enjoyment, perceived fairness, cybersickness, or co-presence.

Compared to avatar appearance, the locomotion technique significantly influenced participants' perception of co-presence and subjectively perceived fairness. In addition, previous studies have shown an impact of the locomotion technique on the individuals themselves, such as differences in task performance and cybersickness [104].

In general, of course, it is beneficial if fairness, game enjoyment, and co-presence are as high as possible while minimizing cybersickness symptoms. However, this study has demonstrated that certain trade-offs have to be made when considering aspects that can be perceived both from the users themselves as well as observers. Here, the continuous joystick technique led to an increase in co-presence while the game is simultaneously perceived as more fair. On the other hand, the joystick technique led to higher levels of cybersickness which in turn led to less game enjoyment. In conclusion, designers have to balance between personal presence, cybersickness and enjoyment versus that of other participants in multi-user scenarios.

While it is useful to present users a choice of locomotion type, designers of VR applications need to keep the aforementioned advantages and disadvantages in mind, from both perspectives of sense of (co-)presence, as well as fairness and enjoyment. This suggests a greater emphasis should be put on the visual representation of teleportation-based navigation in multi-user VR, for example by animating positional transitions. These findings motivate testing the effects of interpolated continuous animations for non-continuous movement from an observers perspective. Implementation of such transitions as well as testing implications on co-presence will be discussed in Chapter 10.

In this part, we showed that visualization cues and natural representation of human locomotion are important aspects of human-to-human communication in shared MR spaces. Thus, in the following part, we will explore methods for providing natural looking locomotion with full-body human avatars from both first- and third-person perspectives.

V Embodied Interactions

9. VR Strider

Locomotion and its visualization are integral aspects of shared MR spaces, from the viewpoint of both the moving user as well as observers. In this part, we introduce a number of novel motion-based techniques for traversing the VE, which are portrayed by procedurally animated full-body human avatars.

Convenience of usage is an important aspect of any form of end-user technology. Prolonged sessions of HMD's quickly become tedious, as users typically have to stand stationary within their tracking space. Because of this, some experiences offer their users to remain seated. However, this has several implications, such as a lower point of view and potentially a reduced sense of presence or limitation of available locomotion types. In this chapter, we explore mechanisms to bridge the difference between sitting and standing modes of hybrid user groups with regard to locomotion and its visualization. The following RQ will be answered:

- *RQ$_5$: How can a locomotion interface for seated users provide the illusion of walking? What are the implications for the sense of spatial and co-presence, and how should it be portrayed to observers?*

9.1 Motivation

Real walking is usually considered as the most intuitive way of locomotion in real and virtual worlds, and has been found to be more presence-enhancing compared to other forms of locomotion such as flying or joystick-based navigation [319]. Furthermore, walking has been shown to be superior over other techniques for complex spatial tasks [261], cognitive map building [262], and cognitive demands [203]. However, there are limitations of using walking as LUI since it is sometimes impossible to use real walking in immersive VEs [331]. The main limitation is the size of the tracked area and walking space. For instance, if the VE is larger than the tracked physical space, the user may eventually leave the tracked space while trying to reach distant locations in the VE. A variety of locomotion techniques have been proposed to cope with such physical restrictions, including walking in place [264, 279, 281], redirected walking [250, 291] or omnidirectional treadmills [79, 145]. Currently, the most common locomotion techniques, which can be found in consumer VR experiences, are teleportation and joystick-based navigation using game controllers [54]. While those LUIs can be used with the supplied input devices of a consumer VR kit like the Oculus Rift or the HTC Vive, they have several drawbacks compared to real walking. Those limitations include a reduced sense of self-motion, limited sense of presence, inferior spatial cognition, or more frequent

Figure 9.1: The *VR Strider* device including mini exercise bike and electronics for pressure tracking and vibrotactile feedback.

occurrences of cybersickness symptoms [54].

So far, most LUIs do not satisfy the full range of partly conflicting consumer needs including cost-efficiency, a small form-factor, presence-enhancement, low cause of cybersickness, appropriate for seated experiences, and being multi-purpose for different kinds of virtual locomotion like walking and driving. To fill this gap of current LUIs, we propose the *VR Strider*, a low-cost and small form-factor LUI that provides an illusion of real walking in VEs for seated experiences. The core idea is to translate motions of a modified mini exercise bike to a synchronized walking animation for a virtual avatar. Moreover, we enhanced the bike machine in such a way that audio-based tactile feedback built into the soles of the pedals simulates virtual ground contacts.

To summarize, the contributions of this paper are as follows:

- construction of a modified mini exercise bike for simulated walking in VR with vibrotactile feedback,
- realistic mappings from cycling to animated walking,
- development of the novel anchor-turning technique for fast and precise rotations,
- a user study to compare established in-place locomotion techniques with *VR Strider* regarding several metrics (including task performance, spatial cognition, sense of presence etc.) in a path-integration task, and
- a confirmatory study to explore the applicability of *VR Strider* to other types of virtual locomotion.

9.2 Background

Walking is a multisensory activity, which provides humans with visual, proprioceptive, vestibular and kinestheic cues while moving through the world [292]. While real walking is generally considered as the most natural and intuitive form of locomotion in the real world, its replication in VR remains a challenging task. However, treadmills and pedaling devices are well-suited for supporting locomotion [145], since they prevent users from stepping outside the tracking space while performing walking-like motions. Indeed, preventing displacements in the physical world limits self-motion cues provided by the vestibular sense while using such locomotion devices. Nevertheless, a large number of locomotion interfaces have been developed for use with specific devices, such as treadmills [48, 283], torus treadmills [154], virtual perambulators [154], foot platforms [155], pedaling devices [10, 42] and spheres [207]. In particular, Uniport was the first device built for lower body locomotion and exertion, which operates in a similar fashion to a unicycle [79]. In addition, different bicycle simulators have been implemented for a translation from real to virtual cycling, in full scale [309] or for smaller exercise devices [133]. For instance, Tran et al. [309] examined translation and rotation gains for such devices and reported that redirection can be applied up to an angle of 1.42° without the user noticing. Allison et al. [10] showed that participants significantly overestimated travel distances when they were moving simultaneously in the real and virtual space with a tricycle. When they only pedaled without moving in real space, or when only visual motion cues were provided, the estimation was more accurate. Xu et al. [337] compared joystick-based navigation, teleportation and walking-in-place techniques with regards to

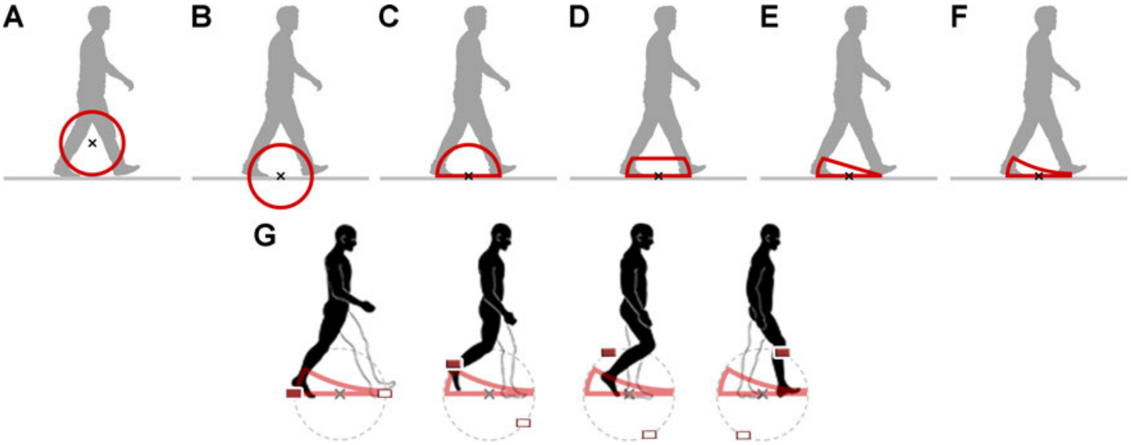


Figure 9.2: A-F: Iterations of the IK layer of the walking animation. G: The result superimposed on the reference diagram [296].

distance estimation, but could not find any significant differences. In a similar experiment, Kruijff et al. [178] demonstrated that multisensory input from audio, video and vibrotactile feedback have a positive effect on users' sense of self-motion and distance traveled. Kruijff et al. [179] further showed that leaning forward can have a positive effect on the perception of self-motion. Lastly, Dorado et al. [87] explored homing and path-integration in VR. They reported increased performance of participants with virtual techniques, which match proprioceptive cues from the real world. In their task, treadmills performed significantly better than joysticks.

In summary, the above discussed LUIs try to approximate or replicate natural human locomotion. Teleportation, joysticks, treadmills, leaning and walking-in-place can simulate several aspects of the natural human gait like proprioception, embodiment, a sense of self-motion, low motion sickness, visuals or haptic feedback, but not all of them at once to a convincing degree. Together with the above mentioned end-user restrictions and requirements such as cost-efficiency, physical space requirements and fatigue, it appears reasonable that no in-place technique could to date establish itself as de-facto standard for universal VR locomotion, and therefore teleportation remains as the prevalent VR navigation technique.

9.3 VR Strider

In this section we describe the design and development of the proposed LUI including the device and the locomotion technique that simulates virtual walking in a seated VR experience. Our goal was to provide a LUI that covers as many aspects of the natural human gait as possible, while having the ability to seamlessly switch to different kinds of virtual locomotion.

The device consists of four main components (cf. Figure 9.1):

1. The base, i. e., a mini exercise bike, that has a fixed position relative to the user's seat.
2. A freely spinning axis with foot pedals on each side.
3. Electronic parts, which are used for tracking the current axis rotation and the pressure that is applied to each pedal.
4. Two actuators embedded in the soles that provide audio-based tactile feedback.

The base consist of an exercise bike with a small form factor similar to [133]. We modified it with 3D printed pedals with a fixture for tracking markers, a large surface area for the vibrotactile soles and a compartment for the electronics. The circular movement pattern of the user's legs can be translated into any locomotion movement. As described above, we were particularly interested in replicating real walking, and therefore focus on a walking animation for a full-body avatar. The following sections discuss the mapping process, multiple turning methods and the provided tactile feedback.

9.3.1 Translating Circular Movements

The basis for all locomotion techniques that can be employed with the *VR Strider* hardware is the virtual replication of the circular movement of the user's feet. As previous studies have shown, full-body avatars have a positive effect on the user's overall sense of presence [279] and distance estimation [211]. Therefore, our goal is to provide a believable realistic mapping between the user's posture when using the device to synchronize the user's proprioceptive cues with the visual feedback. Virtual movement with real bicycles or pedalling devices has previously been used for bicycle simulators [309] and non-animated locomotion [79, 133] using simple tracking techniques like revolution counters or tachometers. As this is fairly imprecise and could not be used to convincingly animate the avatar, we opted to use the reliable tracking system of the HTC Vive with its universal standalone trackers.

To animate the avatar, its torso has to first be aligned with that of the user. The avatar's feet are then controlled with IK by setting IK targets to the coordinates of the HTC Vive trackers. The movement depends on the rotation of the bike's axis. To determine the speed at which the real axis ($axis_R$) rotates, we chose to build virtual representations of the pedals and the axis ($axis_V$). To correctly reproduce the current orientation of $axis_R$, the offset from the tracker to the joint between pedal and $axis_R$ attachment point needs to be considered; the center of the virtual pedals needs to be offset accordingly by the same vector. In our setup the trackers have an unambiguous orientation relative to $axis_R$, as they can only be mounted in one specific orientation to the pedal. From this information the pose of $axis_V$ can be deduced. Its position is the center point between the two virtual pedals, and its orientation is orthogonal to the trackers' orientations.

To infer a movement, the angle between the current forward direction of $axis_V$ and the same value from the last update cycle is calculated. For this, $axis_V$ and one pedal are centered on the X-axis in the local object space of the tracking origin. This allows us to find a vector from $axis_V$ to the pedal position on a plane in space that is perpendicular to $axis_V$' axis of rotation. Now, this vector can be compared to the vector of the last update cycle to find a signed angle, which is necessary to determine if $axis_R$ is spinning forward or backward. The distance the avatar needs to be moved can be calculated with the formula $distance = 2 \times Pi \times radius_R \times \frac{angle}{360}$.

9.3.2 Animating Virtual Walking

In order to provide a convincing illusion of walking in VR, the implementation needs to be extended in two key areas: the avatar animation adapting to the current movement speed, and separating the tracking of head and feet. In this section, we approximate the human gait cycle as described by Stoeckel et al. [296] by mapping the circular pedal motion to a realistic walking animation. Prior works have explored physics-based character animations [112] and tracking-based IK feet control [66]. However, the translation of circular movement patterns to animated walking is a novel approach.

Humans have a variable gait length, which is directly dependent on the movement speed. Under normal circumstances, the faster a person walks, the longer their step lengths [161]. To replicate this correlation, we found that a two layer animation setup is the optimal choice. The basic concept is to use a generic walking animation for the entire avatar body, and apply IK overrides to the feet. We used generic walking animations from Adobe's Mixamo service [5] to demonstrate that our setup will work with any animation for rigged humanoid avatars and can therefore be easily implemented into existing projects. A linear interpolation between a set of standing, walking, jogging and running animations was used for upper body, arms and ankles depending on the speed. For the next step, the animation needs to by synchronized to real leg movement and subsequently the current rotation of the bike $axis_R$. Instead of playing the animation blend at a certain speed, in each frame the progress of a full revolution of $axis_V$ is calculated on a scale of 0 to 1, where 0 refers to a defined default position. That value is then used to manually set the animation to a specific point in its normalized timeline. With this, the avatar animation will match the user's leg movement as far as having the left foot on the ground when the left pedal is at its lowest point and the same for the right side.

The deficit of generic animations is their predefined gait length and optimization for a specific movement speed. In practice this means that if the calculated speed is not exactly as intended for each individual animation, the forward momentum will not match the gait length of the avatar. This results in the feet sliding over the floor. The linear blending reduces this phenomenon, but it is still easily noticeable. The solution to this problem is to again use IK overrides for the avatar's feet.

In contrast to the translation from seated user to seated avatar, setting the IK targets directly to the tracker positions is impractical. Feet and head tracking operate in the same reference system, so their relative position to each other is always that of a seated person. As mentioned above, a separation of head and feet tracking is required. It is achieved by introducing a second virtual axis into the scene ($axis_{V2}$), which copies the configuration of $axis_V$. It is however positioned on ground level directly beneath the avatar and always faces the same direction as the avatar. The feet are then again bound to the virtual pedals by IK targets located at the ankle. This allows us to position the entire tracking space in the VE to align the camera with the standing avatar's eyes. Hence, the view in the HMD is already from the perspective of a standing person of a fixed height.

If those IK targets were to move in the same manner as the pedals of $axis_{V2}$, the feet's movements would be described by the red circle in Figure 9.2 A. To closely match a real gait cycle, $axis_{V2}$ has to be shifted down so that its center is exactly on ground level (Figure 9.2 B). Next, the IK target movement has to be limited on the Y-axis. Restricting the Y-value from being lower than ground level prevents the feet from penetrating the ground surface (Figure 9.2 C). On the resulting half-dome shape, a peak point is defined at maximum step height (Figure 9.2 D), and an impact point at the maximum step length (Figure 9.2 E). The upper limit is finally defined by a circle that tangents the peak point on the heel side and the impact point on the toe side (Figure 9.2 F). The resulting mapping is depicted in Figure 9.2 G as four discrete phases of a single step, where the filled and the blank pedals represent the right and left foot respectively. This shape is akin to a realistic human gait cycle as described by Stoeckel et al. [296].

Furthermore, the radius of $axis_{V2}$ needs to be scaled with the movement speed, keeping the overall shape of the gait cycle intact [161, 296]. To blend the generic animations and our IK layer, a smoothing function was implemented that slows down the scaling of the radius. It takes the last 30 samples of $axis_{V2}$ rotation into account when determining the exact size of the radius. This is especially noticeable when abruptly stopping or starting, when the avatar smoothly puts down its feet on the ground after coming to a halt.

9.3.3 Enhancing the Walking Illusion

After a first pilot test of our LUI, we received feedback that using the *VR Strider* still induces a feeling of bicycling rather than walking, especially when the virtual feet are not in the view of the user. To address this issue, a simple head-bumping function that uses a sine-wave for the horizontal and vertical movement of the camera was implemented and synchronized to the current rotation of the axis and therefore the animation of the avatar [264]. Furthermore, the avatar was slightly manipulated to lean backwards, bringing the virtual feet into the FOV more often. The angle to which the avatar leaned was dependent on the smoothed speed of the 30 last samples, exactly as the virtual leg radius. The result is a feeling of inertia of the FOV, which is in accordance with the real-world experience when transitioning from standing to walking and vice versa.

9.3.4 Turning Mechanisms

We implemented, tested and adjusted multiple turning mechanisms, which were meant to be used in conjunction or for different use cases. Besides a standard joystick input with a maximum rotational speed of $50^\circ/s$, we implemented three additional methods that are (i) purely based on foot pressure, (ii) gaze-directed with foot pressure triggers and (iii) hand-directed only, which are explained in more detail below:

(i) The main idea was to add pressure sensors to the bottom of the feet and detect when the user applies more force to one pedal than the other to allow natural turns while walking. The larger the difference, the sharper the turn. This method enables natural navigation that requires no additional input, similar to real-world walking.

(ii) The gaze-directed solution determines the rotation through a logarithmic function of the view direction. The turning mode is activated when looking to one side and deliberately applying pressure on the same side's pedal. It ends when the user's view direction is within a cone of 30° of the avatar's forward vector. The benefit of this implementation over prior gaze-directed navigation techniques [233, 263] is the ability to look to the side while walking on a straight line, while relieving the hands from all navigation tasks.

(iii) For the hand-directed navigation, a mechanism was implemented that we called anchor-turning. While the trigger is pulled, the wand acts as an anchor for rotations. The world rotates around the user's head in a similar way to a person sitting on a swivel chair, holding on to a fix point and turning themselves. A user can point at any object or reference point in the scene, and drag it into a different location relative to themselves. For example, pointing at a tree that is 45° to the left of the avatar's forward direction, and dragging it 45° to the right, rotates the avatar in such a way that the tree is now to the front of the avatar. The calculation is the same as for the rotation of $axis_V$. In the local object space of the avatar, the initial vector from the HMD to the triggering wand is compared to the current, resulting in the signed angle on the X-Z plane in world space. The angle is then inverted to turn the avatar setup in the opposite direction. For a standing avatar, the wand is anchored in the same position in world space, while the initial point keeps its relative position to the avatar and therefore turns with the user. Compared to bimanual [63, 295], gaze- [42] or pointing-based [65] navigation techniques, anchor-turning enables direct and precise single-handed turns while walking. The anchor is defined relative to the user and not the world, allowing for turns of any angle during movement. It is independent from the viewing direction and does not require a surface to point at. Therefore, it is a universal turning technique that can be applied to any seated or room-scale setup. To our best knowledge, this method of turning has neither been implemented, evaluated or published before.

9.3.5 Sole Haptic Feedback

Previous works have demonstrated the positive effect of multisensory integration of visual, audio and vibrotactile feedback on the users' sense of presence and distance perception [17, 178]. To explore the applicability of tactile feedback in simulated walking, we implemented a pair of soles that provide audio-based tactile feedback in both feet, in the medial plantar area using a haptic reactor[1]. The soles receive commands from the experiment PC, activate the actuator and reproduce pre-recorded audio files matching the properties of the material the user is walking on. For example, we reproduce an amplified audio stream of steps on grass if users walk over a grass surface, simultaneously to the audio feedback on the headphones. The intensity is defined by the walking speed. An Espressif ESP32 microprocessor controls the haptic reactor through an I2S 3W Class D amplifier and also collects pressure values from an array of six force-sensitive resistors (0.5 inches diameter) distributed ergonomically across the sole, transferring pressure data to the computer at a rate of 50Hz over Wi-Fi. The actuator-to-PC latency is approximately 28 ms.

The implementation's source code is available on GitHub [116].

9.4 User Study

In this section we describe the user study that we conducted to evaluate the *VR Strider* LUI. Our hypothesis was that the *VR Strider* would increase the users' perception of self-motion, distance

---

[1] $9x10x22.6mm^3$, https://www.alps.com/prod/info/E/HTML/Haptic

estimation and sense of presence compared to other established in-place locomotion techniques. Teleportation and joystick-based navigation were chosen for comparison, as these are among the most commonly used locomotion techniques in consumer solutions and in research [264]. Verifying our hypothesis required a test setup that allowed us to compare these vastly different interaction techniques. Therefore, we chose to design a user study based on a simple path-integration task, i. e., a triangle completion task as previously used for testing navigation performance [302]. Furthermore, we used metrics from similar previous studies that compared teleportation, joystick- and leaning-based approaches in regards to task performance (completion time and number of errors) [54]. This homing task required participants to walk in a triangle pattern.

As we provide multiple options for turning in the VE, a small focus group with 5 participants was conducted to find the most usable turning mechanism of the three alternatives described in Section 9.3. The results indicated that anchor-turning was the preferred method of turning. Participants reportedly found it difficult to consistently apply more pressure to one pedal than the other while pedalling. As the angle error and the distance error in the triangle completion task can be assumed to not be correlated [64, 302], the performance of real head turning can simultaneously be compared with joystick turning and anchor-turning, while evaluating the distance error for teleportation, joysticks and the *VR Strider*. Thus, it was necessary to disable the other implemented turning mechanisms of the *VR Strider* setup to have exactly one method of turning per trial.

We expected the *VR Strider* LUI to cause less cybersickness relative to the joystick condition due to the matching of avatar animation and real leg movement, in combination with a slight head-bumping. We assumed that this would minimize the sensory conflict between visual and proprioceptive systems, which is known to cause cybersickness [188], while supporting the sensation of walking. Furthermore, we expected tactile feedback to be the most significant factor in the users' ability to estimate distances from the wide range of parameters that can be adjusted for the bike [178], so we separated its usage as a standalone condition. Therefore, we compared *Teleportation*, *Joystick*, the bike with tactile feedback *Bike FOn* and the bike without feedback *Bike FOff*. Due to the four distinct methods, we chose a within-subject design.

Hence, we assumed the following hypotheses:

H_1**:** Illusion of walking improves the sense of self-motion. *Bike FOn* has the highest reported smoothness of movements and lowest difficulty to be accurate in movements.
H_2**:** Matching leg proprioception and avatar animation reduces sensory conflicts. *Bike FOn* and *Bike FOff* cause less cybersickness symptoms than *Joystick*.
H_3**:** Cycling is closer to walking than joystick-based navigation. *Bike FOn* is easier to learn and use than *Joystick*.
H_4**:** Tactile feedback improves distance estimation. *Bike FOn* has a lower distance error than *Bike FOff*.
H_5**:** Illusion of walking improves distance estimation. *Bike FOn* and *Bike FOff* have the lowest distance errors.
H_6**:** Proprioceptive feedback of leg movements improves the user's sense of presence. *Bike FOn* has a higher sense of presence than *Teleportation* and *Joystick*.
H_7**:** Anchor-turning is quick and precise. *Bike FOn* results in similar levels of angular errors as *Teleportation*.

9.4.1 Participants and Apparatus

20 participants ($M = 30.6$, $SD = 6.82$, 7 female) took part in the experiment. The mean time per participant was about 60 minutes. Unity3D was used to render the scene on an HTC Vive Pro and a non-swiveling chair to seat the participants within the tracking space for the bike and joystick conditions, and asked them to stand for the teleportation condition. Standing was required to let the participants use their full head and body proprioception for angular estimation as a baseline to test against. Depending on the condition they used an Xbox One Controller or the HTC Vive Wand for input.

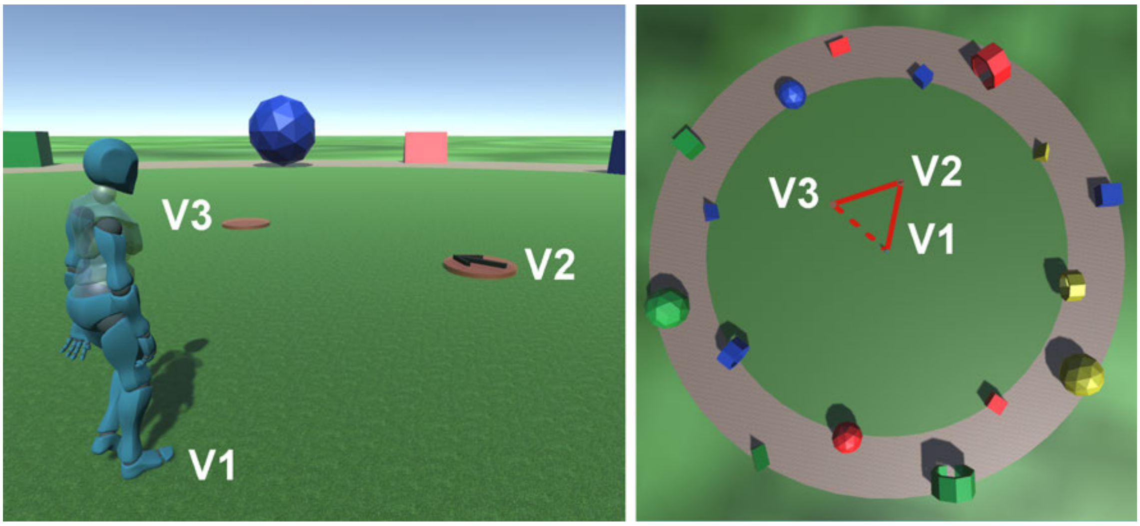


Figure 9.3: A view to the VE for the path-integration task.

9.4.2 Stimuli and Procedure

The experiment started with a demographic and SSQ [163], followed by a briefing about the nature of the task and the different techniques. Before each condition, the participants were given the chance to experiment with the input device in a test trial and were afterwards asked to perform the given task. In each trial, the participant was positioned at the scene's center. This position was defined as vertex 1 (V1) of the triangle (cf. Figure 9.3). V2 was always positioned 10 meters ahead of them and had an arrow marker on top that pointed to V3. After being guided to V2 and V3, participants were then tasked to move back to V1, as precisely as possible. When the participant had arrived at the estimation of V1, they confirmed by pressing a predefined button on the controller or wand. After completing all trials per condition, the participant filled in the questionnaires described in the following section. After the experiment was finished, participants were asked to state their preferred locomotion technique and what they liked or disliked about it.

The order of the techniques was randomized for each participant, as was the order of the different triangle shape configurations. Initially, we planned to test 3 distance levels between V1 and V2, 3 distance levels between V2 and V3 and 3 angle levels between V2 and V3 like [8] for a total of 27 trials per method condition. We ultimately had to reduce this number after the pilot study because of several reports of strong cybersickness effects during the joystick condition that led to premature aborts of the entire experiment. Based on [302], we used two levels for the distance factor between V2 and V3 (5 and 10 meters), and three levels for the angle between them (60, 90 and 120°). The resulting 6 triangles were then mirrored at random to the left or right to reduce learning effects without increasing the overall number of trials. The markers consist of a semitransparent red cylinder with a height of 10 cm and a radius of 25 cm. This height was meant to make the participant look at the ground when close by to bring the avatar's body into the FOV. The markers disappeared when stepped on. The ground in the predefined walking area was uniform as we didn't want to introduce additional cues for orientation in form of textures or vibration patterns, benefiting some conditions more than others. Like [302] and [8], we used salient and non-salient cues as landmarks and reference points for orientation with a ring of unique geometric shapes at a distance of 16 and 22 meters from the center.

9.4.3 Measures

The measures we used to evaluate the techniques are based on a comparative study between teleportation, joystick and leaning for a travel task in VR [54]. This enables an indirect comparison of our technique with their described leaning locomotion technique and extends the performance findings from their travel task by analysing distance and angle estimation.

Performance was measured in terms of time to complete the path-integration task, as well as the absolute distance and angle error from V3 back to the starting point V1. The time for the completion of each trial was logged from the first user input to them pressing the assigned confirmation button. Even though the comparison of the required time between instant teleportation, a joystick movement of arbitrary speed and an emulation of walking is unlikely to give insightful results, a difference between the levels of the haptics factor might be observable and was thus considered. The absolute angle error was defined as the unsigned angle between the two vectors originating from the V3, to the starting position V1 and the current position of the user respectively. To make the distance error independent of the angle error, it was not defined as the distance from the user to the starting position as prior studies did [8, 302]. Instead, the difference of the distances from V3 to the starting position V1 and the current position of the user was considered. This reflects the participant's ability to walk the correct distance after deciding on the direction, independent of the correctness of the chosen direction.

Cybersickness was self-reported with the SSQ [163], the sense of presence with the IPQ [268], and usability with the SUS [47]. Finally, comfort was self-reported with the Device Assessment Questionnaire (DAQ) [88]. It consists of 13 items, of which we adapted items 2, 5 and 6 to be equivalent to the comparative study of [54] and explicitly related to the split of measured performance in distance and rotation (smoothness of movements, difficulty to be accurate in movement and rotations) and added an item to measure leg fatigue. Each item can be rated on a 5-point Likert scale ranging from 1 to 5.

9.5 Results

In this section, we summarize the results from the experiment. We had to exclude four participants from the analysis. One of them misunderstood the task and three of them could not finish the session because of feeling slight nausea during the joystick condition. In all three cases the participants stated to generally be susceptible to VR and motion sickness. We included all other trials from the remaining 16 participants in the analysis. Results were normally distributed according to a Shapiro-Wilk test at the 5% level. We analyzed the results with an ANOVA test and Post-hoc Tukey's multiple comparisons at the 5% significance level with Bonferonni's correction. Whenever the Mauchly's test indicated a violation of the assumption of sphericity, the degrees of freedom were corrected using the Greenhouse-Geisser method. For the questionnaires only significant results are reported. The significance codes for figures are as follows: $^{***} \leq .001$, $^{**} \leq .01$ and $^{*} \leq .05$. The full statistics can be found in the appendix, Section 13.

9.5.0.1 Completion Time

The results for the completion *Time* are illustrated in Figure 9.4(left). The task *Time* was significantly longer for *Bike FOff* and *Bike FOn* in comparison with *Teleportation* and *Joystick*, showing that *Teleportation* ($M = 20.9$ seconds, $SD = 9.79$) was the fastest *Method*. All the resulting details can be found on Table 9.1. This result is not surprising as teleportation provides a discrete means to cover large distances, whereas *VR Strider* is a continuous locomotion method.

9.5.0.2 Distance Error

The results for the *Distance Error* are illustrated in Figure 9.4(center). The *Distance Error* was significantly lower for *Bike FOff* and *Bike FOn* in comparison with *Teleportation* and *Joystick*, revealing that *Bike FOn* ($M = 1.58$ meters, $SD = 1.33$) was the *Method* with less *Distance Error*. All the resulting details can be found on Table 9.2. This finding confirms hypothesis H_5 underlining that the proposed *VR Strider* LUIs is superior regarding distance perception.

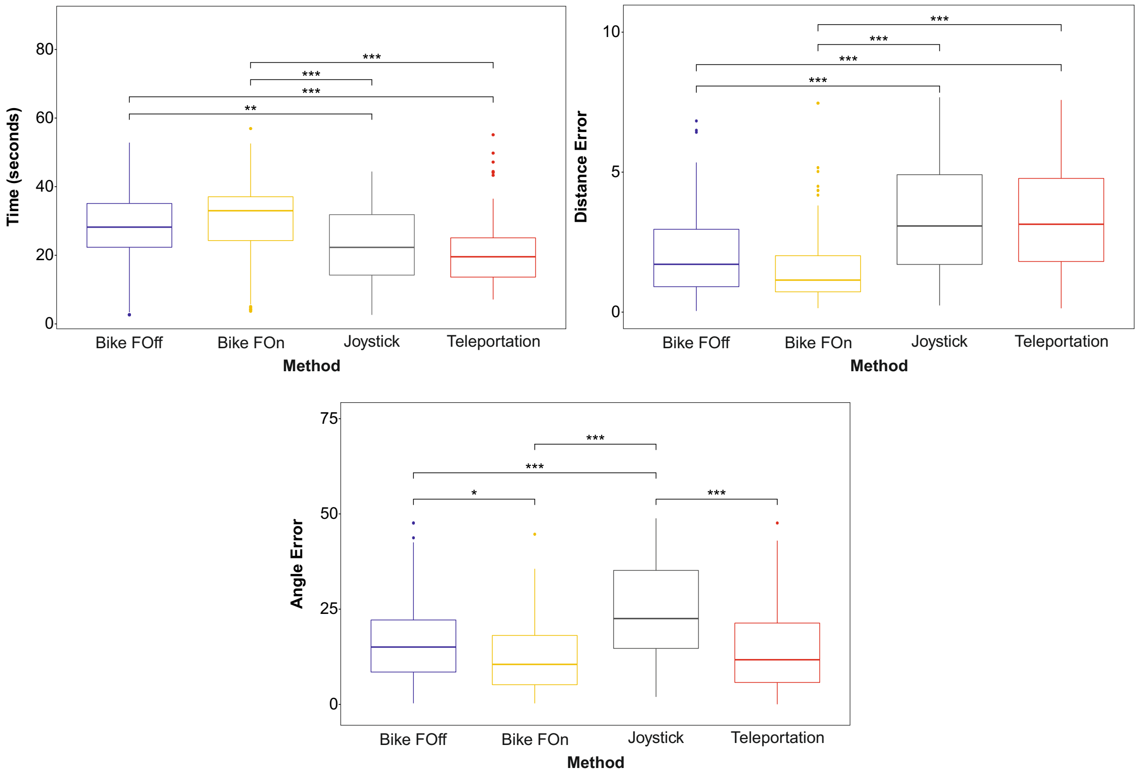


Figure 9.4: Task *Time* (top left), *Distance Error* (top right), and *Angular Error* (bottom). Lower is better in all plots.

9.5.0.3 Angle Error

The results for the *Angle Error* are illustrated in Figure 9.4(right). The *Angle Error* was significantly lower for *Bike FOff*, *Bike FOn*, and *Teleportation* in comparison with *Joystick*, revealing that *Bike FOn* ($M = 12.5°$, $SD = 9.41$) was the *Method* with less *Angle Error*. All the resulting details can be found on Table 9.3. This finding confirms hypothesis H_7 which underlines that the proposed anchor-turning is superior regarding spatial orientation.

9.5.0.4 Cybersickness

For the SSQ, we compared the differences from the questionnaire results gathered before and after every *Method* (POST-PRE). The differences were compared with a Wilcoxon signed-rank test. We found significant differences between the methods; *Bike FOff* ($M = 1.17$, $SD = 25.1$, $p = 0.004$), *Bike FOn* ($M = -6.54$, $SD = 27.3$, $p = 0.003$), and *Teleportation* ($M = -7.01$, $SD = 32.8$, $p = 0.003$) received lower cybersickness scores than *Joystick* ($M = 44.6$, $SD = 43.6$). The SSQ scores are illustrated in Figure 9.5(left). This finding confirms hypothesis H_2, underlining that *Bike FOn* can be used for longer sessions.

9.5.0.5 Sense of Presence

For the IPQ, we used a Wilcoxon signed-rank test to compare the differences between *Methods*. We found significant differences between methods; *Bike FOn* ($M = 62.6$, $SD = 13.9$) presented the highest score, and it is significantly better than *Bike FOff* ($M = 56.7$, $SD = 14.4$, $p = 0.018$), *Teleportation* ($M = 48.6$, $SD = 12.0$, $p = 0.003$), and *Joystick* ($M = 36.4$, $SD = 9.35$, $p < 0.001$). Also, the scores for *Bike FOff* ($p < 0.001$) and *Teleportation* ($p = 0.002$) were significantly higher than *Joystick*. Finally, *Bike FOff* also presented a significantly higher score ($p = 0.019$) than *Teleportation*. The IPQ scores are illustrated in Figure 9.5(center). This finding confirms hypothesis

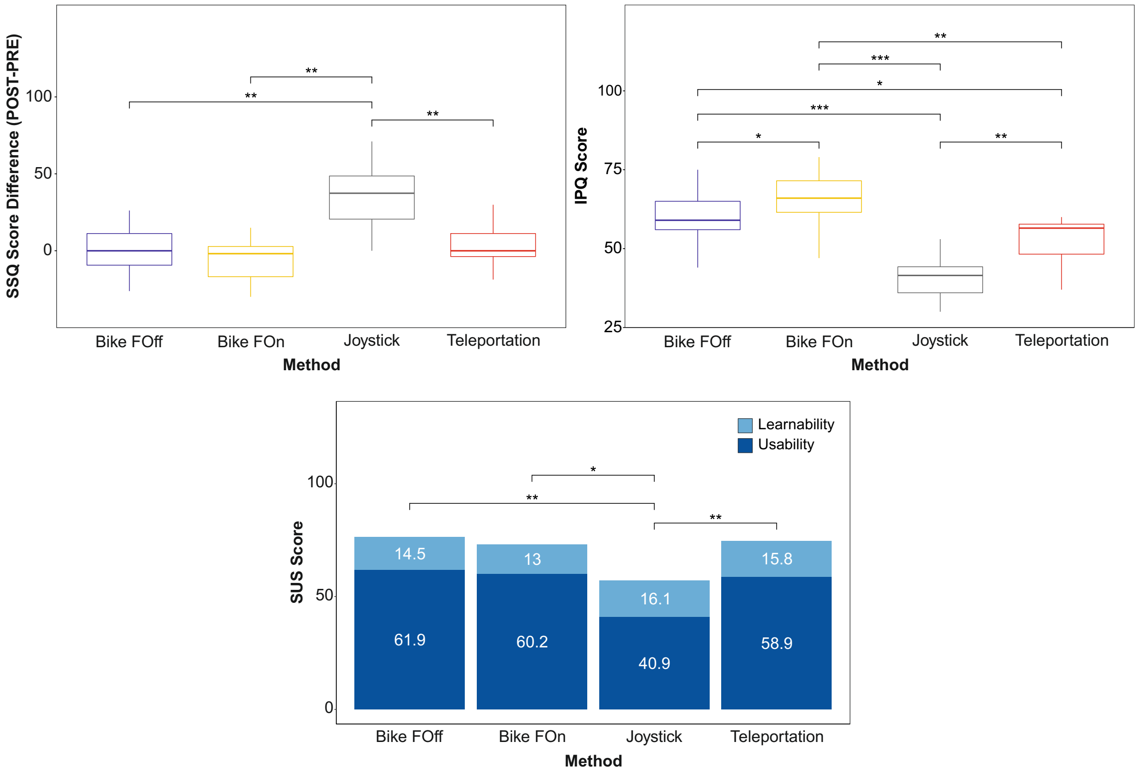


Figure 9.5: Results for SSQ (top left, lower is better), IPQ (top right, higher is better), and SUS (bottom, higher is better).

H_6 which underlines that leg proprioception has a strong impact on the sense of presence.

9.5.0.6 Usability

For the usability scale, we compared the questionnaire results with a Wilcoxon signed-rank test. We found significant differences between the methods; showing that *Bike FOff* ($M = 76.4$, $SD = 14.9$, $p = 0.004$), *Bike FOn* ($M = 73.1$, $SD = 16.5$, $p = 0.014$), and *Teleportation* ($M = 74.7$, $SD = 16.0$, $p = 0.003$) produced higher usability scores than *Joystick* ($M = 57.0$, $SD = 15.0$). Thus, according to the data, the usability scores for *Bike FOff*, *Bike FOn*, and *Teleportation* are similar. The SUS scores for the *Learnability* and *Usability* components are illustrated in Figure 9.5(right). This finding confirms hypothesis H_3 which underlines that the proposed LUI is easy to learn and use.

9.5.0.7 Comfort

We analyzed the Likert ratings (1-5) related to comfort (DAQ) with an ordinal logistic regression in order to find differences between *Methods*. Figure 9.6 illustrates the distribution of responses for items with significant differences. For the *required force*, *Bike FOff* required less effort than *Joystick* ($p = 0.021$). For the *smoothness of movements*, *Bike FOff* ($p < 0.001$), *Bike FOn* ($p < 0.001$), and *Joystick* ($p = 0.003$) produced higher scores than *Teleportation*. Also *Bike FOn* was better than *Joystick* ($p = 0.032$). For the *mental effort*, *Bike FOff* produced a lower score than *Joystick* ($p = 0.016$) and *Teleportation* ($p = 0.034$). For the *difficulty to be accurate in movements*, *Bike FOff* ($p = 0.01$), and *Bike FOn* ($p = 0.004$) produced lower scores than *Joystick*. For the *difficulty to be accurate in rotations*, *Bike FOff* ($p = 0.007$), and *Teleportation* ($p < 0.001$) produced lower scores than *Joystick*. For the *general comfort of usage*, *Bike FOff* ($p < 0.001$), *Bike FOn* ($p < 0.001$) and *Teleportation* ($p < 0.001$) produced higher scores than *Joystick*. The remaining scores did not present significant differences. This finding confirms hypothesis H_1, underlining that *Bike FOn* improves the sense of self-motion.

Finally, we used a matrix of Pearson/Spearman rank coefficients in order to correlate our demographic data with the dependent variables. We found that older people were faster with the methods *Bike FOff* ($p = 0.002$), *Bike FOn* ($p = 0.002$) and the *Joystick* ($p = 0.038$). There was no correlation with 3D/Games experience. Also, participants with high experience in VR/3D/stereoscopic displays presented significantly lower distance errors for *Joystick* ($p = 0.013$) and *Teleportation* ($p = 0.039$), and lower angle errors for *Joystick* ($p = 0.013$). When asked which of the four tested locomotion devices the users preferred overall, 81.25% stated they liked the *VR Strider* with enabled tactile feedback the most, citing the ease of usage, the smoothness of movement and the feeling of embodiment as most important benefits. 18.75% preferred the teleportation, with the most common reasoning being the efficiency in travel, and 0% preferred the *VR Strider* without tactile feedback or the joystick.

9.6 Discussion

The results of the study highlight the usability of the *VR Strider*. Regarding to task performance the proposed LUI had a positive impact on the participants' ability to estimate distances and rotations compared to joystick and teleportation. To begin with, we will compare the two Bike conditions.

The results for the tactile feedback condition do not reflect a positive effect on distance estimation. While the mean of the distance error was favorable with feedback enabled (*Bike FOn*, $M = 1.58$m, $SD = 1.33$m versus *Bike FOff*, $M = 2.08$m, $SD = 1.57$m), it did not amount to a statistically significant effect ($p = 0.15$). Thus, hypothesis H_4 could not be validated. We assume that this circumstance is due to the low intensity the haptic actuator provides. This accords with reports from the open-comments section of our questionnaire, with 5 participants wishing for a stronger haptic feedback as it was too subtle while pedalling actively. We plan to solve the problem by using a more powerful surface transducer in our next hardware iteration. On the contrary, the results indicate an effect of tactile feedback on angular estimations ($p = 0.044$). There was no significant effect for tactile feedback regarding the time to complete the task. The means of these conditions were very close to each other, but the enabled feedback had a slight disadvantage (*Bike FOn*, $M =$ 31.2 s, $SD = 10.8$; *Bike FOff*, $M =$ 28.0 s, $SD = 10.3$). We attribute this to an increased feeling of embodiment, as we observed participants looking down much more frequently and closely observing their virtual feet, testing slow and fast steps during the experiment. Additionally, we found no significant differences for the self-reported cybersickness, usability or comfort for for the Bike conditions. There was however a significant effect of tactile feedback enabled ($M = 62.6$, $SD = 13.9$) versus disabled ($M = 56.7$, $SD = 14.4$, $p = 0.018$) on the sense of presence in the IPQ questionnaire. To summarize, tactile feedback resulted in slightly better orientation and slower walking, but not to a significant margin. The main advantage is an increased sense of presence and sense of embodiment through the feeling of impact that matches the visual representation of the feet touching the ground. Following this observation, we will only discuss the *VR Strider* with feedback enabled (*Bike FOn*) when comparing the device with *Teleportation* and *Joystick*.

As expected, the instantaneous *Teleportation* was the fastest technique with *Joystick* close behind. As previously discussed, the movement speed for the *Joystick* can be set arbitrarily (to a maximum of $2m/s$ in our case), making comparisons futile. It is noteworthy however, that participants chose to frequently observe their virtual feet when using the bike and tried different techniques like counting their steps, resulting in an overall slower progression in the task. We also found no significant difference for the distance estimation; in fact, they were almost identical (*Teleportation*, $M = 3.45$ m, $SD = 2.02$; *Joystick*, $M = 3.42$ m, $SD = 2.10$). *Bike FOn* on the other hand produced significantly lower distance errors ($M = 1.58$ m, $SD = 1.33$). The synchronization of proprioception through real leg movements, the visualization and tactile feedback proved to be greatly beneficial to spatial orientation. This is in accordance with prior studies that examined multisensory input and embodied translation cues [178, 218, 331]. While technical attributes like a technique's accuracy and latency might have confounded the results, we deem them minor factors as the difference between *VR Strider* and teleport/joystick is enormous (means of 1.58m versus 3.45m/3.42m, $p < 0.001$).

For the angle error, the *Joystick* condition was significantly worse than any other condition, while the implemented anchor-turning performed equally as good as real head turning or even slightly better when tactile feedback was enabled. Again, the proprioception of pointing at a reference point and moving it one-to-one with the arm has helped the participants to navigate in the VE. This matches participant comments stating that the anchor-turning feels very natural.

Regarding cybersickness, our hypothesis was confirmed that the joystick would induce the

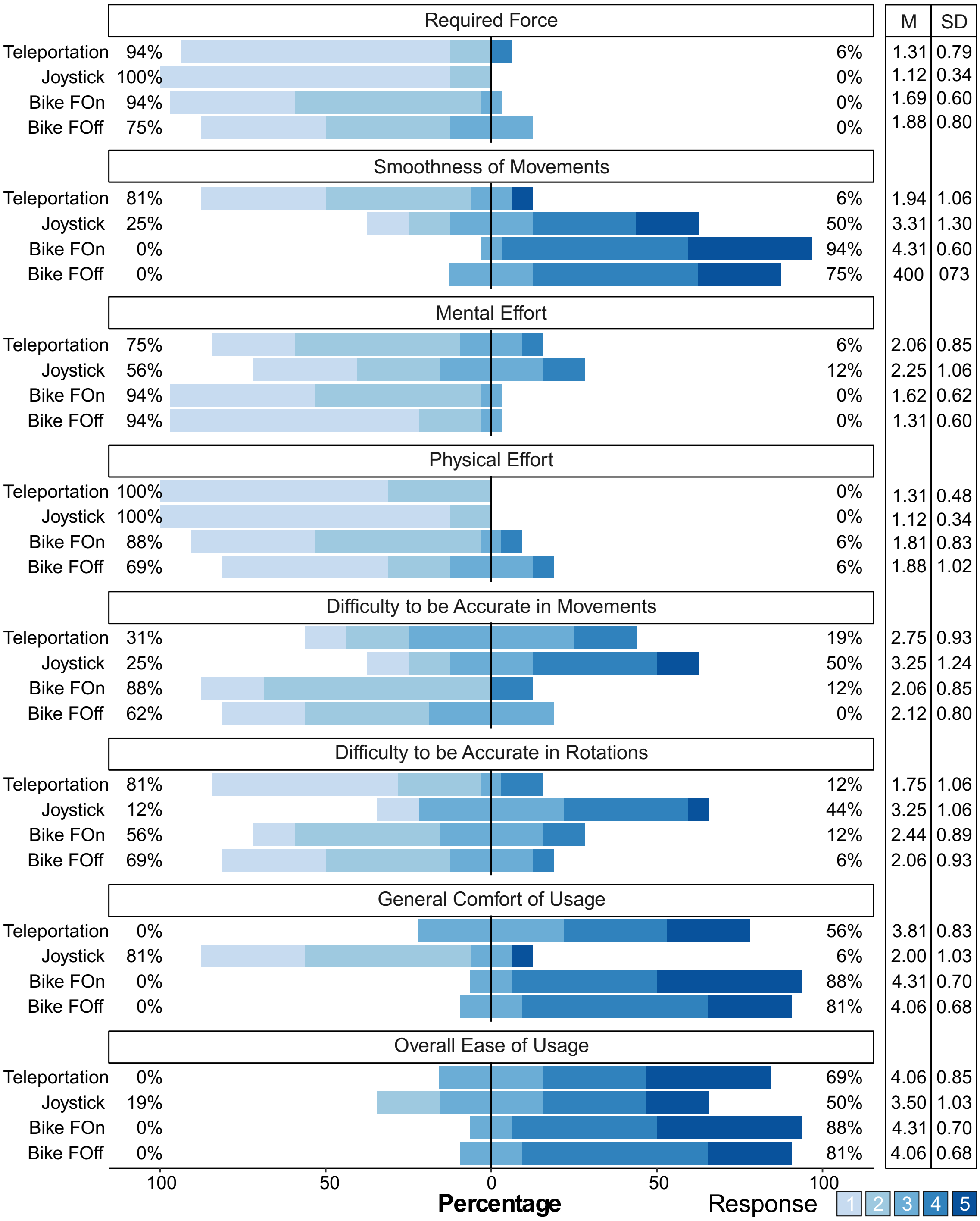


Figure 9.6: Results for the Device Assessment (DAQ).

most amount of cybersickness. There is no overlap of the visual input to the vestibular or proprioceptive system, leading to increased incidences of nausea, disorientation and oculomotor related symptoms. We found no significant increases from pre to post SSQ tests for *Teleportation* or the *Bike FOn*, meaning that both the pedalling and the anchor-turning do not induce cybersickess and are thus efficient and safe to use. This corresponds with studies that showed an improvement in cybersickness-related symptoms when using leaning or teleportation over indirect locomotion techniques [54, 249]. By extension, the *Bike FOn* causes significantly less cybersickness than leaning techniques [54].

The usability for the *Bike FOn* and *Teleportation* conditions were similar. Since the teleportation technique is generally considered to be easy to learn and our device achieved similar results, we can deduce that the translation of circular movement to virtual walking is easy to learn and use. An overall mean SUS score of 73.1 indicates that it can be integrated into a variety of applications for experienced or beginner VR users alike. Sense of presence was reported highest for *Bike FOn*, again confirming that leg proprioception has a strong impact on the sense of presence. Thus, the attempt to create an illusion of walking in VR while in reality being seated was an overall success. The reported comfort shows the *Bike FOn* to be superior in the sense of smoothness of movement, required mental effort, difficulty to be accurate in movements and rotations when compared to *Joystick* and *Teleportation*. The general comfort of usage was tied between *Bike FOn* and *Teleportation*. This is conform with the quantitative measurements of the experiment and supports our claim for the general applicability of our device for VR experiences.

Sarupuri et al. [264] compared Walking-In-Place (WIP) to teleportation, joysticks and Trigger-Walking with respect to usability and cybersickness. They reported similar overall SSQ scores for teleportation, joysticks and WIP, and an SUS score of 69.8 for WIP. TriggerWalking as a specific implementation of WIP achieved similar results as our teleportation and joystick conditions in a near-identical path integration task. This is conform with other studies that found no significant differences in distance estimation between these techniques [337]. In comparison to WIP, it can be assumed that *VR Strider* improves distance estimation due to the precise matching of leg proprioception and momentum. It also received higher usability ratings, in particular for leg fatigue. This suggests that a cycling motion is a closer metaphor for walking than WIP.

To summarize, *Joystick* is not an appropriate method of locomotion for VR experiences. Their usage resulted in high distance and angular estimation errors, and a high amount of cybersickness, while providing a low sense of presence. They lacked in usability and had a low general comfort of usage. *Teleportation* had an equally low rate of success for distance estimation, but real head turning was vastly superior for angular estimation. It showed less overall occurrences of cybersickness, had a better sense of presence and a higher usability score. The sense of smoothness of movements was unsurprisingly very low, but the general comfort of usage was rated significantly higher than the *Joystick*. The *Bike FOn* was the slowest tested locomotion technique, but achieved the best results in distance and angular estimation. cybersickness was comparable to *Teleportation*, as was the usability. However, regarding sense of presence and comfort it was rated higher than *Teleportation*. Overall, *Bike FOn* performed at least as good or better than *Teleportation* and *Joystick* in all quantitative and qualitative measures except time.

9.7 Confirmatory Study

As described above, the *VR Strider* can also be applied to other forms of virtual movement. As an indicator to see how it could be used as a multi-purpose locomotion device, we conducted a confirmatory study. The goal was to test the switch from a virtual walking to a virtual pedalling experience and vice versa. For this, the pedalling implementation from Section 9.3 was used and applied to a virtual GoKart as an example of a direct translation of real to virtual pedalling (see Figure 9.7).

Figure 9.7: A translation of real pedalling to a virtual GoKart.

Seven participants were asked to walk up to the GoKart and enter it via a button press on the wand, drive around a race track and then again exit the vehicle. In the second trial the visual representation of the pedalling was disabled and instead a motor sound was played, even though they still had to peddle to accelerate. The transition was done with a short fade-to-black when entering or leaving the vehicle from either side. With this setup we wanted to test if our device is restricted to walking and pedalling experiences, or to a more broad range of motorized vehicles. The feedback we received was generally in favor of using the manual GoKart over the motorized version. The consensus was that the manual GoKart worked really well and that switching from walking to driving and vice versa was a smooth and pleasant transition. Pedalling to accelerate the motorized vehicle however was described as surreal, awkward or even immersion breaking. These reports most likely infer a reduction in sense of presence.

This unfortunately indicates that *VR Strider*'s applicability is limited to experiences specifically tailored to walking or vehicles that are powered by pedalling like bicycles, GoKarts, paddle boats or airplanes with manual propellers. This needs to be confirmed in a fully controlled study, but already suggests that the hardware needs to be modified and extended to be useful for a wider range of VR experiences and thus be a universal VR input method.

9.8 Conclusion

We presented a novel LUI for VR that proved to be superior to established techniques like joysticks and teleportation in a path-integration task. We contributed the implementation as well as evidence from a user study suggesting that this approach significantly improves the sense of presence and comfort compared to joystick and teleportation. Results also showed that it does not cause more cybersickness than teleportation and is equally as learn- and usable. The provided tactile feedback further increased the users' sense of presence and embodiment. Anchor-turning proved to be a valid and universal technique for turning in seated or standing VR experiences. The tested implementation of pressure-turning however could not satisfy participants in our pilot study and needs to be iterated on. Even though a smooth transition from virtual standing to virtual seated driving is possible, the applicability of the device is limited to virtual walking and specifically made experiences that have a direct translation of real to virtual pedalling. Other forms of virtual locomotion like motorized vehicles need a different approach as the confirmatory study indicated the necessity of synchronized avatar animations to match proprioception and the virtual representation of legs.

Following our findings in the main study, we would like to further evaluate the positive effect of our realistic cycling-to-walking animation mappings compared to generic animations and naive IK implementations. Another major remaining goal is to improve the implementation of pressure turning and strafing. This could be realized through additional joints on the pedals that enable tilting or swiveling of the feet. To address the challenges of controlling virtual motorized vehicles, we plan to modify the base hardware. The main axis of the bike could be split to make the left and right side independent, with an electronic locking mechanism that enables split and joined rotations. This setup could simulate the behaviour of pedals used for acceleration and braking in cars or yaws in aircrafts and helicopters. Building and testing the required hardware remains as a target for future work.

In this chapter, we addressed the hybrid user group aspect of sitting and standing modes with regard to locomotion and its visualization. In the next chapter, we will investigate how full-body avatars can be applied to any type of continuous or noncontinuous locomotion, and how they are perceived by observers.

10. The (Non-)Continuous Toolkit

Teleportation is one of the most-often provided locomotion types, even in shared MR spaces [54]. This is likely due to a high degree of usability and efficiency while typically inducing only few cybersickness-related symptoms. However, continuous movements have been shown to be superior to teleportation with regard to both the sense of presence of its user and the co-presence of observers [54]. This is likely due to the abrupt visual nature of teleportation from both first- and third-person perspectives. In previous chapters we identified the importance of natural looking locomotion visualizations, and presented a technique for full-body avatar animations for continuous movement with *VR Strider*. In this chapter, we explore how full-body avatars can be applied to any kind of continuous or noncontinuous locomotion. The following RQs will be discussed:

- RQ_6: *Can continuous locomotion be modified to incorporate merits of teleportation such as a high usability and reduction of cybersickness, and how can noncontinuous movements be visualized as continuous movements to observers?*

10.1 Motivation

In shared MR spaces, multiple users can meet and interact with each other via their virtual representations. In order for users and their avatars to travel, in particular over long distances, typically teleportation is employed. While such teleportation provides an efficient and comfortable way of moving through VEs, previous work has shown that this technique might induce spatial disorientation and misperception of space and distances [291]. Indeed, a further limitation, which has rarely been studied yet, is involved when using teleportation in shared environments. When avatars abruptly teleport between different locations, it is irritating and challenging for observers to keep track of their movements and trajectories. In this context, we showed that noncontinuous visualizations lead to a reduction in co-presence and perceived fairness in a competitive game scenario [105]. On the other hand, from a first-person perspective, teleportation has several advantages over continuous locomotion, such as joystick control. For example, it has been shown that teleportation causes less cybersickness than joystick-based continuous camera movement that does not correlate to a user's physical movement [54, 186]. This results in a discrepancy between the needs of observers, who would benefit from continuous avatar movement by a user, and the user themselves, who may have a more pleasant VR experience through noncontinuous locomotion.

In this chapter, we approach this mismatch of user requirements from two directions. First, we address the problem from an observer perspective, by developing four novel techniques to visualize noncontinuous movements (e.g., resulting from teleportation) as continuous transitions through an advanced avatar system. Second, we consider the problem from a first-person perspective, by converting conventional continuous movements (e.g., resulting from joystick input) into a series of short-distance teleport steps to reduce cybersickness. This technique is applied to both joystick-based navigation and our novel hand-motion-based *PushPull* technique, that correlates proprioception to virtual movement. All introduced techniques are compared in a multi-stage user study and are bundled in a toolkit that is publicly available via [114].

The remainder of this chapter is organized as follows. Section 10.2 gives an overview of research in avatar visualization and locomotion before setting the current work in context. Section 10.3 presents the proposed avatar visualization and locomotion techniques and provides implementation details. The proposed techniques are compared in a multi-stage user study, which is introduced in Section 10.4, before separately explaining the evaluation of the avatar visualization techniques in Section 10.5 and that of the locomotion techniques in Section 10.6. The chapter concludes with a summary as well as possible future directions in Section 10.7.

To summarize, the contributions of this chapter are as follows:

- *PushPull*: a hand motion-based navigation technique with a dynamic velocity multiplier
- *Stuttered locomotion*: teleport steps for joystick-based navigation and *PushPull*
- *Smart Avatars*: smooth visualizations of continuous and noncontinuous VR locomotion
- A study to compare *Smart Avatar* visualizations from an observer's viewpoint
- A study to investigate the effects of *Stuttered Locomotion*

10.2 Background

In MR, the choice of locomotion has great impact on the user experience and interaction paradigms in general. At the same time, the visual representation of any locomotion technique is usually bound to the avatar of the moving user. The drawback of many avatar animation systems, however, is a lack of plausible locomotion visualizations for those human representations. This circumstance and further related work on locomotion and avatar animation systems has been discussed in-depth in Chapters 2, 8 and 9.

Farmani and Teather [99] presented discrete viewpoint control methods, which either rotate the user's viewpoint by a fixed degree, or translate the user's viewpoint by a fixed distance. The translations are limited to forward/backward movements of each 1 meter, which are triggered through mouse buttons. Their study results suggest that both discrete rotations and discrete lateral movements caused significantly lower cybersickness symptoms than continuous locomotion. Similarly, Adhikari et al. [3, 4] incorporated teleportation into a continuous locomotion technique (controller- or leaning-based). While the user controls simulated self-motion continuously using a given interface, once locomotion speeds exceed a threshold where cybersickness becomes likely, teleportation steps are added periodically in the direction of the continuous locomotion, with teleportation distance increasing with increasing joystick deflection or leaning. They reported similar levels of spatial awareness between continuous and non-continuous movement techniques.

As for the related work on avatar visualization, many works on locomotion techniques primarily focus on the experience of a single user, in this case, the moving user. In a multi-user environment, however, it is equally important to make a user's movements transparent to observers. Therefore, our goal was to combine the positive effects of continuous locomotion, which can be easily tracked by observers, with the previously discussed benefits of teleportation, which has been shown to reduce cybersickness. Our resulting technique is called *Stuttered Locomotion* and will be presented in detail in Section 10.3.2.

10.3 Technique Description

In this section we elaborate on the design and development of the proposed avatar visualization and locomotion techniques' main components. Our goals were twofold: (i) create an avatar system that provides realistic human-like movement visualizations for noncontinuous locomotion and (ii) provide a cybersickness-reducing mode for continuous locomotion techniques through short-distance teleport steps. All implementations are available on [114].

10.3.1 Smart Avatars - Masking Noncontinuous Movements

For noncontinuous locomotion, such as teleportation, conventional avatars would disappear from the initial location and reappear in the target location, which can be a confusing experience for observers in shared MR spaces. To mask these noncontinuous movements, we developed a novel avatar system that relies on IVAs for continuous human representations. The core idea is to assign an IVA to a target user and give them the objective to imitate their assigned user's tracked real world movement when close-by, and to autonomously navigate to their user when the distance is too long, i.e., when the user teleports.

10.3.1.1 Basic Implementation

To make an avatar's movement visually as smooth and natural as possible, the locomotion animation has to be perfectly synchronized to the actual translation and rotation of the avatar. Otherwise, the avatar feet will visually slide over the ground's surface which risks breaking the illusion of looking at an actual human being. To achieve this visual synchronization, we utilized Unity's pathfinding framework 'NavMesh' and their animation rigging system 'Mecanim'. NavMesh defines both walkable surfaces in a given scene and agents which can autonomously navigate and move on these surfaces. These agents are often used to implement AI controlled characters and require a single 3D world-coordinate as input to perform a pathfinding sequence. There are also options for off-mesh links to cross gaps between walkable surfaces that require custom behaviour, such as jumping between isolated surfaces or climbing ladders.

Smart Avatars can be applied to any animation controller, simple or complex, humanoid or non-humanoid, as long as that controller is compatible with autonomously navigating agents and IK passes. For our avatar agents we used Unity's standard assets character controller [316] and a set of root transformation animations obtained from Mixamo [5], which will move the avatar entity through 3D space as opposed to merely transforming its assigned mesh. Rather than having NavMesh move the entity and blending an animation that matches the movement, we take the desired velocity of the NavMesh agent in each frame and interpret this value as input for the Mecanim animator. Based on this input the animator blends between several animations such as walking forward, leaning and turning. The result is a root animation that will move the entity in such a way that the visible feet and the entity movement perfectly match. To summarize, the NavMesh agent determines the required movement, and Mecanim generates a root animation that will satisfy the desired lateral and angular velocity. This procedure results in smoothly and believably animated agents. The pathfinding target is constantly updated with the VR camera rig's position, letting the avatar smoothly follow their assigned user.

While the avatar is close to their user, it should imitate their real-world movements and posture. This requires the avatar to be turned in the same direction as the user, which is achieved through the combination of a gimbal lock and a second set of animations, primarily consisting of strafe-movements. We found that a radius of 2 meters around the camera rig works well as strafing zone. The gimbal lock forces the avatar to be rotated the same as the virtual camera on the X/Z plane. It is engaged over time to co-incite with the animation blending of the animator blend tree. When the walking avatar approaches the position of their user from a medium distance it will quickly, but not abruptly, turn to view the same direction once the threshold is passed. Inside of this strafing zone we defined an even smaller zone of 0.5 meters in which we blend in an IK pass to the animator.

Thus, when the avatar and user position overlap substantially, it begins to imitate the user's head and hand movements.

10.3.1.2 Transition Types

Without applying any special effects, the previously described implementation results in a continuous

Walking animation of the smart avatar whenever its assigned user is teleporting. The limitation of this technique is that the avatar locomotion only looks realistic up to a certain movement speed. By limiting the maximum possible movement speed, there can be a considerable discrepancy between avatar and user position when teleporting over long distances. This could potentially lead to unintentional or inconsistent behaviour and interactions between users. To minimize this positional discrepancy, we developed three long-distance travel visualizations for *Smart Avatars*: *Afterimage*, *Dissolve* and *Foresight*. When the distance between user and avatar is greater than a predefined threshold, a long-distance travel sequence is initiated to quickly realign avatar and user.

Afterimage (cf. Fig. 10.1 A) increases the maximum movement speed of the agent tenfold during long-distance travel. The avatar will additionally leave a trail of ghost-like afterimages that allows observers to visually follow and retrace the path. In order to create a ghost-like image, we bake the avatar's skinned mesh renderer in its current configuration, apply the resulting mesh to a new entity, and add a transparent texture with fresnel effect that fades out over time.

Dissolve (cf. Fig. 10.1 B) likewise creates a copy of the avatar mesh when the sequence is initiated. Unlike this copy, the original avatar is teleported together with the user and made invisible. Both original and copy use a dissolve shader with the same noise map; while the original avatar appears over time, the copy disappears over time. Concurrently a particle system depicts a stream of matter from the initial to the current avatar location. This gives the impression that the avatar is de- and reconstructed.

Foresight (cf. Fig. 10.1 C) uses three agent avatars at the same time for one user: (i) A ghost-like avatar that is teleported with the user, depicting their current position and imitating their pose without delay. (ii) An invisible avatar with the same increased movement speed as the *Afterimage* avatar, which leaves a trail of fading ghost-like images. And (iii) a solid avatar which runs towards the user with default speed. The trail images are timed in such a way that they disappear once the *Smart Avatar* passes through them. When triggered, the effect looks like a normal moving user, but with additional ghost-like projections of the avatar's imminent path.

10.3.2 Stuttered Locomotion - Mapping Continuous Input To Noncontinuous Movements

Numerous cybersickness-reducing techniques for VR locomotion are concerned with the user's viewport, for example by limiting their FOV or dynamically blurring their vision [12, 101]. At the same time, noncontinuous locomotion techniques such as teleportation regularly receive favorable cybersickness ratings when compared to continuous movements like joystick-based locomotion [54, 104]. A commonality among these approaches is the limitation of optical flow, which itself was closely related to the level of elicited cybersickness by multiple authors (for an overview, see [61]). The approach of reducing optical flow also seems to be applied in several commercial applications, such as the home menus of Steam VR and Oculus, which use discrete rotation steps for joystick-based turning mechanisms [99]. From these observations, we conclude that a possible solution to limit the optical flow in continuous locomotion is to decompose a given translation into a series of discrete teleport steps, thus converting a continuous movement into a noncontinuous one. We call this technique *Stuttered Locomotion*. *Stuttered Locomotion* can be applied to any continuous motion control method, which we will exemplify with two of these methods, *Joystick* and *PushPull*. In the following, we will first introduce the techniques without the application of *Stuttered Locomotion* before describing the necessary implementational changes for integrating *Stuttered Locomotion*.

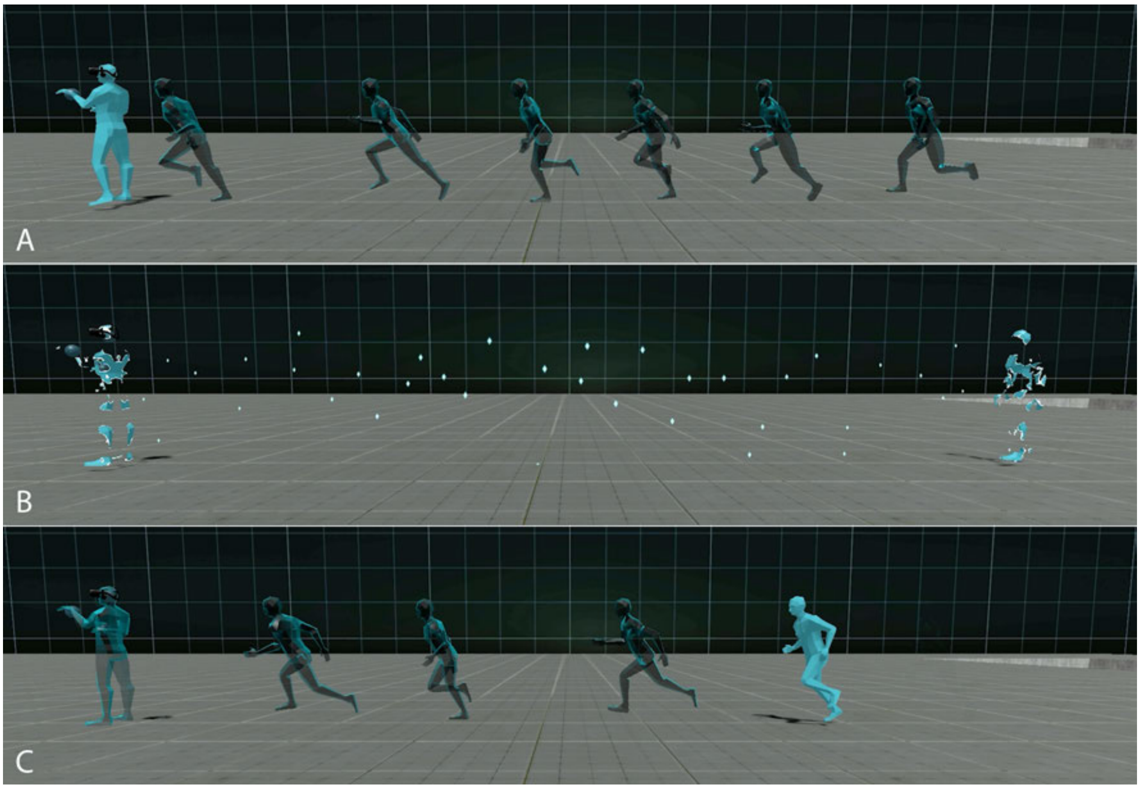


Figure 10.1: The long-distance transition types as seen in the user study's VE: Afterimage (A), Dissolve (B) and Foresight (C).

Smooth Joystick locomotion was included in both the toolkit and the empirical comparison described in Section 10.6, as it can be used in regular VR systems without the need for special sensors, and because it is optimal for observers in shared VEs (due to the continuous movements). In a literature review of 2017, [37] identified controller/joystick as one of the two prevalent continuous locomotion techniques in VR research (besides walking-in-place). The user tilts the joystick on a controller to move in the VE, with the degree of tilt linearly determining the movement speed.

Stuttered Joystick locomotion extends the previous motion control by executing a teleport step with a specified step length on the first frame the left joystick is tilted. While the tilt is continued, another teleport step is triggered each time an internal countdown has passed. This countdown can be divided through an arbitrary user selected multiplier or be made dependent on the degree of joystick tilt. We tested several configurations for step length and fixed/dynamic movement multipliers, and derived a number of presets ranging from a focus on efficiency to maximizing comfort. As part of the provided toolkit, these settings for velocity multipliers, step length, turning steps (i.e., discrete user rotations by a fixed angle), etc. can then be fine-tuned to personal preference by the user in a menu.

Smooth PushPull is a hand motion-based locomotion technique with a dynamic velocity multiplier. It is built upon the Anchor Turning technique described in Chapter 9, which, from the user's perspective, lets them drag-and-drop the virtual world around them. In Anchor Turning, the virtual camera acts as center point of the rotation. While a drag-and-drop motion is performed, the user is rotated in such a way that the dragging hand stays in a fixed position relative to the virtual world.

For *PushPull*, this technique's principle has been transferred to lateral movements. While the right hand performs the rotations as described above, the left hand translates the user in a similar manner. When a drag motion is initiated, the left hand acts as a fix point relative to the world, allowing the user to pull themselves through space. This metaphor has been described as world-drag before [325]. The required movement is achieved by saving the initial coordinates of the dragging hand in the user's local reference frame and comparing this value to the same hand's current position. We project this vector onto the plane that is defined by an upwards pointing normal vector to constrain movements to the XZ-plane (in a left-handed coordinate system). The user's position is then set to its initial position when the drag motion began, offset by the projected vector. As both rotation and translation are calculated in the user's local reference frame, they are independent from each other and can be employed at the same time.

This approach results in fine movements, which is desirable in terms of achieving a high precision, but unsuitable to cover medium to long distances. We therefore implemented a dynamic velocity multiplier based on the hand's position on the Y-axis in the user's reference frame. When the hand is held at chest height or above, the dynamic multiplier is set to 1, resulting in fine control. When the hand is held at hip height or lower, a maximum multiplier of 4, which was determined to result in comfortable and efficient movements in a pilot study, is assumed. Corresponding multipliers for in-between heights are calculated through linear interpolation. This dynamic velocity multiplier brings a noteworthy implication: when the user performs an arm-swing motion starting in front of their body and ending behind their back, their velocity over the movement's duration resembles a parabola with its peak at the halfway point. Essentially, an arm-swing will result in a short visual dash that eases in and out of the motion. In summary, *PushPull* is a continuous motion control technique that allows users to be highly flexible with movement options that range from precision to efficiency and dash-like motions.

Stuttered PushPull behavior is an extension of the previously described behavior, which requires to add a dead zone around the initial drag position. Once a certain threshold of aggregated hand movement has been exceeded, a teleportation is initiated with a distance equal to the threshold's value, which we will subsequently refer to as *step length*. If the vector from initial to current hand position exceeds an exemplary step length of 25cm, a teleport step in the vector's direction is initiated to a point 25cm away from the user's current position. Then, the process resets and the current hand position is assumed to be the new initial position. The resulting movement of a user performing an arm-swing motion with *Stuttered PushPull* thus is a series of short-distance teleport steps that cover the same total distance as the same motion with *Smooth PushPull*. The dynamic velocity multiplier described in the previous section also applies to *Stuttered PushPull*. However, it had to be implemented in such a way that the step length would remain consistent with the selected user settings. This can be achieved by dividing the required threshold through the velocity multiplier, as opposed to multiplying the travel distance.

10.4 User Study Overview

To assess the effectiveness of the introduced techniques, we conducted a multi-stage user study. The study consisted of an observation section (to compare the avatar visualization techniques) and a first-person interaction task (to compare the locomotion techniques). The observer and first-person segments will be described separately in the following two sections.

10.4.1 Participants and Apparatus

24 participants (12 male, 12 female) took part in the study. Age was specified in brackets to comply with general data protection regulation. 12.5% of participants were in the age group of 18 - 24, 58.3% in 25 - 34, 16.6% in 35 - 44, and each 4% in the groups of 45 - 54, 55 - 64 and 65 - 74. The

mean time per participant was about 20 minutes per study segment, adding up to 40 minutes for the entire two-stage study. Unity3D was used to render the scene on an Oculus Quest 2. The full study was conducted remotely through the Remote Study Framework as described in Section 4.1, with detailed instructions for the participants regarding the procedure and requirements for the study environment.

10.5 Observer Study

The observer part of the user study explores the acceptance and perception of our four *Smart Avatar* transition types, *Walking*, *Afterimage*, *Dissolve*, and *Foresight*. It simulates a shared VE scenario in which a user (represented by their avatar) performs teleportation while the study participant assumes the role of the observer. To validate that the advantages of continuous over noncontinuous locomotion demonstrated in Chapter 8 still apply to *Smart Avatars*, we included a *Primitive* condition as a baseline, involving a conventional avatar that disappears and reappears during teleportation. This conventional avatar consisted of a hovering capsule body with detached hands and head, as this is a common human representation that can be found in commercial products such as Oculus Home. For the other four levels, we used a low-poly humanoid mesh with natural human locomotion animations, enabled by the *Smart Avatar* technique. Both types were monocolored. A joystick-controlled full-body avatar was not included in the comparison as it is, from an observer's perspective, indistinguishable from the *Smart Avatar Walking* transition. As the study was already fairly long, we did not include additional conditions that could help to further tease apart the relative contributions of avatar fidelity versus transition types in future studies, such as applying *Afterimage*, *Dissolve*, and *Foresight* to a conventional (capsule body) avatar, or including a *Smart Avatar* condition without any transitions (thus directly teleporting).

10.5.1 Measures

Of particular interest was the observer's spatial awareness of the avatar user's position and movement intentions. We expected the default teleportation visualization, where the avatar instantly moves between locations, to be perceived as difficult to anticipate and to visually track. Our custom spatial awareness questionnaire consisted of three 7-point Likert scales ranging from 1 to 7, asking the participants how strongly they agree with the following statements:

- "I had a sense of where the avatar's user currently is."
- "I had a sense of where the avatar's user intends to move."
- "I could reliably point the laser pointer at the avatar."

These values were aggregated in our analysis as the *Spatial Awareness* score.

In addition to spatial awareness, we wanted to investigate whether or not transitions with special effects, such as *Foresight*, have an effect on the perceived pragmatic and hedonic qualities of the avatar. We expected that the applied visual effects would reduce the humanness of an avatar, but make them visually intriguing and easier to predict and track. To measure the *Pragmatic* and *Hedonic* qualities of the visualizations we employed an English version of the AttrakDiff questionnaire [140, 318]. It uses 7-point Likert scales ranging from -3 to +3, asking participants if they would describe the technique rather with the left term (e.g., *unpredictable*, *isolating*) or right term (e.g., *predictable*, *connective*) of 28 word pairs. There are four sub-scales, each consisting of 7 word pairs - *Pragmatic Quality* (PQ), *Hedonic Quality - Identity* (HQ-I), *Hedonic Quality - Stimulation* (HQ-S) and *Attractiveness* (ATT). Here, *Pragmatic Quality* measures usability while *Hedonic Quality* measures emotional reactions. Additionally, participants were asked to rate their desire to use each visualization technique in a real VR application on a 7-point Likert scale.

	PQ		HQ-I		HQ-S		ATT		Sp. Awareness		Use again	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Primitive	-.37	1.34	-.94	1.22	-1.07	1.21	-.80	1.33	3.76	1.25	2.88	1.51
Walking	2.07	.96	1.30	.88	-.09	1.18	1.40	.90	5.79	1.00	6.17	.94
Afterimage	.31	1.25	.96	.94	1.42	.86	.82	1.24	5.18	1.02	5.00	1.58
Dissolve	1.02	1.15	1.36	.80	1.19	.79	1.45	.97	5.89	.68	6.00	1.08
Foresight	.50	1.35	1.26	.90	1.87	.52	1.05	1.25	5.60	1.10	5.13	1.86

Table 10.1: Means and standard deviations for all measures in the observer study.

10.5.2 Hypotheses

Based on the above described criteria, the study was designed as within-subject, and the following hypotheses were formed:

H_1: Continuous avatar transition techniques are rated higher in spatial awareness than noncontinuous transition techniques.

H_2: The *Primitive* transition is rated lowest in *Pragmatic* and *Hedonic* qualities.

H_3: The *Walking* transition appears the most natural and is thus rated highest in *Pragmatic Quality* and *Hedonic Quality - Identity*.

H_4: The transitions including special effects (i.e., *Afterimage*, *Dissolve*, and *Foresight*) are rated highest in *Hedonic Quality - Stimulation*.

10.5.3 Stimuli and Procedure

In the beginning of a study session, participants filled in a demographic questionnaire and gave their informed consent to participation, followed by a briefing about the nature of the task and the different techniques.

The participant wore an Oculus Quest 2 HMD, and was positioned in a mostly empty VE. In each trial, participants were tasked to observe an avatar moving in a figure eight in front of them. By having the avatar walk in this pattern, the locomotion can be observed from all sides at different distances. The figure eight had dimensions of 50 by 10 meters and was oriented orthogonal to the participant, who was standing at a distance of 10 meters from the figure's center point. The shown avatar animation was a pre-recorded real user sequentially teleporting to 11 points along the figure eight path, over the course of roughly 40 seconds. Participants held an Oculus controller in their preferred hand that was represented virtually as a laser pointer, which they were asked to point at the avatar at all times. The laser's color was green while intersecting the avatar, and red when not. Participants were not embodied beyond visible hands and the attached laser pointer.

There was one trial for each of the five tested transition types. The order of trials was randomized by Latin square to counteract learning effects. After each trial, participants had to rate their experience by filling in the questionnaires described in Section 10.5.1. Then, after all trials had concluded, participants were further asked to indicate their preferred avatar visualization.

10.5.4 Results

In this section the results of the statistical analysis are presented. When the Shapiro-Wilk test indicated a normal distribution of residuals, a repeated-measure ANOVA and post-hoc pairwise comparisons with Bonferroni adjustment were used to test for differences between conditions. Otherwise, the Friedman test and Wilcoxon signed-rank tests with Bonferroni correction were used. Mauchly's test did not indicate a violation of the assumption of sphericity for any of the variables. A 5% significance level was assumed. All aggregated questionnaire results are listed in Table 10.5.3 and illustrated in Figure 10.2.

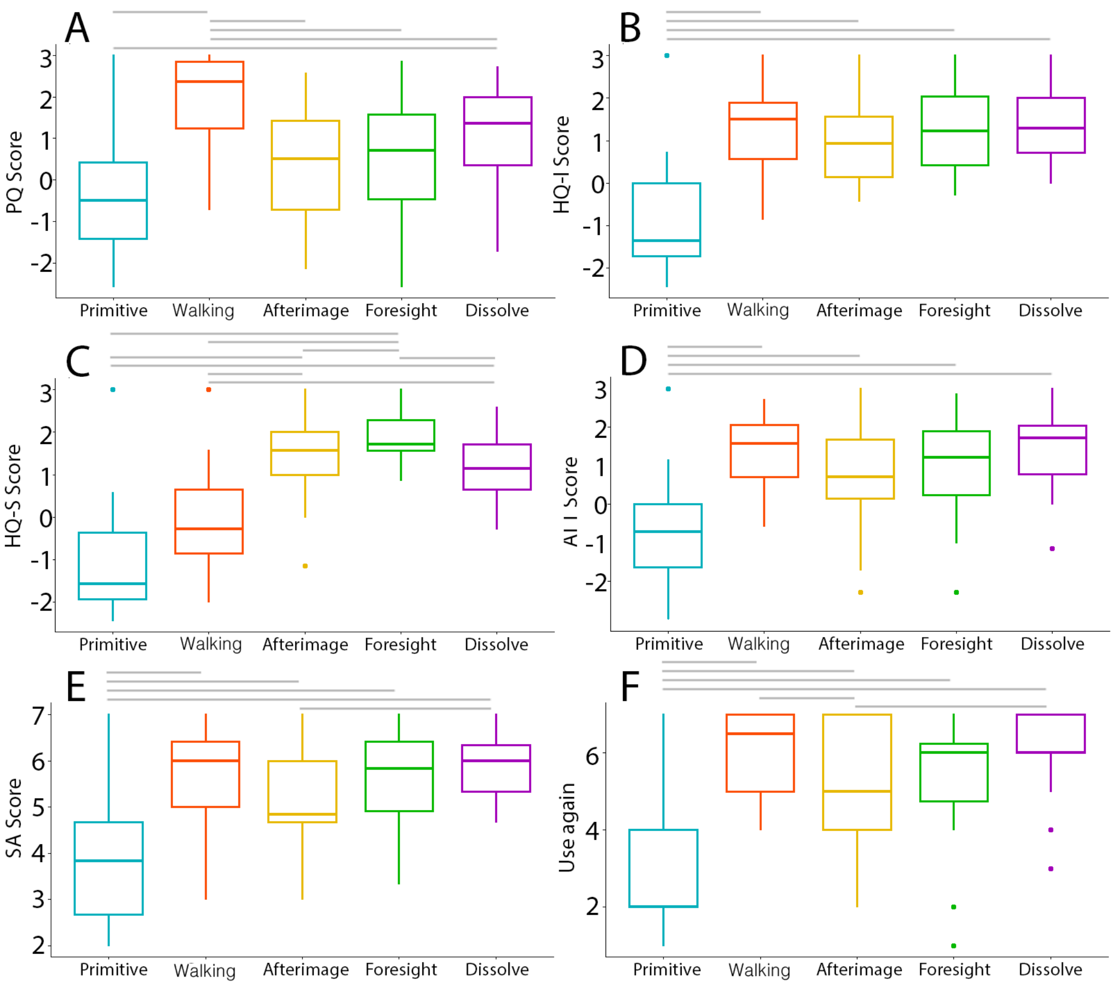


Figure 10.2: Boxplots of the observer study results: Pragmatic Quality (A), Hedonic Quality - Identity (B), Hedonic Quality - Stimulation (C), Attractiveness (D), Spatial Awareness (E), and Desire to use again (F). Boxplots indicate means, quartiles, min/max, and outliers. The gray horizontal lines indicate significant differences between conditions in the order of occurrence in Section 10.5.4.

10.5.4.1 Pragmatic Quality - cf. Fig. 10.2 A

The ANOVA revealed significant differences between the transition types ($F(4,92) = 15.783, p < 0.001, \eta_p^2 = 0.407$). The *Walking* transition was rated significantly higher than *Primitive* ($p < 0.001$), *Afterimage* ($p < 0.001$) and *Foresight* ($p = 0.001$) as well as *Dissolve* ($p = 0.032$). Additionally, *Dissolve* was rated higher than *Primitive* ($p = 0.001$). The depiction of human gait without artificial effects was therefore rated as the most natural and clearly structured visualization type. By virtue of not having a fully humanoid shape, it was unsurprising that the primitive avatar received low ratings for humanness. This, however, is only one item of seven in the *Pragmatic Quality* category.

10.5.4.2 Hedonic Quality - Identity - cf. Fig. 10.2 B

A Friedman test revealed a significant effect of transition type on *Hedonic Quality - Identity* ($\chi^2(4) = 41.624, p < 0.001$) A post-hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.01$. The HQ-I score was significantly lower for the *Primitive* condition in comparison to all other transition types, i.e., *Walking* ($Z = -4.138, p < 0.001$), *Afterimage* ($Z = -4.122, p < 0.001$), *Dissolve* ($Z = -4.198, p < 0.001$) and *Foresight* ($Z = -3.916, p < 0.001$). All variations of the *Smart Avatar* system were rated similarly, highlighting the integrating and connecting properties of naturally and continuously moving full-body avatars.

10.5.4.3 Hedonic Quality - Stimulation - cf. Fig. 10.2 C

A Friedman test revealed significant differences between the transition types ($\chi^2(4) = 65.502, p < 0.001$). Again, we Bonferroni-adjusted the significance level to 0.01 and performed multiple Wilcoxon signed-rank tests. *Foresight* was rated significantly better than all other transitions, i.e., *Primitive* ($Z = -4.173, p < 0.001$), *Walking* ($Z = -3.958, p < 0.001$), *Afterimage* ($Z = -2.994, p = 0.003$) and *Dissolve* ($Z = -2.924, p = 0.003$). In addition, *Primitive* was rated significantly lower than *Afterimage* ($Z = -4.198, p < 0.001$) and *Dissolve* ($Z = -4.138, p < 0.001$). Also, *Walking* was rated significantly lower than *Afterimage* ($Z = -4.199, p < 0.001$) and *Dissolve* ($Z = -3.833, p < 0.001$). The *Foresight* technique was rated highest for the characteristics of *inventive*, *innovative*, *creative* and *captivating*.

10.5.4.4 Attractiveness - cf. Fig. 10.2 D

An ANOVA indicated a significant effect of transition type on *Attractiveness* ($F(4,92) = 18.742, p < 0.001, \eta_p^2 = 0.449$). The ATT score was significantly lower for the *Primitive* condition in comparison to all other transition types ($p < 0.001$). All variations of the *Smart Avatar* system were rated similarly, with the same ranking as for the HQ-I score. Special effects had little influence on the ATT score, signaling that the attractiveness of the locomotion visualization is again directly linked to the continuously moving full-body avatar.

10.5.4.5 Spatial Awareness - cf. Fig. 10.2 E

Regarding *Spatial Awareness*, an ANOVA revealed significant differences between transition types ($F(4,92) = 20.916, p < 0.001, \eta_p^2 = 0.476$). The *Spatial Awareness* score was significantly lower for the *Primitive* condition in comparison to all other transition types ($p < 0.001$). Furthermore, *Afterimage* performed significantly worse than *Dissolve* ($p = 0.011$). There were no other significant differences between the *Smart Avatar* variations. This confirms the findings of Chapter8, showing that continuous locomotion is generally easier to interpret and anticipate than noncontinuous locomotion.

10.5.4.6 Desire to use again - cf. Fig. 10.2 F

After each trial of the observation experiment, participants were asked to rate their desire to use the depicted visualization in a shared MR space with real users on a scale of 1 to 7. A Friedman test revealed significant differences between transitions ($\chi^2(4) = 44.230, p < 0.001$). According to the post-hoc analysis, the *Primitive* visualization was rated significantly lower than all variations of *Smart Avatars*, i.e., *Walking* ($Z = -4.049, p < 0.001$), *Afterimage* ($Z = -3.679, p < 0.001$), *Dissolve* ($Z = -4.126, p < 0.001$) and *Foresight* ($Z = -3.265, p = 0.001$). Additionally, *Afterimage* got significantly lower scores than *Walking* ($Z = -2.987, p = 0.003$) and Dissolve ($Z = -2.857, p = 0.004$). The *Walking* transition received the highest score, however, *Foresight* was rated most often as the favorite avatar type with 37.5% of votes. *Dissolve* received 33.33% of votes, followed by *Walking* with 25%, *Afterimage* with 4.16% and *Primitive* with 0% of votes.

10.5.5 Discussion

In this section we discuss the results of the observer study with regard to the formed hypotheses.

10.5.5.1 Higher Spatial Awareness for Continuous Than for Noncontinuous Transitions (H_1)

The *Primitive* transition received the lowest *Spatial Awareness* score, while *Walking* was rated second highest and significantly better than *Primitive*, confirming hypothesis H_1. This result suggests that the observer's *Spatial Awareness* score is not positively correlated with the relative time an avatar is displayed at the user's actual position, as this would be highest for *Primitive* avatars. Therefore, although the displacement between user and *Smart Avatar* can be substantial for long-distance teleports, the continuous locomotion visualization possibly (falsely) convinced the observers that the user is actually located where the avatar is depicted. Moreover, the *Dissolve* technique visualizes both start and end point of a long-distance travel simultaneously, while the particle flow implies quick movement, therefore providing users with additional cues to accurately predict the avatar's movement.

10.5.5.2 Lowest Pragmatic and Hedonic Quality for Primitive Transition (H_2)

As expected, the *Primitive* transition received the lowest mean ratings for both *Pragmatic* and *Hedonic Quality*, confirming hypothesis H_2. Additionally, it received significantly lower *Attractiveness* scores, making it the consistently lowest rated condition in all measures. It could be argued that the avatar appearance was a major contributing factor to these differences, as it was not identical to the *Smart Avatar* variations. While similar in color, dimensions and overall complexity, the *Primitive* avatar had no legs. However, having smoothly animated legs without the need for tracking devices should not be considered an unfair advantage of the *Smart Avatar* conditions, but one of the key technical contributions that set them apart from conventional VR avatars by design. A comparably animated full-body avatar for the baseline condition would itself have required the application of a technique such as *Smart Avatars*, thus missing the goal of being a baseline to compare the newly created system to. Moreover, the findings of Chapter 8 suggest that a likely cause of the lower *Attractiveness* scores of the *Primitive* avatar is not its appearance, but the lack of motion animation. Nonetheless, we are interested in conducting a similar experiment with different avatar visualizations in the future.

Among the *Smart Avatar* transitions, *Afterimage* performed the worst. From these conditions, it scored lowest in *Spatial Awareness*, *Pragmatic Quality*, *Hedonic Quality - Identity* and *Attractiveness*. In essence, it is conceptually similar to *Foresight* but inferior in every tested measure. We therefore cannot recommend using this technique.

10.5.5.3 Highest Pragmatic Quality and Hedonic Quality - Identity for Walking Transition (H_3)

As hypothesized, *Walking* received significantly higher *Pragmatic Quality* scores, partially confirming hypothesis H_3. Contrary to hypothesis H_3, *Dissolve* received the top scores for both *Hedonic Quality - Identity* and *Attractiveness*, however without being significantly better than the other *Smart Avatar* transitions. The graphically elaborate nature of this technique lead to a higher score in presentability. In terms of AttrakDiff word pairs, it was perceived to be *professional*, *stylish* and *premium* in the *Hedonic Quality - Identity* category, and *pleasant*, *appealing* and *motivating* in *Attractiveness*.

10.5.5.4 Higher Hedonic Quality - Stimulation for Transitions With Special Effects (H_4)

Our interpretation of the subjective preferences between conditions is that *Foresight* was the most memorable technique, which is supported by receiving the highest scores in the *Hedonic Quality - Stimulation* category of the AttrakDiff questionnaire. In contrast, the *Walking* transition received the highest mean rating in *Pragmatic Quality* - a measure of humanness and overall usability in terms of being clearly structured and predictable. In fact, the order of mean scores between *Pragmatic Quality* and desire to reuse a technique were identical. This strongly suggests that there is a correlation between these two measures. Overall, *Foresight*, *Afterimage* and *Dissolve* were the strongest transitions for *Hedonic Quality - Stimulation*, confirming hypothesis H_4.

The results conclusively show that the *Smart Avatar* system and the human representations it enables are better suited for shared MR spaces than conventional *Primitive Avatars* from an observer's perspective. It should be pointed out that it is likely that *Smart Avatars* would be rated similarly to other full-body avatar animation systems during continuous locomotion. Based on the *Spatial Awareness* results, it stands to reason that observers would not be able to differentiate between those systems and *Smart Avatars*. The true benefit of using *Smart Avatars* is that the discussed qualities can also be applied to visualizations of noncontinuous locomotion.

10.6 Interaction Study

In this section, we describe the locomotion part of the study. This part was a direct follow-up to the observation study and was thus conducted with the same participants and likewise as within-subject design. To validate whether *Stuttered Locomotion* indeed reduces the occurrence of cybersickness-related symptoms, we compared it to smooth locomotion for both joystick-based input and our motion-based *PushPull* technique. Locomotion parameters were chosen to result in comparable maximum movement speeds between *PushPull* and *Joystick* and to reflect average human sprinting. Based on a pilot study, the maximum *PushPull* velocity multiplier was set to 4, and the maximum Joystick movement speed was 2.5 meters per second (9 km/h). During the *stuttered* conditions, the step length for both *PushPull* and *Joystick* was set to 0.5 meters and the turning step to 30°. *Stuttered Locomotion* works independently of the movement speed as the step length is fixed.

10.6.1 Measures

To inquire the perceived *Efficiency* and *Pleasantness* of the compared techniques, we used one 7-point Likert scale, ranging from 1 to 7, each.

We further employed a modified *Embodiment* questionnaire that covers the dimensions *Body Ownership*, *Agency and Motor Control*, and *Location of the Body* of the Avatar Embodiment Questionnaire of Gonzales et al. [118]. The individual questionnaire items are listed in Table 10.3. Each item was a 7-point Likert scale ranging from 1 to 7, and items 6 and 8 were inverted when calculating the aggregated *Embodiment* score.

Cybersickness was rated with a condensed form of the SSQ [163]. Similar to [101], it consisted of one item for each of the three SSQ sub-scales (nausea, oculomotor and disorientation). We gave three examples of individual SSQ items belonging to those sub-scales. Every item was represented by a 5-point Likert scale ranging from 0 (not affected) to 4 (strongly affected).

Presence was rated using the SUSP [280], and *Usability* was rated through the SUS [47].

10.6.2 Hypotheses

Based on the above described criteria the following hypotheses were formed:

H_5: *PushPull*'s hand-based movement is directly linked to the user's proprioception and thus receives lower *Cybersickness* scores than *Joystick*.

H_6: *PushPull* is physically more demanding than *Joystick*, leading to lower *Usability*.

H_7: *Stuttered Locomotion* has less optical flow and will receive lower *Cybersickness* scores than *Smooth Locomotion*.

H_8: *Stuttered Locomotion* is unnatural and will be perceived as less *Pleasant* than *Smooth Locomotion*.

10.6.3 Stimuli and Procedure

This experiment directly followed the observation task. At this point participants were reminded they could take breaks between sub-sections. An initial locomotion tutorial prepared the participants to use the *Joystick* and *PushPull* methods by having them reach certain target areas in their vicinity.

After the tutorial concluded, participants were asked to indicate their current level of cybersickness by filling in the condensed SSQ. During the experiment, the participants' task was to find and slice oversized fruits with a sword, and afterwards return to their starting position. There were three sets of three fruits each, arranged in such a way that the participant had to move forward to

	Efficiency		Pleasantness		Embodiment		Cybersickness		Presence		Usability	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Smooth Joystick	4.71	1.90	3.92	2.15	5.45	1.13	10.29	8.91	2.92	1.11	75.52	17.55
Stuttered Joystick	4.13	1.57	4.21	1.56	5.22	1.05	-3.12	8.30	3.17	1.08	74.79	17.37
Smooth PushPull	4.21	1.67	4.29	1.37	5.51	1.07	.16	5.89	3.22	1.26	66.04	16.23
Stuttered PushPull	3.46	1.59	4.38	1.97	5.20	1.27	-3.90	8.21	3.25	1.13	56.35	24.25

Table 10.2: Means and standard deviations for all measures in the interaction study.

reach the first set, turn approximately 90° left to find the second set, turn approximately 180° to see the third set, and finally turn approximately 90° right to return to their highlighted starting point. The next set of fruits was only activated when the previous one was successfully cut. The distance between the starting point and each set of fruits was 5 meters. A black vignette from *VR Tunneling Pro* [274] with default values was faded in during movement or rotation regardless of locomotion technique, which is in line with current state-of-the-art VR experiences [8]. In each trial only one combination of input method (*Joystick* / *PushPull*) and stuttering (*Smooth* / *Stuttered*) was available, while the order of trials was randomized by Latin square. After each trial, participants were asked to fill in the questionnaires described in Section 10.6.1.

10.6.4 Results

We considered the two factors *Input Method* (*Joystick* / *PushPull*) and *Stuttering* (*Smooth* / *Stuttered*), and therefore performed a two-way ANOVA with a 5% significance level for each of the dependent variables. For *Pleasantness* and *Embodiment*, an inspection of both histograms and Q-Q plots showed a small negative skewness (-0.280 and -0.510, respectively) and kurtosis (-0.701 and -0.448, respectively) of the residuals, however, the ANOVA has been shown to be robust against such mild deviations from the normal distribution [221]. All questionnaire results are listed in Table 10.6.3 and illustrated in Figure 10.3.

10.6.4.1 Efficiency - cf. Fig. 10.3 A

A two-way ANOVA revealed a significant main effect of *Stuttering* on self-reported *Efficiency* ($F(1,23) = 7.472, p = 0.012, \eta_p^2 = 0.245$). *Stuttered* locomotion ($M = 3.79, SD = 1.60$) was perceived as significantly less efficient than *Smooth* locomotion ($M = 4.46, SD = 1.79$). We did not find a significant difference between *Joystick* and *PushPull* locomotion.

10.6.4.2 Pleasantness - cf. Fig. 10.3 B

No significant main or interaction effects of *Input Method* and *Stuttering* on *Pleasantness* were found.

10.6.4.3 Embodiment - cf. Fig. 10.3 C

As for *Pleasantness*, we also did not find any significant effects of the two factors on *Embodiment*.

10.6.4.4 Cybersickness - cf. Fig. 10.3 D

For the SSQ, we compared the differences from the questionnaire results gathered before and after every trial (POST-PRE). Therefore, positive results indicate an increase in *Cybersickness*, while negative results indicate a decrease or return to normal upon usage of the method. A two-way ANOVA revealed a significant spreading interaction between *Input Method* and *Stuttering* ($F(1,23) = 10.147, p = 0.004, \eta_p^2 = 0.306$). We also found significant main effects of *Input Method* ($F(1,23) = 17.149, p < 0.001, \eta_p^2 = 0.427$) and *Stuttering* ($F(1,23) = 12.963, p = 0.002, \eta_p^2 = 0.360$) on *Cybersickness*, however, they only have limited conclusiveness in the presence of a significant interaction. We therefore performed a follow-up simple main effects analysis with Sidak-adjusted comparisons, which indicated that for *Joystick* input, *Cybersickness* was rated 13.402 points higher for *Smooth* locomotion than for *Stuttered* locomotion ($p < 0.001$). For *PushPull*

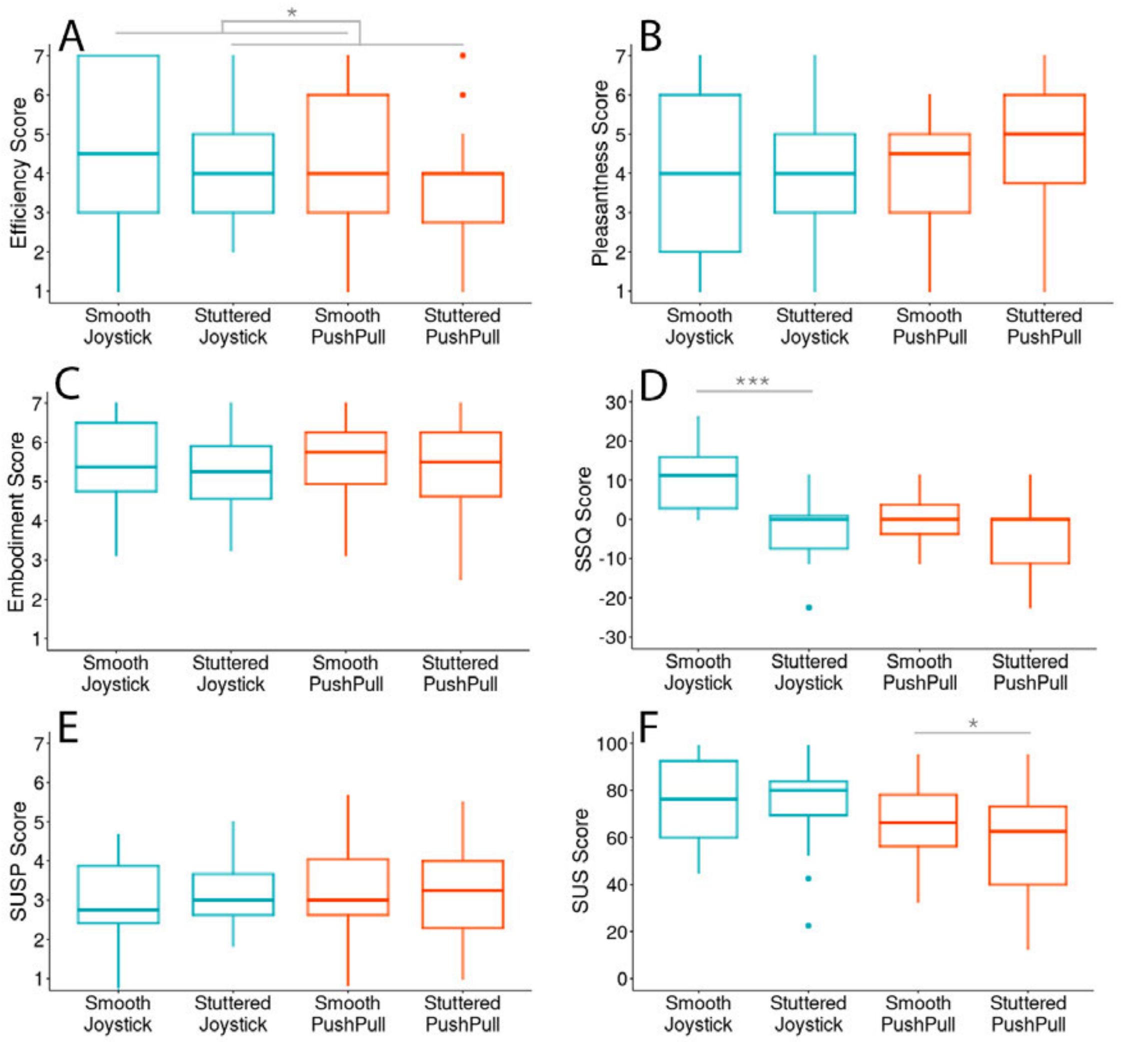


Figure 10.3: Boxplots of the interaction study results: Efficiency (A), Pleasantness (B), Embodiment (C), Cybersickness (D), Presence (E), and Usability (F). The gray horizontal lines indicate significant differences between conditions (for significant interactions, only the significant simple main effects are illustrated).

Code	Item
BO1	I caused the movement and cutting actions.
BO2	I performed the movement and cutting actions myself.
BO3	The visible avatar represented me.
BO4	I felt like I could control the avatar as if it was my own body.
BO5	I felt as if the avatar was my body.
BO6	I felt out of my body.
BO7	I felt as if my body was located where I saw the avatar.
BO8	I felt like the avatar was someone else.

Table 10.3: The *Body Ownership* questionnaire based on [118]. Participants were asked how strongly they agreed with these statements.

input, no significant difference between *Smooth* and *Stuttered* was found in the simple main effects analysis ($p = 0.108$).

10.6.4.5 Presence - cf. Fig. 10.3 E

For the self-reported sense of *Presence*, all conditions received similar ratings, and therefore no significant differences were found.

10.6.4.6 Usability - cf. Fig. 10.3 F

Regarding self-reported *Usability*, a two-way ANOVA revealed a significant main effect of *Input Method* ($F(1,23) = 8.659, p = 0.007, \eta_p^2 = 0.274$) as well as a significant interaction effect between *Input Method* and *Stuttering* ($F(1,23) = 6.386, p = 0.019, \eta_p^2 = 0.217$). We conducted a simple main effects analysis with Sidak adjustment and found a significant difference between *Smooth PushPull* and *Stuttered PushPull* ($p = 0.011$), with the latter scoring 9.687 points worse on average. For *Joystick* input, no significant simple main effect was found ($p = 0.864$).

10.6.5 Discussion

In this section we discuss the results of the interaction study with regard to the formed hypotheses.

10.6.5.1 Lower Cybersickness for PushPull Than for Joystick (H_5)

The analysis of the SSQ revealed that *PushPull* indeed caused fewer occurrences of cybersickness symptoms than *Joystick*, thus confirming hypothesis H_5. The major benefit of motion-based locomotion, the reduction of cybersickness, also applies to *PushPull*, making it an attractive choice for regular usage. This applies in particular to *PushPull* in combination with *Stuttered Locomotion*, as discussed in Section 10.6.5.3. For Joystick-based locomotion, the comparatively higher mean score for cybersickness was primarily due to the *Smooth* condition, while *Stuttered Joystick* received similar scores to both *PushPull* conditions.

10.6.5.2 Lower Usability for PushPull Than for Joystick (H_6)

PushPull's *Usability* score was significantly lower than that of *Joystick*, confirming hypothesis H_6. Participants reported that *PushPull* required more time to learn and master. As it is similar to the drag-and-drop metaphor, we expected it to be natural and easy to learn, which did not appear to be the case. A potential factor for this might be the fact that motion-based input was a novel concept to many participants. This had no effect on the perceived *Efficiency* of the technique, however.

At the same time, *PushPull* and *Joystick* received similar ratings for *Pleasantness*, *Embodiment* and sense of *Presence*. Regarding those measures, other works found significant differences between teleportation and joysticks as well as motion-based locomotion and leaning [54, 104], but no such effects could be found between *PushPull* and *Joystick* in this study.

10.6.5.3 Lower Cybersickness for Stuttered Locomotion Than for Smooth Locomotion (H_7)

Regarding *Stuttered Locomotion*, we predicted a reduction in *Cybersickness* due to fewer instances of optical flow, a hypothesis that is supported by the results of our analysis, in particular for *Joystick* locomotion. Thus, hypothesis H_7 was confirmed. These results are similar to those of [99], who implemented discrete forward/backward teleport steps of each 1 meter that can be triggered via mouse buttons. However, *Stuttered Locomotion* expands on this idea by being compatible to virtually any continuous locomotion technique, while supporting movements in any direction with a multitude of available user settings. In the context of teleport steps, a participant came forward with an interesting analogy. They compared their perception of *Stuttered Locomotion* to the unique gait of birds like pigeons and chickens. During ground movement, those animals keep their heads in a stable position while the body moves and abruptly thrust their head forward to the next stable position. The most likely theory for why these birds bob their heads is for the same reason that we move our eyes around – to stabilize the image of their surroundings while in motion [106]. This is indeed similar to the *Stuttered Locomotion* approach to reducing *Cybersickness* by limiting optical flow, motivating further investigations in the future.

10.6.5.4 Lower Pleasantness for Stuttered Locomotion Than for Smooth Locomotion (H_8)

We found no significant differences for the measure of *Pleasantness* between *Stuttered* and *Smooth Locomotion*. Hence, *Stuttered Locomotion* was received more positively than anticipated and hypothesis H_8 was rejected. Although the analysis suggests that stuttering has no disadvantages over smooth locomotion in this regard, individual participants reported that they perceived this technique to be unpleasant. When analyzing the interaction effect for *Usability*, we indeed noticed that the combination of *Stuttered Locomotion* and *PushPull* was rated strikingly unfavorable. We suspect that this is due to the initial threshold that has to be passed in order for the first translation to take place. Until the threshold is passed, there is no visual feedback on whether or not the technique works correctly, whereas there is immediate feedback for both joystick configurations. One solution to this problem is to visualize the vector from initial to current hand position in form of an arrow inside a sphere with the radius of the required threshold. When the arrow touches the wall of the sphere, a teleport step is initiated and the arrow resets. If this extension is considered, our recommendation based on the presented user study is to offer *Stuttered Locomotion* as an option whenever continuous locomotion is employed.

10.7 Conclusion and Future Work

In this chapter, we presented two approaches for improving shared virtual experiences by reducing confusion of observers while a user is (potentially noncontinuously) moving through the VE. We addressed this challenge (i) from the observer perspective, by visualizing noncontinuous movements of users via continuous avatar transitions, and (ii) from a first-person perspective, by introducing a locomotion mode that allows mapping continuous input to short-interval movements with reduced cybersickness.

In terms of transitions, we introduced and compared four *Smart Avatar* techniques, which use different continuous visualizations for noncontinuous movements. All the techniques have in common that the avatar imitates the assigned user's (head and arm) movements when nearby, but switches to an autonomous navigation as soon as the user performs a noncontinuous movement in the VE (e.g., teleportation). In a user study, observers rated *Smart Avatars* significantly higher in self-reported *Spatial Awareness*, *Pragmatic* and *Hedonic* qualities as well as *Attractiveness* when compared to conventional avatars.

A second approach to ensure that observers can follow a user's avatar movements would be to utilize continuous locomotion techniques, such as joystick-based input. However, since this has been shown to induce higher levels of cybersickness compared to teleportation, we proposed a concept called *Stuttered Locomotion*. It involves the decomposition of a continuous movement into short teleport steps that are easier to track by observers than a single teleport over a long distance, but at the same time can lead to reduced cybersickness levels of the moving user. These hypothesized positive effects were confirmed in a second user study in which *Stuttered Locomotion* was applied to two continuous locomotion techniques. This benefit was more substantial for the joystick-based input, possibly due to *PushPull*'s demonstrated property of causing less cybersickness than joystick movement overall. On the downside, it was also found that *Stuttered Locomotion* was perceived as significantly less efficient than smooth locomotion.

The techniques of *Smart Avatars* and *Stuttered Locomotion* for both joysticks and *PushPull* were bundled into a toolkit that was made openly available for other researchers to use and extend.

While both approaches can separately contribute to a smoother experience for multiple moving users in a shared VE, they can also be combined. Since teleportation has been previously shown to increase spatial disorientation not only for the observer but also for the moving user [291], *Stuttered Locomotion* could provide an alternative, especially for traveling short distances. In fact, the intermittent short-range teleports that were added on top of continuous locomotion in *HyperJump* by Adhikari et al. [3, 4, 256] did not impair spatial orientation, which suggests that *Stuttered Locomotion* might similarly be able to reduce cybersickness without significantly impairing users' spatial orientation. Compared to related techniques like *Dashing* [27] or *Viewpoint Snapping* [99, 100], that can also provide repeated jumps, *Stuttered Locomotion* has the advantage of not needing to be manually triggered, thus presumably reducing cognitive load. By adding the *Smart Avatar* system on top, stuttered motion can be smoothed out to simulate a fully continuous avatar movement for the observers.

Different nuances of interplay between moving users and observers are the subject of future studies. In particular, we only measured whether observers had the impression that they knew where the other user was, which can be the case even if there is a considerable difference between the user's actual position and the position of their avatar. Other objective measures could be included to examine the extent to which the subjectively assumed position and the actual position match. The influence of a possible mismatch on interaction quality between users could be investigated with according tasks.

Furthermore, the current study did not assess how participants would perceive *Smart Avatar* transitions from their own perspective. Instead of seeing their own avatar only at the actual destination, users could be shown the same transitions as observers when they look behind them or via mirrors in the scene. *Smart Avatar* could likewise be employed as tools for distant interactions by performing the required action in the user's stead. As such, the following chapter discusses an extension of *Smart Avatar* that allows the user to remotely control their virtual body.

11. Puppeteer

The (Non-)Continuous Toolkit offers locomotion and avatar visualization techniques for shared MR spaces. While *Smart Avatars* have been demonstrated to be a useful tool for depicting natural-looking human gait for any kind of locomotion, the decoupling of avatar and user could be employed for user interactions as well. In this chapter, we explore how full-body avatars can be used to perform distant interactions. The following RQ will be discussed:

- *RQ_7: How can the user's point of view be transitioned to an outside perspective of their own body in order to replace self-movement?*

11.1 Motivation

MR experiences are often limited by their users' physical abilities. User interactions must be designed around the limitations of the human body, which prohibits intricate actions that can be found in non-immersive media, such as acrobatic maneuvers performed by the protagonist in 3D games. However, while most users would not be able to perform intricate acrobatic actions, a decoupled *Smart Avatars* could do so in their stead. By experimenting with off-mesh traversal to visualize an avatar that crosses a wide gap by performing a summersault, it became evident that *Smart Avatars* could also be used to visualize movements and interactions that would usually not be feasible in MR.

In this chapter we describe an extension to *Smart Avatars* that enables dynamic switching between first- and third-person perspectives, allowing users to perform supernatural interactions. The main principle of this technique is to give the user the capability to send their own virtual body to a specified location, while the virtual camera stays in place. This decoupled body can be seen and remotely controlled by their user, akin to a puppet. Hence, we call them *Puppet Avatars* and the overall technique *Puppeteer*.

Puppeteer replaces the need for self-movement with pseudo-locomotion by having the virtual body perform a distant interaction in the user's stead. Furthermore, *Puppeteer* can visualize a delayed teleportation sequence, where the user's view is realigned with the visible avatar after it performed a navigation task, e.g., after depicting a summersault over a wide gap. In a user study, we compared different input methods and avatar appearances regarding self-reported measures such as the perceived effectiveness and body ownership.

To summarize, the contributions of this chapter are as follows:

- *Puppeteer*: remote and gesture controlled *Smart Avatars* for dynamic switching between first- and third-person perspective VR interactions &
- A study to survey the acceptance of remote controlled *Smart Avatars* from a third-person perspective.

11.2 Background

Third-person interactions in MR can be found in a number of commercial products and research. For instance, the VR game *Moss* uses a classic platformer-style control scheme to move the character, while the VR user observes the scene from a third-person viewpoint [246]. Furthermore, Nintendo Co., Ltd. recently released VR compatibility modes for a selection of their games on Nintendo Switch [220]. Here, the default third-person perspective is simply rendered as stereoscopic images. Similarly, the commercial software *VorpX* allows a plethora of non-immersive games to be experienced in VR by injecting code into the graphics rendering pipeline on the driver level. This includes games with a third-person perspective [13].

Regarding research projects, [165] proposed a third-person VR interaction method which lets users control objects and the VE through hand-based manipulations. They reported an overall similar sense of presence between third- and first-person interactions, and a slight reduction in cybersickness-related symptoms. [212] evaluated how different viewing perspectives relate to psychophysiological engagement, arousal, and valence metrics. There was no conclusive data on which perspective was superior in this regard, however. On the other hand, [108] conducted a study on body ownership during VR experiences with conflicting sensory stimuli. Their results indicate that illusory ownership of a virtual body can be achieved in both first- and third-person perspectives under congruent visio-motor-tactile conditions. However, subjective body ownership and reaction to threat were generally stronger for first-person perspective and alternating conditions than for third-person perspective. According to [241], a sensory illusion can be achieved in which participants look at their own body from another person's point of view, while maintaining a high sense of body ownership of that other person's body. While not an experiment involving avatars, this study demonstrates that the sense of body ownership can be transferred onto other bodies. With respect to the avatar's positioning, a rubber-hand experiment was performed by [240] in which they manipulated the distance between visual and real hand. Synchronous visuo-tactile stimulation and connectivity of the virtual arm with the rest of the virtual body were found to be the main deciding factor for body ownership, while the alignment between the real and virtual arms and the distance between them were less important. [177] further found that body ownership can also be applied to non-humanoid avatars such as animals. In an experiment, they compared different first- and third-person controls for such animal avatars. Their results indicate that body ownership was generally lower during third-person trials. Lastly, [192] showed that full-body tracking can be used as a controller for puppetry of 2D shadow silhouettes, describing it as a mixture of acting and puppeteering.

11.3 Technique Description

We developed two distinct versions of *Puppeteer*: direct *Remote Controls* and *Gesture Controls*. *Remote Controls* let the user give movement commands to their puppet avatars with the same arching ray cast as common teleportation techniques employ for locomotion. When the command button is released, the avatar agent's pathfinding target is set to the command marker's location. Upon arrival, the puppet avatar engages the gimbal lock to always face in the same direction as its user, and enables the IK pass to imitate the user's head and hand poses. In order for this IK

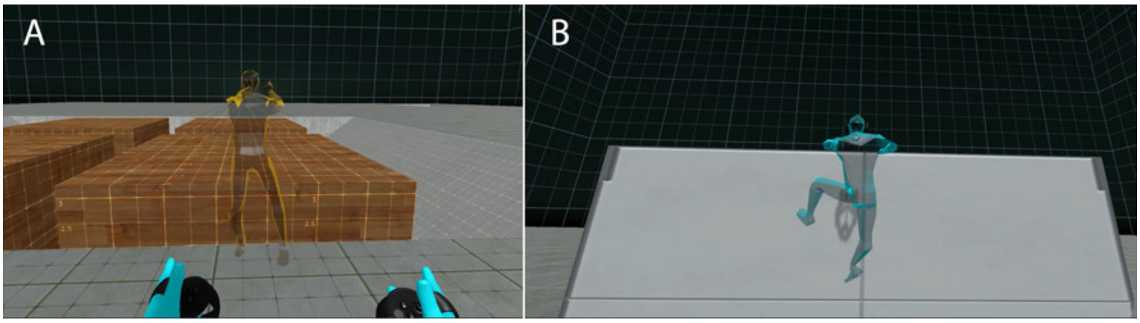


Figure 11.1: (A) The puppet avatar preview, signifying the ability to initiate an off-mesh link navigation task. (B) The puppet avatar performing a climbing task in the user's stead, as seen by the user.

pass to work correctly, we implemented a simple imitator rig consisting of one entity each for both hands and the head. This rig is positioned within the reference frame of the puppet avatar and thus moves long with it. To correctly align these imitations we simply assume that the virtual camera should be positioned in the avatars head, and apply the same relative positional and rotational offset from the user's head to their hands to the imitator equivalents. Now, any real-world motion is visualized through the puppet avatar, albeit in a different location, and thus visible to the user from a third-person perspective. Any item that was held by the user is transferred to the avatar's hand upon decoupling, and transferred back to the user when the avatar returns. To recall the avatar, the user has to either give a movement command to their location or simply press a designated return button. Upon return, the avatar resumes the default *Smart Avatar* behaviour.

Besides direct *Remote Controls*, a variation of *Puppeteer* was implemented that relies on gesture commands. Gestures in general have been shown to be an effective medium for communication in shared VR spaces [330]. this principle can also be applied to a decoupled puppet avatar. *Gesture Controls* do not let the user give specific movement commands to a ground position, but instead use hand gestures to let the avatar play a specific animation. For example, quickly raising one's arms above head level could be the gesture command to initiate a nearby off-mesh link traversal over a gap in the ground. In this scenario, the puppet avatar decouples, becomes observable and performs an animation that might be outside of the user's physical limitations, such as the aforementioned summersault (cf. Figure 11.1 (A)). After the animation finishes, the user is teleported to the location the avatar was last seen in. Available interactions with nearby off-mesh links can optionally be visualized with the same transparent ghost-like shader discussed in Chapter 10. Other animations, not directly related to locomotion, can be incorporated as well, of course. Any action a character in a non-immersive experience can perform, is now also available to VR users by temporarily switching to a third-person perspective (cf. Figure 11.1 (B)).

The implementation of *Puppeteer* is available as part of the (Non-)Continuous Toolkit on GitHub [114].

11.4 User Study

To establish the effectiveness of the introduced technique, we conducted a user study. We compared both versions of *Puppeteer*, *Remote Controls* and *Gesture Controls*, with each other in the same task as in the previously presented locomotion study in Section 10.6. During the introductory section of the experiment, an array of three mirrors was positioned in the VE, reflecting the participant's movement through a *Smart Avatar*. This element was mainly added to prime the participants on how they should expect their virtual body to look like in this remote control experiment stage.

Besides the input method, we were interested in possible effects of the puppet avatar's appearance. We suspected that observing one's own body from the outside might lead to unpleasant out-of-body experiences or uncanny valley effects [213]. A solution to this problem might be to present the avatar as a transparent ghost rather than a fully opaque body. The intent was to imply that the user remote controls an abstract projection of themselves rather than their actual body. The second independent variable therefore was the *Avatar Appearance*, using the same *Ghost* and *Opaque* avatars as employed in the long-distance travel visualizations of the (Non-)Continuous Toolkit.

This study was performed as part of the multi-stage experiment described in Chapter 10 and was thus conducted with the same participants and likewise as within-subject design. The presented task was the same as in the first-person experiment introduced in Section 10.6, giving us the opportunity to contrast first- and third-person controls in VR. Employed measures were the same as described in Section 10.6.1. The mean time for finishing this experiment stage was about 20 minutes per participant.

11.4.1 Hypotheses

Based on the above described criteria the following hypotheses were formed:

H_1: *Gesture controls* require fewer real world movements than any other input technique to perform the given task. It receives the highest *Efficiency* rating.
H_2: *Embodiment* is lower for third-person controls than in first-person controls.
H_3: *Presence* is lower for third-person controls than in first-person controls.
H_4: Third-person controls don't manipulate the virtual camera and thus receive favorable *Cybersickness* scores compared to first-person controls.
H_5: *Embodiment* is higher for the *Ghost* appearance than the *Opaque* appearance.

11.4.2 Stimuli and Procedure

This experiment directly followed the locomotion task introduced in Section 10.6. Therefore, participants already spent an average of 40 minutes in a VE. Between the previous and this study part, participants were reminded they could take breaks between experiment stages.

The study utilized the exact same task as the locomotion study: slicing fruits. However, instead of using first-person controls to reach the fruits, participants had to use *Puppeteer*. Unable to move from the starting position, *Remote Controls* required them to navigate their puppet avatar remotely by giving movement commands through an arching laser pointer, and directly cut the fruits by performing the required motions with their own hands. If positioned correctly, all three fruits of a set could be cut in a single swing. Alternatively, with *Gesture controls* they had to directly look at their target and initiate a gesture command by raising their hands to eye level and swinging down to hip level. The puppet avatar would then decouple from them, dash towards the fruits with a supernatural acrobatic animation of performing multiple summersaults, before slicing all three fruits at once. Then the avatar returned to the participant's location. There were 4 trials for this experiment stage. After each trial participants were asked to fill in the questionnaires described in Section 11.4.3.

11.4.3 Results

In this section the results of the statistical analysis are presented. We considered the two factors *Input Method* (*Remote Controls* / *Gesture Controls*) and *Avatar Appearance* (*Ghost* / *Opaque*), and therefore performed a two-way ANOVA with a 5% significance level for each of the dependent variables. To contrast first- and third-person controls, we further compared individual *Input Method* levels to those employed in the first-person experiment described in Section 10.6 (*Joystick* / *PushPull*) by a series of Tukey's HSD tests. For *Pleasantness* and *Embodiment*, an inspection of both histograms and Q-Q plots showed a small negative skewness (-0.280 and -0.510, respectively) and kurtosis (-0.701 and -0.448, respectively) of the residuals, however, the ANOVA has been shown to be robust against such mild deviations from the normal distribution [221]. All questionnaire results are listed in Table 10.6.3 and illustrated in Figure 11.2.

11.4.3.1 Efficiency - cf. Fig. 11.2 A

The two-way ANOVA revealed significant differences between the levels regarding self-reported *Efficiency* ($F(1,23) = 75.235$, $p < 0.001$, $\eta_p^2 = 0.245$). *Gesture Controls* ($M = 6.44$, $SD = 0.62$) were rated significantly higher than *Remote Controls* ($M = 4.04$, $SD = 1.76$, $p < 0.01$), *Joystick* ($M = 4.42$, $SD = 1.73$, $p < 0.001$) and *PushPull* ($M = 3.83$, $SD = 1.63$, $p < 0.001$). There was no significant difference between *Ghost* ($M = 5.35$, $SD = 1.84$) and *Opaque* appearances ($M = 5.13$, $SD = 1.84$, $p = 0.41$). *Gesture Controls* were therefore the only option that was rated significantly higher than any other.

11.4.3.2 Pleasantness - cf. Fig. 11.2 B

The ANOVA revealed significant differences between the levels regarding self-reported *Pleasantness* ($F(1,23) = 28.9$, $p < 0.001$, $\eta_p^2 = 0.314$). *Gesture Controls* ($M = 5.94$, $SD = 1.05$) were rated significantly higher than *Remote Controls* ($M = 4.42$, $SD = 1.65$, $p < 0.001$), *Joystick* ($M = 4.06$, $SD = 1.84$, $p < 0.001$) and *PushPull* ($M = 4.33$, $SD = 1.66$, $p < 0.001$). The *Ghost* appearance ($M = 5.44$, $SD = 1.53$) was rated significantly higher than the *Opaque* one ($M = 4.92$, $SD = 1.59$, $p = 0.048$). Again, *Gesture Controls* were rated highest among the locomotion techniques.

11.4.3.3 Embodiment - cf. Fig. 11.2 C

The ANOVA revealed significant differences between the levels regarding self-reported *Embodiment* ($F(1,23) = 1.95$, $p < 0.001$, $\eta_p^2 = 0.032$). There were no significant differences between *Gesture Controls* ($M = 4.17$, $SD = 1.65$) and *Remote Controls* ($M = 4.47$, $SD = 1.25$, $p = 0.98$). The *Remote Controls* rating was on the edge of being significantly lower than *Joystick* ($M = 5.45$, $SD = 1.10$, $p = 0.051$), while *Gesture Controls* were significantly lower than both *Joystick* ($p = 0.002$) and *PushPull* ($M = 5.35$, $SD = 1.01$, $p = 0.001$). There were no significant differences between the *Ghost* ($M = 4.53$, $SD = 1.55$) and *Opaque* appearances ($M = 4.11$, $SD = 1.35$, $p = 0.93$).

As we expected *Embodiment* to be the main deciding factor of *Puppeteer*'s viability, a deeper analysis was performed on individual questionnaire items. Here, we found no differences for any individual item between *Joystick* and *PushPull*, as well as between *Smooth* and *Stuttered* locomotion. We will group them together and use the joystick results exclusively to represent traditional first-person input methods. All the mean and standard deviation values can be seen in Table 11.2. Please refer to Table 11.1 to cross-reference the full length questions. For Questionnaire item BO2, both *Joystick* and *Remote Controls* received significantly higher ratings than *Gesture Controls* ($p < 0.01$, $p = 0.044$). For item BO3, *Joystick* was rated higher than both *Gesture Controls* ($p = 0.028$) and *Remote Controls* ($p = 0.030$). In item BO4, *Joystick* was rated higher than *Gesture Controls* ($p = 0.01$). For items BO5 and BO7, *Joystick* received favorable ratings compared to both *Gesture Controls* and *Remote Controls* ($p < 0.01$ each). There were no significant differences for the items BO1, BO6 and BO8.

11.4.3.4 Cybersickness - cf. Fig. 11.2 D

For the SSQ, we compared the differences from the questionnaire results gathered before and after every trial (POST-PRE). The ANOVA revealed significant differences between the levels regarding self-reported *Cybersickness* ($F(1,23) = 29.2$, $p < 0.001$, $\eta_p^2 = 0.532$). There was no significant difference between *Gesture Controls* ($M = -0.55$, $SD = 3.33$) and *Remote Controls* ($M = -0.47$, $SD = 3.63$, $p = 0.99$), while both yielded significantly lower SSQ scores than *Joystick* ($M = 3.58$, $SD = 10.77$, $p < 0.001$). We found no significant differences between the *Ghost* ($M = 0.48$, $SD = 3.72$) and *Opaque* appearances ($M = 0.47$, $SD = 2.92$, $p = 0.97$).

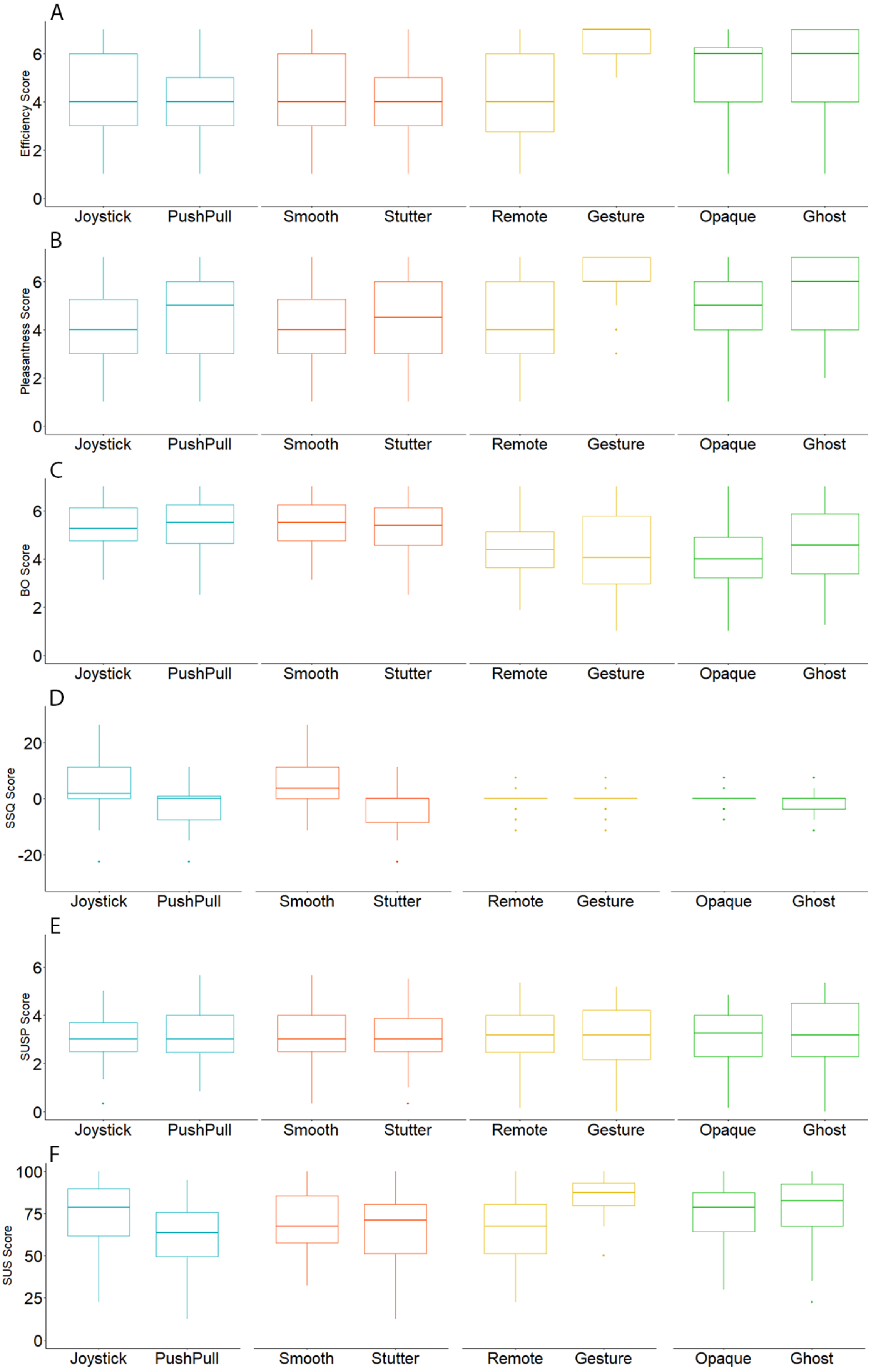


Figure 11.2: Boxplots of the study results: Efficiency (A), Pleasantness (B), Embodiment (C), Cybersickness (D), Presence (E), and Usability (F).

Code	Item
BO1	I caused the movement and cutting actions.
BO2	I performed the movement and cutting actions myself.
BO3	The visible avatar represented me.
BO4	I felt like I could control the avatar as if it was my own body.
BO5	I felt as if the avatar was my body.
BO6	I felt out of my body.
BO7	I felt as if my body was located where I saw the avatar.
BO8	I felt like the avatar was someone else.

Table 11.1: The *Body Ownership* questionnaire based on [118]. Participants were asked how strongly they agreed with these statements.

Input	BO1	BO2*	BO3*	BO4*	BO5*	BO6	BO7*	BO8
Joystick	5.85, 1.27	5.88, 1.18	5.04, 1.40	4.83, 1.56	4.81, 1.55	2.42, 1.34	4.90, 1.56	2.33, 1.70
Gesture	5.46, 1.57	3.73, 1.94	3.79, 1.98	3.40, 2.04	3.38, 2.11	2.54, 1.86	3.44, 2.12	3.29, 2.32
Remote	5.63, 1.28	5.02, 1.53	3.81, 1.74	4.25, 1.79	3.40, 1.85	2.77, 1.65	3.27, 2.11	2.83, 1.82

Table 11.2: *Body Ownership* results for individual items (M, SD). An asterisk indicates a significant difference. Table 11.1 lists the full length questions.

11.4.3.5 Presence - cf. Fig. 11.2 E

The Friedman test showed that there were no significant differences between the levels regarding the self-reported sense of *Presence* ($p = 0.98$).

11.4.3.6 Usability - cf. Fig. 11.2 F

The ANOVA revealed significant differences between the levels regarding self-reported *Usability* ($F(1,23) = 37.06$, $p < 0.001$, $\eta_p^2 = 0.413$). *Gesture Controls* ($M = 85.89$, $SD = 10.11$) were rated significantly higher than *Remote Controls* ($M = 65.31$, $SD = 20.70$, $p < 0.001$) and *Push-Pull* ($M = 61.20$, $SD = 20.77$, $p < 0.01$), but not significantly higher than *Joystick* ($M = 75.16$, $SD = 17.09$, $p = 0.038$). We found no significant differences between the *Ghost* ($M = 77.08$, $SD = 20.07$) and *Opaque* appearances ($M = 74.11$, $SD = 18.30$, $p = 0.38$).

11.5 Discussion

In this section we discuss the results of our user study with regards to the formed hypotheses. In the third experiment stage, both versions of *Puppeteer* and two puppet avatar appearances were evaluated in the same task as the first-person interactions. In this section we will therefore not only compare third-person controls with each other, but also with first-person controls.

Efficiency Regarding *Efficiency*, *Gesture controls* were rated significantly higher than *Remote Controls*. In fact, they were rated even higher than any of the first-person controls. Instead of manually moving to the indicated targets and interacting with them directly, *Gesture controls* allowed the participants to give their body a single command, upon which it performs the required task on its own. Naturally, these were also the quickest trials. Thus, hypothesis H_1 (*Gesture controls* require fewer real world movements than any other input technique to perform the given task. It receives the highest *Efficiency* rating) was confirmed. We found no significant difference between the *Remote Controls* and first-person controls, however.

Usability The *Usability* ranking mirrors that of *Efficiency*. *Gesture controls* were rated with an outstanding mean Usability score of 85.89, placing it between "excellent" and "best imaginable" on the adjective rating scale according to the SUS evaluation guidelines of [24]. *Joystick*'s score

(75.16) is "excellent", while *Remote Controls* (65.31) and *PushPull* (61.20) both fall under the "good" Usability category.

Pleasantness For *Pleasantness*, *Gesture controls* again received the highest ratings among all input methods. As with *Efficiency*, *Remote Controls* received similar ratings compared to both *PushPull* and *Joystick*. Participants enjoyed using both versions of *Puppeteer* as much or more than an established locomotion technique, indicating that there is great potential for third-person interactions in VR.

Of particular note is that the *Ghost* appearance was rated significantly higher than the *Opaque* puppet avatar appearance, but only by a slight margin. We believe that this is mainly due to the fact that *Ghost* puppet avatars don't obscure the sight of their user, as they are mostly transparent. Thus, when the puppet avatar is positioned between user and interaction target, the target remains in sight. This is the only significant difference the analysis showed for the appearance. As such, hypothesis H_5 (*Embodiment* is higher for the *Ghost* appearance than the *Opaque* appearance) has to be rejected. There were no measurable drawbacks to using the *Ghost* over the *Opaque* appearance, making it a safe choice to represent puppet avatars.

Body Ownership We expected *Embodiment* to be the main deciding factor for *Puppeteer*'s viability as a VR interaction paradigm. The results were difficult to predict, especially considering the out-of-body nature of this technique. Yet, the results were surprisingly positive. *Remote Controls* received ratings that were fairly close to first-person controls, with a mean value of 4.47 versus 5.45 of *Joystick*. The difference was on the edge of being significantly lower ($p = 0.051$). However, this value still exceeded our expectations for *Embodiment*. *Gesture controls* and its mean value of 4.17 on the other hand was significantly lower than both *Joystick* and *PushPull*. It is likely that third-person controls in general provide a lower sense of *Embodiment*, and thus H_2 (*Embodiment* is lower for third-person controls than in first-person controls) can be assumed to be correct. Even though *Remote Controls* mimicked the user's hands and head pose perfectly, resulting in a slightly higher *Embodiment* mean score than *Gesture controls*, it did not amount to a significant effect. These findings are in line with other studies regarding third-person interactions [108].

As this measure was of special interest to us, we performed further detailed analytic comparisons between individual items of the *Embodiment* questionnaire (cf. Tables 11.1 & 11.2). Here, we found that participants had perceived to have caused the visible actions in a similar manner between first- and third-person controls. They felt like they had performed the movement and interactions themselves almost equally between first-person controls and *Remote Controls*, but less so with *Gesture controls*. The same is true for the perceived ability to control the avatar as if it was their own body. The impression that the avatar was their body and located where they saw it was similar between both versions of *Puppeteer*, but higher for *Joysticks*. Interestingly, we found no indication that *Puppeteer* causes a feeling of having an out-of-body experience. By the same token, the avatar was not seen as a different person. In summary, there was a noticeable difference for *Embodiment* between first- and third-person controls, but arguably well within the range of acceptable embodiment for VR experiences.

Cybersickness We expected third-person controls in general to cause fewer *Cybersickness* related symptoms, simply by virtue of not having to move through the VE. The results indicate that *Puppeteer* neither increased nor decreased *Cybersickness*, making it a system that can be used for prolonged sessions and thus confirming hypotheses H_4 (third-person controls don't manipulate the virtual camera and thus receive favorable *Cybersickness* scores compared to first-person controls). It must be noted, however, that there undoubtedly were order effects in place when directly comparing first- and third-person controls, as those trials were separated from each other. It stands to reason, though, that these results would either have been nearly identical or in favor of *Puppeteer* if first- and third-person trials had been shuffled, as we compared the pre and post SSQ scores for each trial.

Sense of Presence Another prediction was that usage of third-person controls has a negative effect on the sense of *Presence*. The analysis revealed no such effect. [254] outlined that the sense of *Presence* directly relates to a user's sense of agency, which is reflected on the *Embodiment* questionnaire items BO1, BO2 and BO4. Here, *Remote Controls* received similar ratings to first-person controls, implying equivalent agency. Subsequently, hypothesis H_3 (*Presence* is lower for third-person controls than in first-person controls) has to be rejected. This is another factor that further cements the viability of third-person controls.

Guidelines The data suggests that experience designers have creative freedom of how the inclusion of puppet avatars should be framed. It could be introduced as either the user's own body or a projection of it. However, as item BO7 of the *Embodiment* questionnaire indicates that the puppet avatars were not seen as another person, it is questionable if they can be established as an independent companion. This is likely due to the fact that the puppet avatar emerges from the position of the user, and returns to them upon recall. Although, the *Puppeteer* technique could be repurposed for commanding external agents with minor modifications if so desired.

Both Remote and Gesture Controls have their pros and cons, and both seem to be promising and viable interaction forms for VR experiences. One of Gesture Controls' potential use-cases is the visualization of off-mesh links, e.g., climbing a ladder or jumping over gaps. While the is in reach of an off-mesh link, performing a gesture can trigger a delayed teleportation. The avatar decouples, and visualizes the required motion from a third-person perspective. Afterwards, the user is teleported to again coincide with the *Puppet Avatar*.

The question that still remains is if experiences have to be deliberately designed around *Puppeteer*, or if it can be easily incorporated into existing solutions. From a technical standpoint, a given project only has to support autonomously navigating agents and IK passes for avatars in order to use both *Smart Avatars* and *Puppeteer*. Therefore, we believe that *Smart avatars* and *Remote Controls* can be easily added to existing solutions. *Gesture Controls*, on the other hand, require the implementation of specific interactions for the puppet avatar, making them less universally applicable.

Overall, *Puppeteer* demonstrated that third-person interactions in MR are not only possible, but highly enjoyable. The ability to perform superhuman actions by temporarily switching to a third-person perspective opens up VR experience design immensely, allowing the incorporation of mechanics that were though to be exclusive to non-immersive interactive media. Potential applications include elaborate close quarters combat systems, parkour-esque locomotion or simply switching to a bird's eye perspective on the user's location. Because of the demonstrated qualities of *Puppeteer*, it was added as integral part of the (Non-)Continuous Toolkit.

11.6 Conclusion and Future Work

In this chapter, we presented an extension to the *Smart Avatars* system, which replaces the need for self-movement by having the virtual body perform distant interactions in the user's stead. *Remote Controlled* avatars received *Embodiment* ratings close to first-person controls, while *Gesture controls* convinced participants with outstanding *Usability*, albeit at the cost of lower embodiment.

To combine the benefits of *Remote* and *Gesture Controls*, we would like to implement a hybrid *Puppeteer* solution. Based on participant feedback, the marker-based movement commands will be replaced by gesture commands to decouple and recall the avatar. While decoupled, the avatar takes on the *Remote Control* behaviour, directly mimicking its user. At any time, gesture commands for off-mesh link navigation or interactions can be triggered. However, these commands will only trigger a portion of the interaction animation. At certain key moments in the interaction, the user has to perform another predefined gesture command in order for the animation to continue. Missing the time window will abort the interaction. With this approach, we intent for the users to have a stronger feeling of directly being in control of the puppet avatar, even during supernatural animations. Implementing and testing this hybrid technique remains as a target for future work.

In addition to enabling distant interactions, *Puppeteer* can further visualize sitting users as standing avatars regardless of input method. This conversion is applied automatically, as the spatial relation of the user's real hands and head are transferred to the puppet avatar when nearby. Depending on the underlying set of animations, the puppet avatar can thus be depicted as standing, sitting or even flying while imitating their assigned user. Overall, the (Non-)Continuous Toolkit proved to be a functional solution for the hybrid user group aspect of different locomotion types and their visualization from both first- and third-person perspectives.

VI

Conclusion

12. Summary

In this dissertation, novel shared MR spaces and corresponding interaction techniques have been introduced, including an in-depth discussion of potential use cases, fundamental components and their technical implementation. Due to additional layers of information visualization, shared MR spaces can provide a high degree of remote human-to-human information transmission, which might surpass that of face-to-face communication. Such shared MR spaces are an ideal for the *new work* paradigm, for instance, by providing remote training in medicine, mechanical engineering or education. A notable use case that gained enormous media attention recently is the metaverse, which connects users through an alternative virtual world, consisting of many individual or linked shared MR spaces. However, several technical aspects of hybrid MR setups still remain largely unexplored.

This dissertation discussed research questions on how the state-of-the-art can be further advanced. These research questions cover aspects of hybrid user groups such as different hardware device types, sitting and standing modes, avatar appearance, locomotion types and different visualizations of said locomotion. The main contributions were grouped into three parts: (i) enabling hybrid collaborations, (ii) shared experiences, and (iii) interactions through embodied avatars. A number of novel techniques in the realm of interaction and visualization were introduced, and their usefulness was demonstrated through a series of user studies.

Two generic tools that facilitate the administration of MR user studies have been presented. The first being a questionnaire, which predicts the likeliness of a user experiencing cybersickness symptoms when exposed to MR, the *CSSQ*. The items on this questionnaire were compiled from several previous works that have demonstrated the influence of various biological, chemical and psychological factors on the human senses and nervous system. Using this questionnaire, users can have a greater degree of confidence entering MR for the first time, and researchers performing studies in MR can select participants based on the study's purpose. A study showed a correlation between *CSSQ* items and the perceived cybersickness during and after motion in a VE. A high mean value in the *CSSQ* indicates a stronger perception of cybersickness during and after the exposure to MR.

Next, the *Remote Study Framework* was introduced to enable researchers to collect data remotely without direct supervision. It divides a user study into a series of Study States, which are grouped into Study Sequences to represent individual trials, and finally arranges Study Sequences into Study Blocks that have to be completed in a fixed order. The Study States define the behaviour of the VE, properties of the interaction and visualization techniques, the given instructions to the participant, as well as the state of an interactive questionnaire. This questionnaire can be handled through an external service like LimeSurvey or Google Forms, while being displayed in the VE through a web browser extension. This framework has been demonstrated to be an effective tool in the user study described in Chapter 10.

Hybrid Collaborations

In order to enable collaborations between hybrid user groups with regard to the variance of MR device types, we integrated mobile devices and VST HMDs into shared MR spaces. To address the limitation of immersive experiences often being limited to those users wearing an MR HMD, we presented a project-independent smartphone application called *VR Invite*, which allows multiple bystanders to observe and interact with the VE and the HMD users through handheld VST AR viewports. The required viewports are rendered on a host computer, which transmits the data via wireless LAN to the mobile devices. Moreover, the position and orientation of the smartphone is tracked to adjust the viewpoints accordingly. We conducted a user study in the context of rehabilitation for older adults, with a focus on integrating bystanders. We compared *VR Invite* with a TV-gamepad-combination as interaction medium for the bystander regarding sense of presence, social presence, workload and usability, both with purely verbal and active assistance capabilities. The results indicate that the opportunity for direct interaction positively influences the bystander's sense of presence in the VE and the reported usability of the smartphone application. However, social presence was rated higher in passive conditions in which the real person was the center of attention, as opposed to the avatar on the screen. Furthermore, users valued the comfort of sitting down over active participation and agency with room-scale locomotion. This streaming technology can be applied to any VST device, regardless of proximity to the user.

The principle of VST AR can likewise be applied to the VST HMD, which supports stereoscopic rendering of virtual objects in the real-world environment, e.g., the avatars of other users. Compared to real humans within one's FOV, virtual human representations are easy to identify in VST AR, even when using high resolution human 3D scans. Their position is not stable within the visible real world reference frame due to camera-to-photon latency, which is a detriment to both the technical immersion and psychological sense of presence. In order to reduce the registration error between captured real-world videos and rendered virtual images, we introduced *Camera Time Warp*, a novel reprojection technique for VST AR. Instead of rendering the image plane locked to the virtual camera, *Camera Time Warp* renders the image plane at the real-world position it was captured at, and compensates for potential artifacts. We conducted two experiments to evaluate the effectiveness of *Camera Time Warp*. In the first experiment, participants were asked to report subjective discomfort while moving their head in a pattern inspired by the ISO 9241-9 Fitts' Law task at different speeds while the video feed was rendered at varying frame rates. The results show that the technique can significantly reduce subjective levels of discomfort and cybersickness-related symptoms for all tested configurations. In the second experiment, participants were asked to move physical objects on a projected path as quickly and precisely as possible. Results indicate a positive effect of *Camera Time Warp* on speed and accuracy.

VR Invite and *Camera Time Warp* increase the accessibility of shared MR spaces for VST mobile devices and HMDs. Thus, the hybrid user group aspect of heterogeneous device setups could be addressed successfully.

Shared Experiences

With the ability to integrate MR devices from the entire reality-virtuality continuum into shared MR spaces, we further researched aspects of human-to-human interactions. We focused on visualization techniques with regard to shared spatial awareness as well as the sense of spatial and co-presence. This includes rendering techniques for points of interest, avatar animations, and visualization of locomotion.

In shared MR spaces, passive information visualization mechanisms can support the users' ability to forecast imminent actions of others. Visual indicators of users' perspectives can reduce the required verbal communication and therefore increase the efficiency of work within remote teams. We compared three distinct types of visual indicators for a user's point of interest in a VE regarding task completion time, precision and error rate. These perspective conveying methods include a 3D cone indicating the boundaries of a user's FOV, highlighting the object a user is looking at, and displaying a direct video mirror of the user's viewport. We used these methods to transmit information from one person to another, that would otherwise be inaccessible. In a virtual warehouse scenario, participants moved visually obstructed boxes to a target area with a crane, while an experiment conductor provided visual cues as to where these boxes are, solely by looking at them. The results indicate that object highlighting is significantly inferior regarding precision and error rate, while the video mirror proved to be the most reliable. The view cone was however perceived as the method with the highest degree of social presence. These findings suggest that users of *VR Invite* should be represented as either physical representations of the handheld device, or in form of a video mirror.

Furthermore, a user study was performed that was designed to investigate the impact of avatar appearance and locomotion techniques on observers in shared MR spaces. In our study we compared continuous joystick-based and non-continuous teleportation-based locomotion with regard to the observer's perceived (co-)presence, cybersickness, fairness and enjoyment in a competitive game environment. Likewise, we compared the effects of humanoid and abstract avatar appearances. The results revealed significant effects of the locomotion type on perceived co-presence and subjective fairness, such that continuous locomotion yielded significantly higher perceived co-presence and fairness. However, there were no significant differences between the avatar appearances except for a weak positive effect of realistic avatars on mutual awareness. These findings suggest that a greater emphasis should be put on the visual representation of teleportation-based navigation in shared MR spaces, for instance, by animating positional transitions. They further suggest that a distinction for the effect of avatar appearance has to be made based on the context of the MR application. While previous work showed a positive effect of realistic avatars on co-presence during cooperative tasks, we found no such effect in a competitive setting.

The hybrid user group aspects of avatar appearance, locomotion and its visualization were confirmed as integral factors of human-to-human interactions in shared MR spaces.

Embodied Interactions

Following the confirmation of the importance of locomotion and its visualization, we presented a number of techniques that address these aspects of hybrid user groups. These novel techniques are centered around procedurally animated full-body avatars, which portray natural-looking human gait from both first- and third-person perspectives.

We introduced *VR Strider*, a novel LUI for seated MR experiences, which maps cycling biomechanics of the user's legs to virtual walking movements. The core idea is to translate the motion of pedaling on a mini exercise bike to a corresponding walking animation of a virtual avatar while providing audio-based tactile feedback on virtual ground contacts. We conducted an experiment to evaluate the LUI and our novel anchor-turning rotation control method regarding task performance, spatial cognition, cybersickness, sense of presence, usability and comfort in a path-integration task. The results show that *VR Strider* has a significant positive effect on the participants' angular and distance estimation, sense of presence and feeling of comfort compared to other established locomotion techniques, such as teleportation and joystick-based navigation. A confirmatory study further stressed the necessity of synchronized avatar animations for virtual locomotion. We have demonstrated that this technique offers a solution for the hybrid user group aspect of sitting and standing modes.

Finally, we presented a toolkit for continuous and non-continuous locomotion with matching human representations based on IVAs. this toolkit includes two techniques each for avatar visualization, i.e., *Smart Avatars* and *Puppeteer*, as well as locomotion, i.e., *PushPull* and *Stuttered Locomotion*. *Smart Avatars* deliver continuous full-body human representations for non-continuous locomotion in shared MR spaces. They are controlled by IVAs that will either imitate their user's real-world movements when close-by, or autonomously navigating to their assigned user when the distance between them exceeds a certain threshold, i.e., after the user teleports. As part of the *Smart Avatar* system, we implemented four avatar transition techniques and compared them to conventional avatar locomotion in a user study, revealing significant positive effects on the observers' spatial awareness, as well as pragmatic and hedonic quality scores. Participants favored the *Foresight* technique, which provides a projection of where the user will end their movement with several in-between ghost-like images, that a naturally moving *Smart Avatar* passes through. Smart Avatars were developed to smoothly visualize *Stuttered Locomotion*, a technique which transforms continuous user movement into a series of short-distance teleport steps. In a second experiment, we compared smooth and *Stuttered Locomotion* for both joysticks and the novel motion-based *PushPull* technique, which employs a dynamic velocity multiplier based on the user's hand pose. The results show that both *PushPull* and *Stuttered Locomotion* significantly reduce cybersickness symptoms. In a third experiment, we evaluated the acceptance of employing a third-person perspective on one's own body through *Puppeteer*, a remote control extension of *Smart Avatars*. Here, direct remote controls showed body ownership close to first-person interactions, while gesture controls received higher usability scores than first-person controls. Furthermore, *Puppeteer* can visualize sitting users as standing avatars, thus addressing the hybrid user group aspect of sitting and standing modes without the need for proprietary LUIs.

Through our user studies and the proposed suite of techniques, the hybrid user group aspects of avatars, locomotion and its visualization were addressed for both first- and third-person perspectives. In conclusion, this work offers novel techniques and guidelines for the implementation of hybrid MR setups with respect to virtual interactions and their visualization.

13. Outlook

In this dissertation, research questions related to shared MR spaces, enabling hybrid MR setups, shared experiences and embodied interactions were explored and evaluated. This dissertation has revealed and inspired further research, which could be addressed in future studies. For instance, shared MR spaces could greatly benefit from a sophisticated system that increases the expressiveness of avatars through facial and hand tracking. Such a system could further increase the capacity of non-verbal communication channels. In a prior work, we have demonstrated that hand tracking in combination with detailed haptic feedback can increase the body ownership of visible avatars, even if the proportions of limbs are distorted drastically [17]. However, multi-user hand tracking requires a relatively high network bandwidth, as the rotations for every joint have to be synchronized to all other clients. This is acceptable for applications with a low number of concurrent users, but is unlikely to scale well enough for use cases such as the metaverse. Therefore, an effective compression algorithm for hand poses would be beneficial. One potential approach for this compression is to utilize an unsigned 32 bit integer, which can hold numbers of up to 10 digits, and map each finger's pose to two of those digits. This way, each finger can be represented through 100 different poses.

In order to map the angular configuration of three joints to two digits, we can define a finger's pose as a combination of a single digit base joint angle and a single digit curling factor. For example, the index finger's base joint has a range of motion of approximately 135°, which equates to ten increments, or one digit, of each 13.5°. However, the increments do not have to be mapped linearly, allowing for a more dense mapping of typical poses. The second and third joints are co-linked and can thus be described through a single digit, which similarly divides the range of motion into ten segments.

While this technique results in a reduction of visual fidelity, compared to a naive implementation, it reduces the bandwidth by a factor of 60 (15 joints, each represented through a quaternion consisting of four 32-bit floating point numbers). To ensure a smooth visualization, the transition between possible finger configurations can be interpolated. In order to test this approach's applicability, we would like to study the detection threshold for finger movement on other users' avatars. Depending on the distance at which the reduction of fidelity becomes detectable, this technique can either be used for universal performance optimization or exclusively for distant avatars in scenarios with a high number of concurrent users.

The expressiveness of avatar hands could be further improved by amplifying detected gestures. This is helpful in scenarios with large distances between avatars or occluding geometry between them. Moreover, the intended meaning of a gesture could be emphasized by displaying a matching iconography, which opens an opportunity to deepen intercultural communication in the metaverse. If the metaverse concept proves to be successful, it will likely become a global platform for users from many different countries and cultures. However, the interpretation of both facial expressions and hand gestures have been reported to vary depending on cultural factors of observers [156]. Therefore, it is crucial for providers and users of MR social platforms to be aware of cultural differences in body language [76, 300]. A visualization tool for translation of gestures between users of different cultures would thus present an interesting research opportunity.

In the metaphor that is suggested in the metaverse's name, each multi-user MR application is a singular universe within the meta-universe, the universe of universes. However, the aspect of transitioning between universes, and more specifically shared MR spaces, remains largely unexplored. As such, further research should be conducted on both the technical implementation and the user experience of this central aspect of the metaverse. Different visualizations of this transition are conceivable, ranging from simple camera fading techniques to interactable portals that lead to another realm. Depending on the content of the applications, techniques to blend together different shared MR spaces could be employed as well. Here, two universes could momentarily overlap, before the transition is completed and the universes separate. Additionally, it would be interesting to explore the aspect of transitioning between applications that belong to different categories of the reality-virtuality continuum, such as smoothly transitioning from an AR to a VR application. Computer vision algorithms used for camera-based inside-out 3D tracking such as SLAM build 3D maps of the user's surroundings, indicating the possibility to visually interpolate between the geometry of the user's real-world environment and the VE of the VR application.

Ultimately, the metaverse poses an enormous potential to change several aspects of our daily lives. Until the required infrastructure is built, research continues to develop and refine the visualization of and interactions within shared MR spaces. If this technology is used appropriately, it could transform the way we think about work, leisure and socializing. Thus, once the technical foundation is laid, it becomes important to consider not only the technical advancement of this new-found digital world, but also the humans that will live in it. Developers and designers of shared MR spaces will need to reflect on ethical, social and moral perspectives in order for the metaverse to thrive.

Bibliography

[1] 3DVision-Blog. 2014. First impressions from the new oculus rift dk2 3d vr hmd. Retrieved 03/24/2022 from `http://3dvision-blog.com/9326-first-impressions-from-the-new-oculus-rift-dk2-3d-vr-hmd/`.

[2] Matt Adcock, Stuart Anderson, and Bruce Thomas. 2013. Remotefusion: real time depth camera fusion for remote collaboration on physical tasks. In *Proceedings of the 12th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry*, 235–242.

[3] Ashu Adhikari. 2021. *Improving Spatial Orientation in Virtual Reality with Leaning-based Interfaces*. Master's thesis. Simon Fraser University, Surrey, BC, Canada.

[4] Ashu Adhikari, Daniel Zielasko, Alexander Bretin, Markus von der Heyde, Ernst Kruijff, and Bernhard E Riecke. 2021. Integrating continuous and teleporting vr locomotion into a seamless" hyperjump" paradigm. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 370–372.

[5] Adobe Systems Incorporated. 2022. Mixamo. Retrieved 01/18/2022 from `https://www.mixamo.com/`.

[6] sarvesh agrawal sarvesh, adèle simon adèle, søren bech søren, klaus bæntsen klaus, and søren forchhammer søren. 2020. Defining immersion: literature review and implications for research on audiovisual experiences. *journal of the audio engineering society*, 68, 6, 404–417.

[7] Sun Joo Ahn, Joshua Bostick, Elise Ogle, Kristine L. Nowak, Kara T. McGillicuddy, and Jeremy N. Bailenson. 2016. Experiencing nature: embodying animals in immersive virtual environments increases inclusion of nature in self and involvement with nature. *Journal of Computer-Mediated Communication*, 21, 6, 399–419.

[8] Majed Al Zayer, Isayas B. Adhanom, Paul MacNeilage, and Eelke Folmer. 2019. The effect of field-of-view restriction on sex bias in vr sickness and spatial navigation performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19) Article 354. ACM, Glasgow, Scotland Uk, 354:1–354:12. ISBN: 978-1-4503-5970-2.

[9] Alex Albert, Matthew R. Hallowell, Brian Kleiner, Ao Chen, and Mani Golparvar-Fard. 2014. Enhancing construction hazard recognition with high-fidelity augmented virtuality. *Journal of Construction Engineering and Management*, 140, 7, 04014024.

[10] Robert S. Allison, Laurence R. Harris, Michael Jenkin, Greg Pintilie, Fara Redlick, and Daniel C. Zikovitz. 2000. First steps with a rideable computer. In *Proceedings IEEE Virtual Reality 2000 (Cat. No. 00CB37048)*. IEEE, 169–175.

[11] Ivan Alsina-Jurnet and José Gutiérrez-Maldonado. 2010. Influence of personality and individual abilities on the sense of presence experienced in anxiety triggering virtual environments. *International journal of human-computer studies*, 68, 10, 788–801.

[12] Samuel Ang and John Quarles. 2020. Gingervr: an open source repository of cybersickness reduction techniques for unity. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 460–463.

[13] Animation Labs. 2022. Vorpx. Retrieved 01/18/2022 from `https://www.vorpx.com/`.

[14] Ferran Argelaguet, Alexander Kulik, André Kunert, Carlos Andujar, and Bernd Froehlich. 2011. See-through techniques for referential awareness in collaborative virtual reality. *International Journal of Human-Computer Studies*, 69, 6, 387–400. ISSN: 1071-5819.

[15] Andreas Aristidou and Joan Lasenby. 2011. Fabrik: a fast, iterative solver for the inverse kinematics problem. *Graphical Models*, 73, 5, 243–260.

[16] Andreas Aristidou, Joan Lasenby, Yiorgos. Chrysanthou, and A. Shamir. 2018. Inverse kinematics techniques in computer graphics: a survey. *Computer Graphics Forum*, 37, 6, 35–58.

[17] Oscar Ariza, Jann P. Freiwald, Nadine Laage, Michaela Feist, Mariam Salloum, Gerd Bruder, and Frank Steinicke. 2016. Inducing body-transfer illusions in vr by providing brief phases of visual-tactile stimulation. In *Proceedings of the 2016 Symposium on Spatial User Interaction*. ACM, 61–68.

[18] 2019. *Around the (virtual) world: infinite walking in virtual reality using electrical muscle stimulation. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–8. ISBN: 9781450359702.

[19] Steve Aukstakalnis. 2016. *Practical augmented reality: A guide to the technologies, applications, and human factors for AR and VR*. Addison-Wesley Professional.

[20] Jeremy N Bailenson, Kim Swinth, Crystal Hoyt, Susan Persky, Alex Dimov, and Jim Blascovich. 2005. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence: Teleoperators & Virtual Environments*, 14, 4, 379–393.

[21] Jeremy N Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. 2006. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments*, 15, 4, 359–372.

[22] Michael Bajura and Ulrich Neumann. 1995. Dynamic registration correction in video-based augmented reality systems. *IEEE Computer Graphics and Applications*, 15, 5, 52–60.

[23] Stacy A Balk, Mary Anne Bertola, and Vaughan W Inman. 2013. Simulator sickness questionnaire: twenty years later. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 257–263.

[24] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual sus scores mean: adding an adjective rating scale. *Journal of usability studies*, 4, 3, 114–123.

[25] Rosa Mar*ı*a Baños, Cristina Botella, Mariano Alcañiz, V*ı*ictor Liaño, Belén Guerrero, and Beatriz Rey. 2004. Immersion and emotion: their impact on the sense of presence. *Cyberpsychology & behavior*, 7, 6, 734–741.

[26] Gary Bente and Nicole C Krämer. 2011. Virtual gestures: embodiment and nonverbal behavior in computer-mediated communication. *Face-to-face communication over the internet: issues, research, challenges*, 176–209.

[27] Jiwan Bhandari, Paul MacNeilage, and Eelke Folmer. 2018. Teleportation without Spatial Disorientation using Optical Flow Cues. en. In *Graphics Interface 2018*. Toronto, Canada, 153–158. Retrieved 10/22/2018 from.

[28] Mark Billinghurst and Hirokazu Kato. 1999. Collaborative mixed reality. In *Proceedings of the First International Symposium on Mixed Reality*, 261–284.

[29] Mark Billinghurst and Hirokazu Kato. 2002. Collaborative augmented reality. *Commun. ACM*, 45, 7, 64–70. ISSN: 0001-0782.

[30] Mark Billinghurst, Ivan Poupyrev, Hirokazu Kato, and Richard May. 2000. Mixing realities in shared space: an augmented reality interface for collaborative computing. In *2000 IEEE international conference on multimedia and expo. ICME2000. Proceedings. Latest advances in the fast changing world of multimedia (Cat. No. 00TH8532)*. Volume 3. IEEE, 1641–1644.

[31] Oliver Bimber and Ramesh Raskar. 2005. *Spatial augmented reality: merging real and virtual worlds*. CRC press.

[32] Frank Biocca and Chad Harms. 2003. Networked minds social presence inventory:|(scales only, version 1.2) measures of co-presence, social presence, subjective symmetry, and intersubjective symmetry.

[33] Frank Biocca, Chad Harms, and Jenn Gregg. 2001. The networked minds measure of social presence: pilot test of the factor structure and concurrent validity. In *4th annual international workshop on presence, Philadelphia, PA*, 1–9.

[34] Jim Blascovich and Jeremy Bailenson. 2011. *Infinite reality: Avatars, eternal life, new worlds, and the dawn of the virtual revolution*. William Morrow & Co.

[35] Jim Blascovich, Jack Loomis, Andrew C.Beall, Kimberly R. Swinth, Crystal L. Hoyt, and Jeremy N.Bailenson. 2002. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13, 2, 103–124.

[36] Jan Bobeth, Susanne Schmehl, Ernst Kruijff, Stephanie Deutsch, and Manfred Tscheligi. 2012. Evaluating performance and acceptance of older adults using freehand gestures for tv menu control. In *Proceedings of the 10th European Conference on Interactive TV and Video* (EuroITV '12). Association for Computing Machinery, Berlin, Germany, 35–44. ISBN: 9781450311076.

[37] Costas Boletsis. 2017. The new era of virtual reality locomotion: a systematic literature review of techniques and a proposed typology. *Multimodal Technologies and Interaction*, 1, 4, 24.

[38] Costas Boletsis and Jarl Erik Cedergren. 2019. Vr locomotion in the new era of virtual reality: an empirical comparison of prevalent techniques. *Advances in Human-Computer Interaction*, 2019.

[39] Elizabeth M Bonsignore, Derek L Hansen, Zachary O Toups, Lennart E Nacke, Anastasia Salter, and Wayne Lutters. 2012. Mixed reality games. In *Proceedings of the ACM 2012 conference on computer supported cooperative work companion*, 7–8.

[40] Stéphane Bouchard, Julie St-Jacques, Geneviève Robillard, and Patrice Renaud. 2008. Anxiety increases the feeling of presence in virtual reality. *Presence: Teleoperators and Virtual Environments*, 17, 4, 376–391.

[41] Laroussi Bouguila and Makoto Sato. 2002. Virtual locomotion system for large-scale virtual environment. In *Proceedings IEEE Virtual Reality 2002*. IEEE, 291–292.

[42] Doug Bowman, Ernst Kruijff, Joseph J LaViola Jr, and Ivan P Poupyrev. 2004. *3D User interfaces: theory and practice*. Addison-Wesley.

[43] Doug A Bowman, David Koller, and Larry F Hodges. 1998. A methodology for the evaluation of travel techniques for immersive virtual environments. *Virtual reality*, 3, 2, 120–131.

[44] Doug A Bowman and Ryan P McMahan. 2007. Virtual reality: how much immersion is enough? *Computer*, 40, 7, 36–43.

[45] Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. 2016. Point & teleport locomotion technique for virtual reality. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, 205–216.

[46] Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. *International journal of human-computer studies*, 59, 1-2, 119–155.

[47] John Brooke et al. 1996. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189, 194, 4–7.

[48] Frederick P. Brooks Jr. 1987. Walkthrough a dynamic graphics system for simulating virtual buildings. In *Proceedings of the 1986 workshop on Interactive 3D graphics*. ACM, 9–21.

[49] Emily Brown and Paul Cairns. 2004. A grounded investigation of game immersion. In *CHI'04 extended abstracts on Human factors in computing systems*, 1297–1300.

[50] Susan Bruck and Paul A Watters. 2009. Estimating cybersickness of simulated motion using the simulator sickness questionnaire (ssq): a controlled study. In *2009 sixth international conference on computer graphics, imaging and visualization*. IEEE, 486–488.

[51] Sean Buckley. 2015. This is how valve's amazing lighthouse tracking technology works. Retrieved 03/24/2022 from `https://gizmodo.com/this-is-how-valve-s-amazing-lighthouse-tracking-technol-1705356768`.

[52] Timothy J Buker, Dennis A Vincenzi, and John E Deaton. 2012. The effect of apparent latency on simulator sickness while using a see-through helmet-mounted display: reducing apparent latency with predictive compensation. *Human factors*, 54, 2, 235–249.

[53] Saniye Tugba Bulu. 2012. Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education*, 58, 1, 154–161. ISSN: 0360-1315.

[54] Fabio Buttussi and Luca Chittaro. 2019. Locomotion in place in virtual reality: a comparative evaluation of joystick, teleport, and leaning. *IEEE Transactions on Visualization and Computer Graphics*.

[55] Gordon Calleja. 2011. *In-game: From immersion to incorporation*. mit Press.

[56] Jorge CS Cardoso. 2016. Comparison of gesture, gamepad, and gaze-based locomotion for vr worlds. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*, 319–320.

[57] Christer Carlsson and Olof Hagsand. 1993. Dive a multi-user virtual reality system. In *Proceedings of IEEE Virtual Reality Annual International Symposium*. IEEE, 394–400.

[58] Juan S Casaneuva. 2001. *Presence and co-presence in collaborative virtual environments*. PhD thesis. University of Cape Town.

[59] Joe Cecil, Parmesh Ramanathan, Vahid Rahneshin, Aditya Prakash, and Miguel Pirela-Cruz. 2013. Collaborative virtual environments for orthopedic surgery. In *2013 IEEE international conference on automation science and engineering (CASE)*. IEEE, 133–137.

[60] Chih-Hui Chang, Wu-Wen Pan, Li-Ya Tseng, and Thomas A Stoffregen. 2012. Postural activity and motion sickness during video game play in children and adults. *Experimental brain research*, 217, 2, 299–309.

[61] Eunhee Chang, Hyun Taek Kim, and Byounghyun Yoo. 2020. Virtual reality sickness: a review of causes and measurements. *International Journal of Human–Computer Interaction*, 36, 17, 1658–1682.

[62] Heni Cherni, Natacha Métayer, and Nicolas Souliman. 2020. Literature review of locomotion techniques in virtual reality. *International Journal of Virtual Reality*.

[63] Isaac Cho and Zachary Wartell. 2015. Evaluation of a bimanual simultaneous 7dof interaction technique in virtual environments. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 133–136.

[64] Elizabeth Chrastil and William Warren. 2008. Testing models of path integration in a triangle completion task. *Journal of Vision*, 8, 6, 1153–1153.

[65] Chris Christou, Aimilia Tzanavari, Kyriakos Herakleous, and Charalambos Poullis. 2016. Navigation in virtual reality: comparison of gaze-directed and pointing motion control. In *2016 18th Mediterranean Electrotechnical Conference (MELECON)*. IEEE, 1–6.

[66] Shih-kai Chung and James K Hahn. 1999. Animation of human walking in virtual environments. In *Proceedings Computer Animation 1999*. IEEE, 4–15.

[67] Elizabeth F Churchill and Dave Snowdon. 1998. Collaborative virtual environments: an introductory review of issues and systems. *virtual reality*, 3, 1, 3–15.

[68] Ernest Cline. 2011. *Ready Player One*. Crown Publishing Group.

[69] Arzu Çöltekin, Ian Lochhead, Marguerite Madden, Sidonie Christophe, Alexandre Devaux, Christopher Pettit, Oliver Lock, Shashwat Shukla, Lukáš Herman, Zdeněk Stachoň, et al. 2020. Extended reality in spatial sciences: a review of research challenges and future directions. *ISPRS International Journal of Geo-Information*, 9, 7, 439.

[70] Karin Coninx, Frank Van Reeth, and Eddy Flerackers. 1997. A hybrid 2d/3d user interface for immersive object modeling. In *Proceedings Computer Graphics International*. IEEE, 47–55.

[71] HTC Corporation. 2021. Vive facial tracker. Retrieved 03/24/2022 from `https://business.vive.com/eu/product/vive-facial-tracker/`.

[72] HTC Corporation. 2022. Viverse. Retrieved 03/24/2022 from `https://www.vive.com/us/viverse/`.

[73] Valve Corporation. 2020. Half-life: alyx. Retrieved 03/24/2022 from `https://www.half-life.com/en/alyx/`.

[74] Richard L. Daft and Robert H. Lengel. 1983. Information richness. A new approach to managerial behavior and organization design. Technical report. Texas A and M Univ College Station Coll of Business Administration.

[75] Richard L. Daft and Robert H. Lengel. 1986. Organizational information requirements, media richness and structural design. *Management science*, 32, 5, 554–571.

[76] Emma Dai and Jay Bal. 2009. Harmonising culture in co-operative business ventures: using a simulation in a metaverse. In *Proceedings of the 3rd European Conference on Games Based Learning*. Academic Conferences Ltd., 412–420.

[77] Natasha Dailey. 2021. Metaverse is a buzzword. Retrieved 03/24/2022 from `https://markets.businessinsider.com/news/currencies/elon-musk-metaverse-is-a-marketing-buzzword-crypto-2021-12`.

[78] Sébastien Dalibard, Nadia Magnenat-Talmann, and Daniel Thalmann. 2012. Anthropomorphism of artificial agents: a comparative survey of expressive design and motion of virtual characters and social robots. In *Workshop on autonomous social robots and virtual humans at the 25th annual conference on computer animation and social agents (CASA 2012)*.

[79] Rudolph P. Darken, William R. Cockayne, and David Carmein. 1997. The omni-directional treadmill: a locomotion device for virtual worlds.

[80] Bradley Austin Davis, Karen Bryla, and Philips Alexander Benton. 2015. *Oculus Rift in Action*. Manning Publications Company.

[81] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. 2014. A systematic review of cybersickness. In *Proceedings of the 2014 Conference on Interactive Entertainment*. ACM, 1–9.

[82] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. 2015. Comparing the onset of cybersickness using the oculus rift and two virtual roller coasters. In *Proceedings of the 11th Australasian Conference on Interactive Entertainment (IE 2015)*. Volume 27, 30.

[83] Nicolas J Dedual, Ohan Oda, and Steven K Feiner. 2011. Creating hybrid user interfaces with a 2d multi-touch tabletop and a 3d see-through head-worn display. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 231–232.

[84] Darleen M DeRosa, Donald A Hantula, Ned Kock, and John D'Arcy. 2004. Trust and leadership in virtual teamwork: a media naturalness perspective. *Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management*, 43, 2-3, 219–232.

[85] Doc-Ok. 2014. Hacking the oculus rift dk2. Retrieved 03/24/2022 from `http://doc-ok.org/?p=1138`.

[86] Trevor J. Dodds and Roy A. Ruddle. 2008. Using teleporting, awareness and multiple views to improve teamwork in collaborative virtual environments. In *Virtual Environments 2008*. Eurographics Association, 81–88.

[87] José Dorado, Pablo Figueroa, Jean-Rémy Chardonnet, Frédéric Merienne, and Tiberio Hernandez. 2019. Homing by triangle completion in consumer-oriented virtual reality environments. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1652–1657.

[88] Sarah A. Douglas, Arthur E. Kirkpatrick, and I. Scott MacKenzie. 1999. Testing pointing device performance and user assessment with the iso 9241, part 9 standard. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 215–222.

[89] Andrew T. Duchowski, Nathan Cournia, Brian Cumming, Daniel McCallum, Anand Gramopadhye, Joel Greenstein, Sajay Sadasivan, and Richard A Tyrrell. 2004. Visual deictic reference in a collaborative virtual environment. In *Proceedings of the 2004 symposium on Eye tracking research & applications*, 35–40.

[90] Thierry Duval and Cédric Fleury. 2009. An asymmetric 2d pointer/3d ray for 3d interaction within collaborative virtual environments. In *Proceedings of the 14th international Conference on 3D Web Technology*, 33–41.

[91] David W. Eccles and Gershon Tenenbaum. 2004. Why an expert team is more than a team of experts: a social-cognitive conceptualization of team coordination and communication in sport. *Journal of Sport and Exercise Psychology*, 26, 4, 542–560.

[92] Mica R. Endsley. 1988. Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society annual meeting* number 2. Volume 32. SAGE Publications Sage CA: Los Angeles, CA, 97–101.

[93] Mica R. Endsley. 1995. A taxonomy of situation awareness errors. *Human factors in aviation operations*, 3, 2, 287–292.

[94] Elliot E. Entin and Daniel Serfaty. 1999. Adaptive team coordination. *Human factors*, 41, 2, 312–325.

[95] Goffman Erving. 1963. Behavior in public places: notes on the social organization of gatherings. *New York*, 3.

[96] J. Alberto Espinosa, .F Javier Lerch, and Robert E. Kraut. 2004. Explicit versus implicit coordination mechanisms and task dependencies: one size does not fit all.

[97] Exit Games. 2020. Photon unity networking 2. Retrieved 11/20/2020 from `https://doc-api.photonengine.com/en/pun/v2/index.html`.

[98] Marc Fabri and David Moore. 2005. The use of emotionally expressive avatars in collaborative virtual environments. *Virtual Social Agents*, 88, April, 88–94.

[99] Yasin Farmani and Robert J Teather. 2020. Evaluating discrete viewpoint control to reduce cybersickness in virtual reality. *Virtual Reality*, 24, 4, 645–664.

[100] Yasin Farmani and Robert J. Teather. 2018. Viewpoint Snapping to Reduce Cybersickness in Virtual Reality. In *Proceedings of the 44th Graphics Interface Conference* (GI '18). Canadian Human-Computer Communications Society, Toronto, Canada, 168–175. ISBN: 978-0-9947868-3-8. Retrieved 07/14/2020 from.

[101] Ajoy S Fernandes and Steven K. Feiner. 2016. Combating vr sickness through subtle dynamic field-of-view modification. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, 201–210.

[102] International Organization for Standardization. 2000. Iso/dis 9241-9 ergonomic requirements for office work with visual display terminals (vdts) - part 9: requirements for non-keyboard input devices.

[103] Daniel Freeman, Sarah Reeve, A Robinson, Anke Ehlers, David Clark, Bernhard Spanlang, and Mel Slater. 2017. Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological medicine*, 47, 14, 2393–2400.

[104] Jann Philipp Freiwald, Oscar Ariza, Omar Janeh, and Frank Steinicke. 2020. Walking by cycling: a novel in-place locomotion user interface for seated virtual reality experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12.

[105] Jann Philipp Freiwald, Julius Schenke, Nale Lehmann-Willenbrock, and Frank Steinicke. 2021. Effects of avatar appearance and locomotion on co-presence in virtual reality collaborations. In *Mensch Und Computer 2021* (MuC '21). Association for Computing Machinery, Ingolstadt, Germany, 393–401. ISBN: 9781450386456.

[106] Barrie Frost. 1978. The optokinetic basis of head-bobbing in the pigeon. *Journal of Experimental Biology*, 74, 1, 187–195.

[107] 2019. *Assessing the accuracy of point & teleport locomotion with orientation indication for virtual reality using curved trajectories. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–12. ISBN: 9781450359702.

[108] Henrique Galvan Debarba, Sidney Bovet, Roy Salomon, Olaf Blanke, Bruno Herbelin, and Ronan Boulic. 2017. Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality. *PLOS ONE*, 12, 12, 1–19.

[109] Matthias Gamer and Heiko Hecht. 2007. Are you looking at me? measuring the cone of gaze. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 3, 705.

[110] Beat Games. 2019. Beat saber. Retrieved 03/24/2022 from `https://beatsaber.com/`.

[111] O. Serdar Gedik and A. Aydn Alatan. 2013. 3-d rigid body tracking using vision and depth sensors. *IEEE transactions on cybernetics*, 43, 5, 1395–1405.

[112] Thomas Geijtenbeek and Nicolas Pronost. 2012. Interactive character animation using simulated physics: a state-of-the-art review. In *Computer Graphics Forum* number 8. Volume 31. Wiley Online Library, 2492–2515.

[113] Dedre Gentner and Albert L Stevens. 1983. Mental models lawrence erlbaum associates. *Hillsdale, New Jersey*.

[114] GitHub. 2022. Smart avatars github repository. Retrieved 03/24/2022 from `https://github.com/uhhhci/TheNonContinuousToolkit`.

[115] GitHub. 2022. Vr invite github repository. Retrieved 03/24/2022 from `https://github.com/uhhhci/VRInvite`.

[116] GitHub. 2022. Vr strider github repository. Retrieved 03/24/2022 from `https://github.com/uhhhci/VRStrider`.

[117] Statista GmbH. 2021. Pokémon go: daily active users. Retrieved 03/24/2022 from `https://www.statista.com/statistics/604551/pokemon-go-daily-active-users-in-europe/`.

[118] Mar Gonzalez-Franco and Tabitha C Peck. 2018. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5, 74.

[119] Google. 2021. Arcore overview. Retrieved 03/24/2022 from `https://developers.google.com/ar/discover`.

[120] Daniel W Gower Jr and Jennifer Fowlkes. 1989. Simulator sickness in the UH-60 (Black Hawk) flight simulator. Technical report. ARMY AEROMEDICAL RESEARCH LAB FORT RUCKER AL.

[121] Simone Grassini and Karin Laumann. 2020. Questionnaire measures and physiological correlates of presence: a systematic review. *Frontiers in psychology*, 11, 349.

[122] Oliver Grau. 2004. *Virtual Art: from illusion to immersion*. MIT press.

[123] Chris Greenhalgh. 2012. *Large scale collaborative virtual environments*. Springer Science & Business Media.

[124] Keith Grochow, Steven L Martin, Aaron Hertzmann, and Zoran Popović. 2004. Style-based inverse kinematics. In *ACM SIGGRAPH 2004 Papers*, 522–531.

[125] Victoria Groom, Clifford Nass, Tina Chen, Alexia Nielsen, James K Scarborough, and Erica Robles. 2009. Evaluating the effects of behavioral realism in embodied agents. *International Journal of Human-Computer Studies*, 67, 10, 842–849.

[126] The Guardian. 2021. Facebook announces name change to meta in rebranding effort. Retrieved 03/24/2022 from `https://www.theguardian.com/technology/2021/oct/28/facebook-name-change-rebrand-meta`.

[127] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 2017. Sharevr: enabling co-located experiences for virtual reality between hmd and non-hmd users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 4021–4033.

[128] Jan Gugenheimer, Evgeny Stemasov, Harpreet Sareen, and Enrico Rukzio. 2018. Facedisplay: towards asymmetric multi-user interaction for nomadic virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13.

[129] Simon Gunkel, Martin Prins, Hans Stokking, and Omar Niamut. 2017. *WebVR meets WebRTC: Towards 360-degree social VR experiences*. IEEE.

[130] Carl Gutwin and Saul Greenberg. 2002. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work (CSCW)*, 11.

[131] Carl Gutwin, Reagan Penner, and Kevin Schneider. 2004. Group awareness in distributed software development. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (CSCW '04). Association for Computing Machinery, Chicago, Illinois, USA, 72–81. ISBN: 1581138105.

[132] Heiko Häckelmann, Hans J Petzold, and Susanne Strahringer. 2011. *Kommunikationssysteme: Technik und Anwendungen*. Springer-Verlag.

[133] Sam Halperin. 2016. Exploring Bicycle-Based Virtual Reality Exergames as a Design Space. Retrieved 03/24/2022 from `https://unitylist.com/p/m6q/vr-bike-game-controller`.

[134] Universität Hamburg. 2019. Limesurvey. Retrieved 03/24/2022 from `https://www.limesurvey.uni-hamburg.de/`.

[135] Mark A Hamilton and Kristine L. Nowak. 2010. Advancing a model of avatar evaluation and selection. *PsychNology Journal*, 8, 1.

[136] Alyssa Harris, Kevin Nguyen, Preston Tunnell Wilson, Matthew Jackoski, and Betsy Williams. 2014. Human joystick: wii-leaning to translate in large virtual environments. In *Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*, 231–234.

[137] Sandra G Hart and Lowell E Staveland. 1988. Development of nasa-tlx (task load index): results of empirical and theoretical research. In *Advances in psychology*. Volume 52. Elsevier, 139–183.

[138] Tilo Hartmann, Werner Wirth, Peter Vorderer, Christoph Klimmt, Holger Schramm, and Saskia Böcking. 2013. *Spatial Presence Theory: State of the Art and Challenges Ahead*. ISBN: 978-3-319-10189-7.

[139] Béatrice S Hasler and Doron A Friedman. 2012. Sociocultural conventions in avatar-mediated nonverbal communication: a cross-cultural analysis of virtual proxemics. *Journal of Intercultural Communication Research*, 41, 3, 238–259.

[140] Marc Hassenzahl, Michael Burmester, and Franz Koller. 2003. Attrakdiff: ein fragebogen zur messung wahrgenommener hedonischer und pragmatischer qualität. In *Mensch & computer 2003*. Springer, 187–196.

[141] Alex Heath. 2022. Meta's social vr platform horizon hits 300,000 users. Retrieved 03/24/2022 from `https://www.theverge.com/2022/2/17/22939297/meta-social-vr-platform-horizon-300000-users`.

[142] Dimitri Hein, Christian Mai, and Heinrich Hußmann. 2018. The usage of presence measurements in research: a review. In *Proceedings of the International Society for Presence Research Annual Conference (Presence)*. The International Society for Presence Research Prague, 21–22.

[143] Teresa Hirzle, Maurice Cordts, Enrico Rukzio, Jan Gugenheimer, and Andreas Bulling. 2021. A critical assessment of the use of ssq as a measure of general discomfort in vr head-mounted displays. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14.

[144] Elizabeth M Hodge, MHN Tabrizi, Mary A Farwell, and Karl L Wuensch. 2008. Virtual reality classrooms: strategies for creating a social presence. *International Journal of Social Sciences*, 2, 2, 105–109.

[145] John M. Hollerbach. 2002. Locomotion interfaces. *Handbook of virtual environments: Design, implementation, and applications*, 239–254.

[146] Roland Holm, Erwin Stauder, Roland Wagner, Markus Priglinger, and Jens Volkert. 2002. A combined immersive and desktop authoring tool for virtual environments. In *Proceedings IEEE Virtual Reality 2002*. IEEE, 93–100.

[147] Tina Iachini, Luigi Maffei, Massimiliano Masullo, Vincenzo Paolo Senese, Mariachiara Rapuano, Aniello Pascale, Francesco Sorrentino, and Gennaro Ruggiero. 2019. The experience of virtual reality: are individual differences in mental imagery associated with sense of presence? *Cognitive processing*, 20, 3, 291–298.

[148] Hikaru Ibayashi, Yuta Sugiura, Daisuke Sakamoto, Natsuki Miyata, Mitsunori Tada, Takashi Okuma, Takeshi Kurata, Masaaki Mochimaru, and Takeo Igarashi. 2015. Dollhouse vr: a multi-view, multi-user collaborative design workspace with vr technology. In *SIGGRAPH Asia 2015 Emerging Technologies*, 1–2.

[149] Apple Inc. 2022. Arkit overview. Retrieved 03/24/2022 from `https://developer.apple.com/augmented-reality/arkit/`.

[150] Epic Games Inc. 2021. Rift tour, a musical experience unlike any other. Retrieved 03/24/2022 from `https://www.epicgames.com/fortnite/en-US/rift-tour`.

[151] Google Inc. 2022. Google lens. Retrieved 03/24/2022 from `https://lens.google/`.

[152] Mighty Immersion Inc. 2022. Managexr. Retrieved 03/24/2022 from `https://www.managexr.com/`.

[153] Niantic Inc. 2016. Pokémon go. Retrieved 03/24/2022 from `https://pokemongolive.com/en/`.

[154] Hiroo Iwata. 1999. Walking about virtual environments on an infinite floor. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. IEEE, 286–293.

[155] Hiroo Iwata, Hiroaki Yano, and Fumitaka Nakaizumi. 2001. Gait master: a versatile locomotion interface for uneven virtual terrain. In *Proceedings IEEE Virtual Reality 2001*. IEEE, 131–137.

[156] Rachael E. Jack, Oliver G. B. Garrod, Hui Yu, Roberto Caldara, and Philippe G. Schyns. 2012. Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109, 19, 7241–7244.

[157] Jason Jerald. 2015. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.

[158] Dongsik Jo, Ki-Hong Kim, and Gerard Jounghyun Kim. 2016. Effects of avatar and background representation forms to co-presence in mixed reality (mr) tele-conference systems. In *SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments* (SA '16) Article 12. Association for Computing Machinery, Macau, 4 pages. ISBN: 9781450345484.

[159] David M Johnson. 2005. Introduction to and review of simulator sickness research. Technical report. Army Research Inst Field Unut Fort Rucker AL.

[160] Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. 2014. Roomalive: magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, 637–644.

[161] Kimberlee Jordan, John H. Challis, and Karl M. Newell. 2007. Walking speed influences on gait cycle variability. *Gait & Posture*, 26, 1, 128–134.

[162] Pekka Kallioniemi, Kimmo Ronkainen, Jussi Karhu, Sumita Sharma, Jaakko Hakulinen, and Markku Turunen. 2019. Citycompass vr-a collaborative virtual language learning environment. In *IFIP Conference on Human-Computer Interaction*. Springer, 540–543.

[163] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3, 3, 203–220.

[164] Erika Kerruish. 2019. Arranging sensations: smell and taste in augmented and virtual reality. *The Senses and Society*, 14, 1, 31–45.

[165] Mingyu Kim, Jiwon Lee, Changhun Kim, and Jinmo Kim. 2018. Tpvr: user interaction of third person virtual reality for new presence and experience. *Symmetry*, 10, 4. ISSN: 2073-8994.

[166] Peter Kim, Jason Orlosky, Kiyoshi Kiyokawa, Photchara Ratsamee, and Tomohiro Mashita. 2017. [poster] dotwarp: dynamic object timewarping for video see-through augmented reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, 184–185.

[167] Sohee Kim. 2021. Metaverse is a multitrillion-dollar opportunity, epic ceo says. Retrieved 03/24/2022 from `https://www.bloomberg.com/news/articles/2021-11-17/metaverse-is-a-multitrillion-dollar-opportunity-epic-ceo-says?sref=f8taTPHn`.

[168] Yeongkyu Kim. 2018. How to design an emotionally intelligent machine: in terms of visual aspect considering ‘uncanny valley’ theory. Retrieved 03/24/2022 from `https://medium.com/advanced-design-for-ai/how-to-make-an-emotionally-intelligent-machine-in-terms-of-visual-design-considering-uncanny-9bc473b0267f`.

[169] Youjeong Kim and S Shyam Sundar. 2012. Anthropomorphism of computers: is it mindful or mindless? *Computers in Human Behavior*, 28, 1, 241–250.

[170] Young Youn Kim, Hyun Ju Kim, Eun Nam Kim, Hee Dong Ko, and Hyun Taek Kim. 2005. Characteristic changes in the physiological components of cybersickness. *Psychophysiology*, 42, 5, 616–625.

[171] IKEA United Kingdom. 2022. Ikea place. Retrieved 03/24/2022 from `https://www.ikea.com/gb/en/customer-service/mobile-apps/`.

[172] Silvia Erika Kober and Christa Neuper. 2013. Personality and presence in virtual reality: does their relationship depend on the used presence measure? *International Journal of Human-Computer Interaction*, 29, 1, 13–25.

[173] Ned Kock. 2004. The psychobiological model: towards a new theory of computer-mediated communication based on darwinian evolution. *Organization science*, 15, 3, 327–348.

[174] Eugenia M Kolasinski. 1995. *Simulator sickness in virtual environments*. Volume 1027. US Army Research Institute for the Behavioral and Social Sciences.

[175] Christopher Kommetter and Martin Ebner. 2019. A pedagogical framework for mixed reality in classrooms based on a literature review. *EdMedia+ Innovate Learning*, 919–929.

[176] Igor Kotlyar and Dan Ariely. 2013. The effect of nonverbal cues on relationship formation. *Computers in Human Behavior*, 29, 3, 544–551.

[177] Andrey Krekhov, Sebastian Cmentowski, and Jens Krüger. 2018. Vr animals: surreal body ownership in virtual reality games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (CHI PLAY ’18 Extended Abstracts). Association for Computing Machinery, Melbourne, VIC, Australia, 503–511. ISBN: 9781450359689.

[178] Ernst Kruijff, Alexander Marquardt, Christina Trepkowski, Robert W. Lindeman, Andre Hinkenjann, Jens Maiero, and Bernhard E. Riecke. 2016. On your feet!: enhancing vection in leaning-based interfaces through multisensory stimuli. In *Proceedings of the 2016 Symposium on Spatial User Interaction* (SUI ’16). ACM, Tokyo, Japan, 149–158. ISBN: 978-1-4503-4068-7.

[179] Ernst Kruijff, Bernhard Riecke, Christina Trekowski, and Alexandra Kitson. 2015. Upper body leaning can affect forward self-motion perception in virtual environments. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, 103–112.

[180] Ernst Kruijff, J Edward Swan, and Steven Feiner. 2010. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 3–12.

[181] Owlchemy Labs. 2021. Mobile spectator. Retrieved 03/24/2022 from `https://owlchemylabs.com/owlchemy-mobile-spectator-ar-spectator-camera/`.

[182] James R Lackner. 2014. Motion sickness: more than nausea and vomiting. *Experimental brain research*, 232, 8, 2493–2510.

[183] Philipp Ladwig and Christian Geiger. 2018. A literature review on collaboration in mixed reality. In *International Conference on Remote Engineering and Virtual Instrumentation*. Springer, 591–600.

[184] Eike Langbehn, Paul Lubos, and Frank Steinicke. 2018. Evaluation of locomotion techniques for room-scale vr: joystick, teleportation, and redirected walking. In *Proceedings of the Virtual Reality International Conference-Laval Virtual*, 1–9.

[185] Joseph J. Laviola. 2000. A discussion of cybersickness in virtual environments. *SIGCHI Bulletin*.

[186] Joseph J LaViola Jr. 2000. A discussion of cybersickness in virtual environments. *ACM Sigchi Bulletin*, 32, 1, 47–56.

[187] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional.

[188] Joseph J. LaViola Jr. 2000. A discussion of cybersickness in virtual environments. *SIGCHI Bull.*, 32, 1, 47–56. ISSN: 0736-6906.

[189] Eun-Ju Lee. 2007. Character-based team identification and referent informational influence in computer-mediated communication. *Media Psychology*, 9, 1, 135–155.

[190] Geonsun Lee, HyeongYeop Kang, JongMin Lee, and JungHyun Han. [n. d.] A user study on view-sharing techniques for one-to-many mixed reality collaborations.

[191] Hojun Lee, Gyutae Ha, Sangho Lee, and Shiho Kim. 2017. A mixed reality tele-presence platform to exchange emotion and sensory information based on mpeg-v standard. In *2017 IEEE Virtual Reality (VR)*. IEEE, 349–350.

[192] Luıis Leite and Veronica Orvalho. 2012. Shape your body: control a virtual silhouette using body motion. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '12). Association for Computing Machinery, Austin, Texas, USA, 1913–1918. ISBN: 9781450310161.

[193] Siân E Lindley, James Le Couteur, and Nadia L Berthouze. 2008. Stirring up experience through movement in game play: effects on engagement and social behaviour. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 511–514.

[194] Facebook Technologies LLC. 2020. Oculus quest 2. Retrieved 03/24/2022 from `https://www.oculus.com/quest-2/`.

[195] Facebook Technologies LLC. 2022. Horizon worlds. Retrieved 03/24/2022 from `https://www.oculus.com/horizon-worlds/`.

[196] Facebook Technologies LLC. 2022. Oculus integration. Retrieved 03/24/2022 from `https://developer.oculus.com/documentation/unity/`.

[197] Christoph H. Loch, D. Charles Galunic, and Susan Schneider. 2006. Balancing cooperation and competition in human groups: the role of emotional algorithms and evolution. *Managerial and Decision Economics*, 27, 2/3, 217–233. ISSN: 01436570, 10991468.

[198] I. Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Hum.-Comput. Interact.*, 7, 1, 91–139. ISSN: 0737-0024.

[199] Kevin J. MacKenzie and Simon J. Watt. 2010. Eliminating accommodation-convergence conflicts in stereoscopic displays: Can multiple-focal-plane displays elicit continuous and consistent vergence and accommodation responses? In *Stereoscopic Displays and Applications XXI*. Andrew J. Woods, Nicolas S. Holliman, and Neil A. Dodgson, editors. Volume 7524. International Society for Optics and Photonics. SPIE, 414–423.

[200] Jean MacMillan, Elliot E. Entin, and Daniel Serfaty. 2004. Communication overhead: the hidden cost of team cognition.

[201] Allyn Malventano. 2016. Lighthouse tracking system dissected and explored. Retrieved 03/24/2022 from `https://pcper.com/2016/04/steamvr-htc-vive-in-depth-lighthouse-tracking-system-dissected-and-explored/2/`.

[202] Tony Manninen. 2002. Interaction forms in multiplayer desktop virtual reality games. In *VRIC2002 Conference*. Volume 223, 232.

[203] William E. Marsh, Jonathan W Kelly, Veronica J. Dark, and James H. Oliver. 2013. Cognitive demands of semi-natural virtual locomotion. *Presence: Teleoperators and Virtual Environments*, 22, 3, 216–234.

[204] Michael E McCauley and Thomas J Sharkey. 1992. Cybersickness: perception of self-motion in virtual environments. *Presence: Teleoperators & Virtual Environments*, 1, 3, 311–318.

[205] Morgan McCullough, Hong Xu, Joel Michelson, Matthew Jackoski, Wyatt Pease, William Cobb, William Kalescky, Joshua Ladd, and Betsy Williams. 2015. Myo arm: swinging to explore a ve. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, 107–113.

[206] Alison McMahan. 2003. Immersion, engagement and presence. *The video game theory reader*, 67, S 86.

[207] Eliana Medina, Ruth Fruland, and Suzanne Weghorst. 2008. Virtusphere: walking in a human size vr hamster ball. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* number 27. Volume 52. SAGE Publications Sage CA: Los Angeles, CA, 2102–2106.

[208] Mereriam-Webster. 2021. What is the 'metaverse'? a real word for a virtual world. Retrieved 03/24/2022 from `https://www.merriam-webster.com/words-at-play/meaning-of-metaverse`.

[209] Microsoft. 2017. Microsoft hololens. Retrieved 03/24/2022 from `https://www.microsoft.com/en-us/hololens`.

[210] Paul Milgram, Haruo Takemura, Akira Utsumi, Fumio Kishino, et al. 1994. Augmented reality: a class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies* number 11. Volume 2351, 282–292.

[211] Betty J. Mohler, Sarah H. Creem-Regehr, William B. Thompson, and Heinrich H. Buelthoff. 2010. The effect of viewing a self-avatar on distance judgments in an hmd-based virtual environment. *Presence: Teleoperators and Virtual Environments*, 19, 3, 230–242.

[212] Diego Monteiro, Hai-Ning Liang, Andrew Abel, Nilufar Bahaei, and Rita de Cassia Monteiro. 2018. Evaluating engagement of virtual reality games based on first and third person perspective using eeg and subjective metrics. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 53–60.

[213] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19, 2, 98–100.

[214] Fariba Mostajeran, Nikolaos Katzakis, Oscar Ariza, Jann Philipp Freiwald, and Frank Steinicke. 2019. Welcoming a holographic virtual coach for balance training at home: two focus groups with older adults. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1465–1470.

[215] Lisa Nakamura. 2013. *Cybertypes: Race, ethnicity, and identity on the Internet*. Routledge.

[216] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: social responses to computers. *Journal of Social Issues*, 56, 1, 81–103.

[217] W. Todd Nelson, Merry M. Roe, Robert S. Bolia, and Rebecca M. Morley. 2000. Assessing simulator sickness in a see-through HMD: Effects of time delay, time on task, and task complexity. Technical report. AIR FORCE RESEARCH LAB WRIGHT-PATTERSON AFB OH.

[218] Thinh Nguyen-Vo, Bernhard E. Riecke, Wolfgang Stuerzlinger, Duc-Minh Pham, and Ernst Kruijff. 2019. Naviboard and navichair: limited translation combined with full rotation for efficient virtual locomotion. *IEEE transactions on visualization and computer graphics*. ISSN: 1077-2626.

[219] G. Nie, Y. Liu, and Y. Wang. 2017. [poster] prevention of visually induced motion sickness based on dynamic real-time content-aware non-salient area blurring. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, 75–78.

[220] Nintendo Co., Ltd. 2019. Nintendo switch vr mode. Retrieved 01/18/2022 from `https://www.zelda.com/breath-of-the-wild/news/explore-hyrule-with-nintendo-labo-vr-kit/`.

[221] Geoff Norman. 2010. Likert scales, levels of measurement and the “laws” of statistics. *Advances in health sciences education*, 15, 5, 625–632.

[222] Kristine L. Nowak. 2001. Defining and differentiating copresence, social presence and presence as transportation. In *Presence 2001 Conference, Philadelphia, PA*. Citeseer, 1–23.

[223] Kristine L. Nowak. 2004. The influence of anthropomorphism and agency on social judgment in virtual environments. *Journal of Computer-Mediated Communication*, 9, 2, JCMC925.

[224] Kristine L. Nowak. 2015. Examining perception and identification in avatar-mediated interaction. *The handbook of the psychology of communication technology*, 87–114.

[225] Kristine L. Nowak and Frank Biocca. 2003. The effect of the agency and anthropomorphism on users’ sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 12, 5, 481–494.

[226] Kristine L. Nowak and Jesse Fox. 2018. Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, 6, 30–53.

[227] Kristine L. Nowak, Mark A. Hamilton, and Chelsea C. Hammond. 2009. The effect of image features on judgments of homophily, credibility, and intention to use as avatars in future interactions. *Media Psychology*, 12, 1, 50–76.

[228] Kristine L. Nowak and Christian Rauh. 2005. The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. *Journal of Computer-Mediated Communication*, 11, 1, 153–178.

[229] Adam O'Camb. 2018. Htc vive pro teardown. Retrieved 03/24/2022 from `https://ifixit.com/Teardown/HTC+Vive+Pro+Teardown/106064`.

[230] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. 2015. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, 405–415.

[231] Master of Shapes. 2021. Mobile room scale. Retrieved 03/24/2022 from `https://masterofshapes.com/work/cover-me/`.

[232] Jauvane C Oliveira, Xiaojun Shen, and Nicolas D Georganas. 2000. Collaborative virtual environment for industrial training and e-commerce. *IEEE VRTS*, 288.

[233] Yoshikazu Onuki and Itsuo Kumazawa. 2019. Reorient the gazed scene towards the center: novel virtual turning using head and gaze motions and blink. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1864–1871.

[234] Jason Orlosky, Peter Kim, Kiyoshi Kiyokawa, Tomohiro Mashita, Photchara Ratsamee, Yuki Uranishi, and Haruo Takemura. 2017. Vismerge: light adaptive vision augmentation via spectral and temporal fusion of non-visible light. In *Mixed and Augmented Reality (ISMAR), 2017 IEEE International Symposium on*. IEEE, 22–31.

[235] Jason Orlosky, Qifan Wu, Kiyoshi Kiyokawa, Haruo Takemura, and Christian Nitschke. 2014. Fisheye vision: peripheral spatial compression for improved field of view in head mounted displays. In *Proceedings of the 2Nd ACM Symposium on Spatial User Interaction* (SUI '14). ACM, Honolulu, Hawaii, USA, 54–61. ISBN: 978-1-4503-2820-3.

[236] Sergio Orts-Escolano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L Davidson, Sameh Khamis, Mingsong Dou, et al. 2016. Holoportation: virtual 3d teleportation in real-time. In *Proceedings of the 29th annual symposium on user interface software and technology*, 741–754.

[237] Ovrvision. 2016. Ovrision stereo camera. Retrieved 03/24/2022 from `http://ovrvision.com/`.

[238] Claudio Passos, Márcio H Da Silva, Antonio CA Mol, and Paulo VR Carvalho. 2017. Design of a collaborative virtual environment for training security agents in big events. *Cognition, Technology & Work*, 19, 2-3, 315–328.

[239] Jan Pečiva. 2007. *Active transactions in collaborative virtual environments*. PhD thesis. Faculty of Information Technology, University of Brno.

[240] Daniel Perez-Marcos, Maria V Sanchez-Vives, and Mel Slater. 2012. Is my hand connected to my body? the impact of body continuity and arm alignment on the virtual hand illusion. *Cognitive neurodynamics*, 6, 4, 295–305.

[241] Valeria I. Petkova and H. Henrik Ehrsson. 2008. If i were you: perceptual illusion of body swapping. *PLOS ONE*, 3, 12, 1–9.

[242] Thies Pfeiffer, Marc E Latoschik, and Ipke Wachsmuth. 2008. Evaluation of binocular eye trackers and algorithms for 3d gaze interaction in virtual reality environments. *JVRB-Journal of Virtual Reality and Broadcasting*, 5, 16.

[243] Thammathip Piumsomboon, Gun A. Lee, Jonathon D. Hart, Barrett Ens, Robert W. Lindeman, Bruce H. Thomas, and Mark Billinghurst. 2018. Mini-me: an adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18). Association for Computing Machinery, Montreal QC, Canada, 1–13. ISBN: 9781450356206.

[244] Iana Podkosova and Hannes Kaufmann. 2018. Co-presence and proxemics in shared walkable virtual environments with mixed colocation. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, 1–11.

[245] Domagoj Polančec and Igor Mekterović. 2017. Developing moba games using the unity game engine. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 1510–1515.

[246] Polyarc Games. 2020. Moss. Retrieved 01/18/2022 from `https://www.polyarcgames.com/`.

[247] Kirill Ragozin, Kai Kunze, Karola Marky, and Yun Suen Pai. 2020. Mazerunvr: an open benchmark for vr locomotion performance, preference and sickness in the wild. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI EA '20). Association for Computing Machinery, Honolulu, HI, USA, 1–8. ISBN: 9781450368193.

[248] Kasra Rahimi, Colin Banigan, and Eric D. Ragan. 2020. Scene transitions and teleportation in virtual reality and the implications for spatial awareness and sickness. *IEEE Transactions on Visualization and Computer Graphics*, 26, 6, 2273–2287.

[249] Stanislava Rangelova, Simon Flutura, Tobias Huber, Daniel Motus, and Elisabeth André. 2019. Exploration of physiological signals using different locomotion techniques in a vr adventure game. In *International Conference on Human-Computer Interaction*. Springer, 601–616.

[250] Sharif Razzaque, Zachariah Kohn, and Mary C Whitton. 2005. *Redirected walking*. Citeseer.

[251] Lisa Rebenitsch and Charles Owen. 2014. Individual variation in susceptibility to cybersickness. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (UIST '14). ACM, Honolulu, Hawaii, USA, 309–317. ISBN: 978-1-4503-3069-5.

[252] Clare Regan. 1995. An investigation into nausea and other side-effects of head-coupled immersive virtual reality. *Virtual Reality*, 1, 1, 17–31.

[253] Gary E Riccio and Thomas A Stoffregen. 1991. An ecological theory of motion sickness and postural instability. *Ecological psychology*, 3, 3, 195–240.

[254] Simon Riches, Soha Elghany, Philippa Garety, Mar Rus-Calafell, and Lucia Valmaggia. 2019. Factors affecting sense of presence in a virtual reality social environment: a qualitative study. *Cyberpsychology, Behavior, and Social Networking*, 22, 4, 288–292.

[255] Bernhard E Riecke, Bobby Bodenheimer, Timothy P McNamara, Betsy Williams, Peng Peng, and Daniel Feuereissen. 2010. Do we need to walk for effective virtual reality navigation? physical rotations alone may suffice. In *International Conference on Spatial Cognition*. Springer, 234–247.

[256] Bernhard E Riecke, David Clement, Denise Quesnel, Ashu Adhikari, Daniel Zielasko, and Markus von der Heyde. 2022. HyperJumping in Virtual Vancouver: Combating Motion Sickness by Merging Teleporting and Continuous VR Locomotion in an Embodied Hands-Free VR Flying Paradigm. en. In *Siggraph '22 Immersive Pavilion*. ACM, Vancouver, BC, Canada, 1–2. ISBN: 978-1-4503-9369-0.

[257] Sebastian Rings, Frank Steinicke, Bastian Dewitz, Fabian Büntig, and Christian Geiger. 2019. EXGAVINE - Exergames as Novel Form of Therapy in Virtual Reality for the Treatment of Neurological Diseases.

[258] Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz. 2007. Affective interactions using virtual reality: the link between presence and emotions. *Cyberpsychology & behavior*, 10, 1, 45–56.

[259] Kennedy Robert, Lane Norman, Berbaum Kevin, and Lilienthal Michael. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 203–220.

[260] Richard M. Royall. 1986. The effect of sample size on the meaning of significance tests. *The American Statistician*, 40, 4, 313–315.

[261] Roy A. Ruddle and Simon Lessels. 2009. The benefits of using a walking interface to navigate virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 16, 1, 5.

[262] Roy A. Ruddle, Ekaterina Volkova, and Heinrich H Bülthoff. 2011. Walking improves your cognitive map in environments that are large-scale and large in extent. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18, 2, 10.

[263] Shyam P. Sargunam, Kasra R. Moghadam, Mohamed Suhail, and Eric D. Ragan. 2017. Guided head rotation and amplified head rotation: evaluating semi-natural travel and viewing techniques in virtual reality. In *2017 IEEE Virtual Reality (VR)*, 19–28.

[264] Bhuvaneswari Sarupuri, Miriam Chipana, and Robert Lindeman. 2017. Trigger walking: a low-fatigue travel technique for immersive virtual reality. In 227–228.

[265] H. Scherer and S. Holtmann. 1983. Die beeinflussung der vestibulären untersuchung durch alkohol. *Laryngologie, Rhinologie, Otologie und ihre Grenzgebiete*, 62, 12, 558–560.

[266] Dieter Schmalstieg and Tobias Hollerer. 2016. *Augmented reality: principles and practice*. Addison-Wesley Professional.

[267] Ralph Schroeder. 2002. Copresence and interaction in virtual environments: an overview of the range of issues. In *Presence 2002: Fifth international workshop*, 274–295.

[268] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The experience of presence: factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10, 3, 266–281.

[269] Philip Schuchardt and Doug A Bowman. 2007. The benefits of immersion for spatial understanding of complex underground cave systems. In *Proceedings of the 2007 ACM symposium on Virtual reality software and technology*, 121–124.

[270] Martijn J. Schuemie, Peter Van Der Straaten, Merel Krijn, and Charles APG Van Der Mast. 2001. Research on presence in virtual reality: a survey. *Cyberpsychology & behavior*, 4, 2, 183–201.

[271] Jun'ichiro Seyama and Ruth S. Nagayama. 2007. The uncanny valley: effect of realism on the impression of artificial human faces. *Presence: Teleoper. Virtual Environ.*, 16, 4, 337–351. ISSN: 1054-7460.

[272] Ari Shapiro. 2011. Building a character animation system. In *International conference on motion in games*. Springer, 98–109.

[273] Craig R Sherman. 2002. Motion sickness: review of causes and preventive strategies. *Journal of travel medicine*, 9, 5, 251–256.

[274] Sigtrap Ltd. 2022. Vr tunneling pro. Retrieved 01/18/2022 from `http://www.sigtrapgames.com/VrTunnellingPro/html`.

[275] Richard Skarbez, Missie Smith, and Mary C. Whitton. 2021. Revisiting milgram and kishino's reality-virtuality continuum. *Frontiers in Virtual Reality*, 2. ISSN: 2673-4192.

[276] Mel Slater. 2002. Presence and the sixth sense. *Presence*, 11, 4, 435–439.

[277] Mel Slater and Anthony Steed. 2000. A Virtual Presence Counter. *Presence: Teleoperators and Virtual Environments*, 9, 5, 413–434. ISSN: 1054-7460.

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

[279] Mel Slater, Anthony Steed, and Martin Usoh. 1995. The virtual treadmill: a naturalistic metaphor for navigation in immersive virtual environments. In *Virtual Environments 95*. Springer, 135–148.

[280] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 3, 2, 130–144.

[281] Mel Slater, Martin Usoh, and Anthony Steed. 1995. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2, 3, 201–219.

[282] Diane H. Sonnenwald, Mary C. Whitton, and Kelly L. Maglaughlin. 2003. Evaluating a scientific collaboratory: results of a controlled experiment. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 10, 2, 150–176.

[283] Jan L. Souman, P. Robuffo Giordano, Martin Schwaiger, Ilja Frissen, Thomas Thümmel, Heinz Ulbrich, A De Luca, Heinrich H. Bülthoff, and Marc O. Ernst. 2011. Cyberwalk: enabling unconstrained omnidirectional walking through virtual environments. *ACM Transactions on Applied Perception (TAP)*, 8, 4, 25.

[284] Kathleen Späth, Sven Storck, Claudia Nass Bauer, and Frank Elberzhager. 2021. Fulfilling remote collaboration needs for new work. *Procedia Computer Science*, 191, 168–175. The 18th International Conference on Mobile Systems and Pervasive Computing (MobiSPC), The 16th International Conference on Future Networks and Communications (FNC), The 11th International Conference on Sustainable Energy Information Technology. ISSN: 1877-0509.

[285] Stephen Spielberg. 2018. *Ready Player One*. Warner Bros. Pictures.

[286] Misha Sra and Chris Schmandt. 2015. Metaspace: full-body tracking for immersive multiperson virtual reality. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, 47–48.

[287] Kay M. Stanney, Robert S. Kennedy, and Julie M. Drexler. 1997. Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society annual meeting* number 2. Volume 41. SAGE Publications Sage CA: Los Angeles, CA, 1138–1142.

[288] A. Steed, Y. W. Adipradana, and S. Friston. 2017. The ar-rift 2 prototype. In *2017 IEEE Virtual Reality (VR)*, 231–232.

[289] Anthony Steed, Sebastian Frlston, Maria Murcia Lopez, Jason Drummond, Ye Pan, and David Swapp. 2016. An 'in the wild'experiment on presence and embodiment using consumer virtual reality equipment. *IEEE transactions on visualization and computer graphics*, 22, 4, 1406–1414.

[290] Frank Steinicke. 2016. *Being really virtual*. Springer.

[291] Frank Steinicke, Gerd Bruder, Jason Jerald, Harald Frenz, and Markus Lappe. 2009. Estimation of detection thresholds for redirected walking techniques. *IEEE transactions on visualization and computer graphics*, 16, 1, 17–27.

[292] Frank Steinicke, Yon Visell, Jennifer Campos, and Anatole Lécuyer. 2013. *Human walking in virtual environments*. Springer.

[293] Neal Stephenson. 1992. *Snow Crash*. Bantam Books.

[294] Jonathan Steuer. 1992. Defining virtual reality: dimensions determining telepresence. *Journal of communication*, 42, 4, 73–93.

[295] Richard Stoakley, Matthew J Conway, and Randy Pausch. 1995. Virtual reality on a wim: interactive worlds in miniature. In *CHI*. Volume 95, 265–272.

[296] Tino Stoeckel, Robert Jacksteit, Martin Behrens, Ralf Skripitz, Rainer Bader, and Anett Mau-Moeller. 2015. The mental representation of the human gait in young and older adults. *Frontiers in Psychology*, 6, 943.

[297] Jennifer Stromer-Galley and Rosa Mikeal Martey. 2009. Visual spaces, norm governed places: the influence of spatial context online. *New Media & Society*, 11, 6, 1041–1060.

[298] Evan A. Suma, Zachary Lipps, Samantha Finkelstein, David M. Krum, and Mark Bolas. 2012. Impossible spaces: maximizing natural walking in virtual environments with self-overlapping architecture. *IEEE Transactions on Visualization and Computer Graphics*, 18, 4, 555–564.

[299] Un-Jae Sung, Jae-Heon Yang, and Kwang-Yun Wohn. 1999. Concurrency control in ciao. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*. IEEE, 22–28.

[300] Michiru Tamai, Mitsuyuki Inaba, Koichi Hosoi, Ruck Thawonmas, Masayuki Uemura, and Akinori Nakamura. 2011. Constructing situated learning platform for japanese language and culture in 3d metaverse. In *2011 Second International Conference on Culture and Computing*, 189–190.

[301] Rana M Tamim, Robert M Bernard, Eugene Borokhovski, Philip C Abrami, and Richard F Schmid. 2011. What forty years of research says about the impact of technology on learning: a second-order meta-analysis and validation study. *Review of Educational research*, 81, 1, 4–28.

[302] Desney S. Tan, Darren Gergle, Peter G. Scupelli, and Randy Pausch. 2004. Physically large displays improve path integration in 3d virtual navigation tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '04). ACM, Vienna, Austria, 439–446. ISBN: 1-58113-702-8.

[303] Unity Technologies. 2016. Unet networking layers. Retrieved 03/24/2022 from `https://docs.unity3d.com/Manual/UNetUsingHLAPI.html`.

[304] Unity Technologies. 2021. Unity market share. Retrieved 03/24/2022 from `https://unity.com/our-company`.

[305] Theophilus Teo, Louise Lawrence, Gun A Lee, Mark Billinghurst, and Matt Adcock. 2019. Mixed reality remote collaboration combining 360 video and 3d reconstruction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, 1–14.

[306] Thomas Terkildsen and Guido Makransky. 2019. Measuring presence in video games: an investigation of the potential use of physiological measures as indicators of presence. *International journal of human-computer studies*, 126, 64–80.

[307] Miguel A Teruel, Elena Navarro, Pascual González, Vıictor López-Jaquero, and Francisco Montero. 2016. Applying thematic analysis to define an awareness interpretation for collaborative computer games. *Information and Software Technology*, 74, 17–44.

[308] Zachary O Toups, Andruid Kerne, and William A Hamilton. 2011. The team coordination game: zero-fidelity simulation abstracted from fire emergency response practice. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18, 4, 1–37.

[309] Tanh Quang Tran, Holger Regenbrecht, and Minh-Triet Tran. 2019. Am i moving along a curve? a study on bicycle traveling-in-place techniques in virtual environments. In *IFIP Conference on Human-Computer Interaction*. Springer, 345–363.

[310] Matthias Trapp, Christian Beesk, Sebastian Pasewaldt, and Jürgen Döllner. 2011. Interactive rendering techniques for highlighting in 3d geovirtual environments. In *Advances in 3D Geo-Information Sciences*. Springer, 197–210.

[311] Sam Tregillus and Eelke Folmer. 2016. Vr-step: walking-in-place using inertial sensing for hands free navigation in mobile vr environments. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1250–1255.

[312] Michel Treisman. 1977. Motion sickness: an evolutionary hypothesis. *Science*, 197, 4302, 493–495.

[313] Daria Tsoupikova, Kristen Triandafilou, Saumya Solanki, Alex Barry, Fabian Preuss, and Derek Kamper. 2016. Real-time diagnostic data in multi-user virtual reality post-stroke therapy. In *SIGGRAPH ASIA 2016 VR Showcase*, 1–2.

[314] TurkCheeps. 2014. Snowman avatar. Retrieved 11/23/2020 from `https://assetstore.unity.com/packages/3d/props/cute-snowman-12477`.

[315] Unity. 2020. Asset store: steam vr. Retrieved 03/24/2022 from `https://assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647`.

[316] Unity Technologies. 2019. Unity standard assets: character package. Retrieved 01/18/2022 from `https://github.com/Unity-Technologies/Standard-Assets-Characters`.

[317] Unity Technologies. 2020. Xr interaction toolkit. Retrieved 11/20/2020 from `https://docs.unity3d.com/Packages/com.unity.xr.interaction.toolkit@1.0/manual/index.html`.

[318] User Interface Design. 2022. Attrakdiff. Retrieved 01/18/2022 from `http://www.attrakdiff.de/index-en.html/`.

[319] Martin Usoh, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P. Brooks Jr. 1999. Walking > walking-in-place > flying, in virtual environments. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 359–364.

[320] Brandon Van Der Heide, Erin M. Schumaker, Ashley M. Peterson, and Elizabeth B. Jones. 2013. The proteus effect in dyadic communication: examining the effect of avatar appearance in computer-mediated dyadic interaction. *Communication Research*, 40, 6, 838–860.

[321] Sarah F Van Der Land, Alexander P Schouten, Frans Feldberg, Marleen Huysman, and Bart van den Hooff. 2015. Does avatar appearance matter? how team visual similarity and member–avatar similarity influence virtual team performance. *Human Communication Research*, 41, 1, 128–153.

[322] Johannes Marinus Paulus Van Waveren. 2016. The asynchronous time warp for virtual reality on consumer hardware. In *Proceedings of the 22Nd ACM Conference on Virtual Reality Software and Technology* (VRST '16). ACM, Munich, Germany, 37–46. ISBN: 978-1-4503-4491-3.

[323] Roel Vertegaal. 1999. The gaze groupware system: mediating joint attention in multiparty communication and collaboration. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 294–301.

[324] VRChat. 2022. Avatars 3.0. Retrieved 01/18/2022 from `https://docs.vrchat.com/docs/avatars-30`.

[325] VRTK. 2022. Vrtk dragworld. Retrieved 01/18/2022 from `https://vrtoolkit.readme.io/docs/vrtk_dragworld`.

[326] Vuplex. 2020. 3d webview: cross-platform web browser for unity. Retrieved 03/24/2022 from `https://developer.vuplex.com/webview/overview`.

[327] Lana Wachowski and Lilly Wachowski. 1999. *The Matrix*. Warner Bros. Pictures.

[328] T Franklin Waddell and James D Ivory. 2015. It's not easy trying to be one of the guys: the effect of avatar attractiveness, avatar sex, and user sex on the success of help-seeking requests in an online game. *Journal of Broadcasting & Electronic Media*, 59, 1, 112–129.

[329] Thomas Waltemate, Dominik Gall, Daniel Roth, Mario Botsch, and Marc Erich Latoschik. 2018. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, 24, 4, 1643–1652.

[330] Isaac Wang, Pradyumna Narayana, Dhruva Patil, Gururaj Mulay, Rahul Bangar, Bruce Draper, Ross Beveridge, and Jaime Ruiz. 2017. Exploring the use of gesture in collaborative tasks. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (CHI EA '17). Association for Computing Machinery, Denver, Colorado, USA, 2990–2997. ISBN: 9781450346566.

[331] Mary C. Whitton, Joseph V. Cohn, Jeff Feasel, Paul Zimmons, Sharif Razzaque, Sarah J. Poulton, Brandi McLeod, and Frederick P. Brooks. 2005. Comparing ve locomotion interfaces. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.* IEEE, 123–130.

[332] Preston Tunnell Wilson, William Kalescky, Ansel MacLaughlin, and Betsy Williams. 2016. Vr locomotion: walking> walking in place> arm swinging. In *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry-Volume 1*, 243–249.

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

[334] Andrea Stevenson Won, Jeremy Bailenson, Jimmy Lee, and Jaron Lanier. 2015. Homuncular flexibility in virtual reality. *Journal of Computer-Mediated Communication*, 20, 3, 241–259.

[335] Uwe Wössner, Jürgen P. Schulze, Steffen P. Walz, and Ulrich Lang. 2002. Evaluation of a collaborative volume rendering application in a distributed virtual environment. In *EGVE*. Volume 2, 113–122.

[336] Jason Wuertz, Sultan A Alharthi, William A Hamilton, Scott Bateman, Carl Gutwin, Anthony Tang, Zachary Toups, and Jessica Hammer. 2018. A design framework for awareness cues in distributed multiplayer games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14.

[337] Mengxin Xu, Marıia Murcia-López, and Anthony Steed. 2017. Object location memory error in virtual and real environments. In *2017 IEEE Virtual Reality (VR)*. IEEE, 315–316.

[338] Nick Yee. 2014. *he Proteus paradox: How online games and virtual worlds change us- and how they don't.* Yale University Press.

[339] Nick Yee and Jeremy Bailenson. 2007. The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human Communication Research*, 33, 3, 271–290. ISSN: 0360-3989.

[340] Nick Yee and Jeremy Bailenson. 2009. The difference between being and seeing: the relative contribution of self-perception and priming to behavioral changes via digital self-representation. *Media Psychology*, 12, 2, 195–209.

[341] Yasuyoshi Yokokohji, Yusuke Sugawara, and Tsuneo Yoshikawa. 2000. Accurate image overlay on video see-through hmds using vision and accelerometers. In *Proceedings IEEE Virtual Reality 2000 (Cat. No.00CB37048)*, 247–254.

[342] Shanyang Zhao. 2003. Toward a taxonomy of copresence. *Presence: Teleoperators & Virtual Environments*, 12, 5, 445–455.

[343] Michael Zyda. 1999. *Networked virtual environments: design and implementation*. Addison-Wesley.

Appendices

Statistics

Perspective Study

Main Effect						
Effect	DFn	DFd	F	p	p<.05	η_p^2
Method	2	46	6.956	0.002	*	0.078
Environment Size	2	46	4.801	0.01	*	0.039
Method × Environment Size	4	92	0.254	0.907		0.003
Post-hoc Test for *Method* (Multiple Comparisons of Means: paired t test)						
Pair			t	Df	p	p<.05
Object Highlighting - View Cone			2.533	97.967	0.013	*
Object Highlighting - Video Mirror			3.439	79.554	0.001	*
View Mirror - View Cone			-2.016	127.71	0.045	*
Descriptive Statistics						
Method	N	Mean	SD	SE	Min	Max
Object Highlighting	216	4.806	6.856	0.808	1	34
View Cone	216	2.639	2.387	0.281	1	16
Video Mirror	216	1.944	1.686	0.198	1	9

Table 7.1: Perspective study Omnibus/Post-hoc and descriptive results for *Error Rate*.

Main Effect						
Effect	DFn	DFd	F	p	$p<.05$	η_p^2
Method	2	46	4.264	0.02	*	0.038
Environment Size	2	46	13.190	>0.001	*	0.105
Method × Environment Size	4	92	0.328	0.858		0.003
Post-hoc Test for *Method* (Multiple Comparisons of Means: paired t test)						
Pair			t	Df	p	$p<.05$
Object Highlighting - View Cone			1.039	137.29	0.301	
Object Highlighting - Video Mirror			3.145	126.74	0.002	*
View Mirror - View Cone			-1.583	113.59	0.116	
Descriptive Statistics						
Method	N	Mean	SD	SE	Min	Max
Object Highlighting	216	2.769	2.064	0.243	0.244	13.045
View Cone	216	2.374	2.489	0.293	0.151	16.057
Video Mirror	216	1.837	1.437	0.169	0.235	6.089

Table 7.2: Perspective study Omnibus/Post-hoc and descriptive results for *Precision*.

Main Effect						
Effect	DFn	DFd	F	p	$p<.05$	η_p^2
Method	2	46	1.012	0.372		0.012
Environment Size	2	46	20.315	>0.001	*	0.120
Method × Environment Size	4	92	0.892	0.472		0.009
Post-hoc Test for *Method* (Multiple Comparisons of Means: paired t test)						
Pair			t	Df	p	$p<.05$
Object Highlighting - View Cone			-1.347	142.0	0.179	
Object Highlighting - Video Mirror			-0.635	129.26	0.567	
View Mirror - View Cone			-0.909	129.59	0.365	
Descriptive Statistics						
Method	N	Mean	SD	SE	Min	Max
Object Highlighting	216	119.988	60.789	7.164	43.370	350.650
View Cone	216	133.608	60.481	7.128	58.280	302.024
Video Mirror	216	125.598	43.922	5.176	59.290	281.359

Table 7.3: Perspective study Omnibus/Post-hoc and descriptive results for *Time*.

VR Strider Study

Main Effect							
Effect	DFn	DFd	F	p	p<.05	η_p^2	ε
Method	3	45	12.454	0.000	***	0.454	0.752
Distance	1	15	1.567	0.230	.23	0.095	0.000
Angle	2	30	1.493	0.243	.24	0.091	0.809
Post-hoc Test for *Method* (Multiple Comparisons of Means: Tukey)							
Method Pair			Estimate	SE	t Value	$Pr(>\|t\|)$	p<.05
Bike FOn - Bike FOff			3.22	1.50	2.14	0.1415	
Joystick - Bike FOff			-5.61	1.50	-3.73	0.0012	**
Teleport - Bike FOff			-7.10	1.50	-4.72	<0.001	***
Joystick - Bike FOn			-8.83	1.50	-5.87	<0.001	***
Teleport - Bike FOn			-10.32	1.50	-6.87	<0.001	***
Teleport - Joystick			-1.49	1.50	-0.99	0.7539	
Descriptive Statistics							
Method	N	Mean	SD	SE	CI	Min	Max
Bike FOff	96	28.0	10.3	1.051	[25.92, 30.1]	2.63	52.8
Bike FOn	96	31.2	10.8	1.099	[29.05, 33.4]	3.69	56.9
Joystick	96	22.4	10.8	1.099	[20.22, 24.6]	2.63	44.4
Teleport	96	20.9	9.79	0.999	[18.93, 22.9]	7.12	55.1

Table 9.1: VR Strider study Omnibus/Post-hoc and descriptive results for *Time*.

Main Effect

Effect	DFn	DFd	F	p	$p<.05$	η_p^2	ε
Method	3	45	40.109	0.000	***	0.728	0.644
Distance	1	15	7.802	0.014	*	0.342	0.000
Angle	2	30	16.635	0.000	***	0.526	0.883

Post-hoc Test (Multiple Comparisons of Means: Tukey)

Method Pair	Estimate	SE	t Value	$Pr(>\|t\|)$	$p<.05$
Bike FOn - Bike FOff	-0.171	0.081	-2.12	0.15	
Joystick - Bike FOff	0.421	0.081	5.22	<0.001	***
Teleport - Bike FOff	0.432	0.081	5.37	<0.001	***
Joystick - Bike FOn	0.591	0.081	7.35	<0.001	***
Teleport - Bike FOn	0.603	0.081	7.49	<0.001	***
Teleport - Joystick	0.012	0.081	0.15	1.00	

Descriptive Statistics

Method	N	Mean	SD	SE	CI	Min	Max
Bike FOff	96	2.08	1.57	0.1607	[1.754, 2.4]	0.0402	6.83
Bike FOn	96	1.58	1.33	0.1357	[1.304, 1.85]	0.139	7.46
Joystick	96	3.42	2.10	0.2144	[2.991, 3.85]	0.237	7.67
Teleport	96	3.45	2.02	0.2059	[3.041, 3.86]	0.134	7.58

Table 9.2: VR Strider study Omnibus/Post-hoc and descriptive results for *Distance Error*.

Main Effect

Effect	DFn	DFd	F	p	p<.05	η_p^2	ε
Method	3	45	12.953	0.000	***	0.463	0.851
Distance	1	15	0.013	0.911		0.001	0.000
Angle	2	30	7.752	0.002	**	0.341	0.936

Post-hoc Test (Multiple Comparisons of Means: Tukey)

Method Pair	Estimate	SE	t Value	$Pr(> \lvert t \rvert)$	p<.05
Bike FOn - Bike FOff	-0.533	0.203	-2.63	0.044	*
Joystick - Bike FOff	1.008	0.203	4.97	<0.001	***
Teleport - Bike FOff	-0.219	0.203	-1.08	0.701	
Joystick - Bike FOn	1.541	0.203	7.60	<0.001	***
Teleport - Bike FOn	0.314	0.203	1.55	0.409	
Teleport - Joystick	-1.227	0.203	-6.05	<0.001	***

Descriptive Statistics

Method	N	Mean	SD	SE	CI	Min	Max
Bike FOff	96	16.4	10.7	1.091	[14.2, 18.6]	0.286	47.6
Bike FOn	96	12.5	9.41	0.9603	[10.63, 14.5]	0.268	44.7
Joystick	96	25.4	14.1	1.443	[22.48, 28.2]	1.96	48.8
Teleport	96	14.9	10.8	1.099	[12.71, 17.1]	0.0280	47.6

Table 9.3: VR Strider study Omnibus/Post-hoc and descriptive results for *Angle Error*.

Questionnaires

The following questionnaires were used in this thesis, whose items are summarized below.

AttrakDiff

7-point Likert scale (-3 - +3)

> Would you describe this technique rather with the left (-3) or right (+3) term?
>
> Pragmatic Quality
> - technical - human
> - complicated - simple
> - impractical - practical
> - cumbersome - straightforward
> - unpredictable - predictable
> - confusing - clearly structured
> - unruly - manageable
>
> Hedonic Quality - Identity
> - isolating - connective
> - unprofessional - professional
> - tacky - stylish
> - cheap - premium
> - alienating - integrating
> - separates me - brings me closer
> - unpresentable - presentable
>
> Hedonic Quality - Stimulation
> - conventional - inventive
> - unimaginative - creative
> - cautious - bold
> - conservative - innovative
> - dull - captivating
> - undemanding - challenging
> - ordinary - novel
>
> Attractiveness
> - unpleasant - pleasant
> - ugly - attractive
> - disagreeable - likeable
> - rejecting - inviting
> - bad - good
> - repelling - appealing
> - discouraging - motivating

Device Assessment Questionnaire

5-point Likert scale (1 - 5)

> - Required force
> - Smoothness of movements

- Mental effort
- Physical effort
- Difficulty to be accurate in movements
- Difficulty to be accurate in rotations
- Finger fatigue
- Wrist fatigue
- Arm fatigue
- Shoulder fatigue
- Neck fatigue
- Leg fatigue
- General comfort of usage

Igroup Presence Questionnaire

7-point Likert scale (1 - 7)

- In the computer generated world I had a sense of "being there".
- Somehow I felt that the virtual world surrounded me.
- I felt like I was just perceiving pictures.
- I did not feel present in the virtual space.
- I had a sense of acting in the virtual space, rather than operating something from outside.
- I felt present in the virtual space.
- How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?
- I was not aware of my real environment.
- I still paid attention to the real environment.
- I was completely captivated by the virtual world.
- How real did the virtual world seem to you?
- How much did your experience in the virtual environment seem consistent with your real world experience?
- How real did the virtual world seem to you?
- The virtual world seemed more realistic than the real world
- Stomach awareness
- Burping

NASA Task Load Index

20-point Likert scale (0 - 100 in increments of 5)

- Mental demand - How mentally demanding was the task?
- Physical demand - How physically demanding was the task?
- Temporal demand - How hurried or rushed was the pace of the task?
- Performance - How successful were you in accomplishing what you were asked to do?
- Effort - How hard did you have to work to accomplish your level of performance?
- Frustration - How insecure, discouraged, irritated, stressed and annoyed were you?

Networked Minds Social Presence Inventory

7-point Likert scale (1-7)

Co-presence

- I often felt as if (my partner) and I were in the same (room) together.
- I think (my partner) often felt as if we were in the same room together.

- was often aware of (my partner) in the (room).
- (My partner) was often aware of me in the (room).
- I hardly noticed (my partner) in the (room).
- (My partner) didn't notice me in the (room).
- I often felt as if we were in different places rather than together in same (room).
- I think (my partner) often felt as if we were in different places rather than together in the same (room).

Psycho-behavioral interaction: Perceived psychological engagement

- I paid close attention to (my partner).
- My partner) paid close attention to me.
- I was easily distracted from (my partner) when other things were going on.
- (My partner) was easily distracted from me when other things were going on.
- I tended to ignore (my partner).
- My partner) tended to ignore me.
- I was sometimes influenced by (my partner's) moods.
- (My partner) was sometimes influenced by my moods.
- When I was happy, (my partner) tended to be happy.
- When (my partner) was happy, I tended to be happy.
- When I was feeling sad (my partner) also seemed to be down.
- When (my partner) was feeling sad, (my partner) I tended to be sad.
- When I was feeling nervous, (my partner) also seemed to be nervous.
- When (my partner) was nervous, (my partner) I tended to be nervous.
- I was able to communicate my intentions clearly to (my partner).
- (My partner) was able to communicate their intentions clearly to me.
- My thoughts were clear to (my partner).
- (My partner's) thoughts were clear to me.
- I was able to understand what (my partner) meant.
- (My partner) was able to understand what I meant.

Psycho-behavioral interaction: Perceived behavioral interdependence

- My actions were often dependent on (my partner's) actions.
- (My partner's) actions were often dependent on my actions.
- My behavior was often in direct response to (my partner's) behavior.
- The behavior of (my partner) was often in direct response to my behavior.
- What I did often affected what (my partner) did.
- What (my partner) did often affected what I did.

Simulator Sickness Questionnaire

4-point Likert scale (0 - 3)

- General discomfort
- Fatigue
- Headache
- Eye strain
- Difficulty focussing
- Increased salivation
- Sweating
- Nausea
- Difficulty Concentrating
- Fullness of head
- Blurred vision
- Dizziness with eyes open
- Dizziness with eyes closed
- Vertigo
- Stomach awareness
- Burping

Slater-Usoh-Steed Presence Questionnaire

7-point Likert scale (0 - 6)

- Please rate your sense of "being in" the virtual environment, where 6 represents your normal experience of being in a place
- To what extent were there times during the experience when the virtual environment was the reality for you?
- Did you perceive the virtual environment more as images (0) that you saw or more as somewhere that you visited (6)?
- During the time of the experience, which was the strongest on the whole, your sense of being in the virtual environment (0) or of being elsewhere (6)?
- When you think back to the experience, do you remember it like other places you have been today?
- During the time of your experience, did you often think to yourself that you were actually in the virtual environment?

System Usability Scale

5-point Likert scale (1 - 5)

- I think that I would like to use this system again.
- I found the system unnecessarily complex.
- I thought the system was easy to use.
- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.
- I thought there was too much inconsistency in this system.
- I would imagine that most people would learn to use this system very quickly.
- I found the system very awkward to use.
- I felt very confident using the system.
- I needed to learn a lot of things before I could get going with this system.

I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated in the dissertation.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, May 20, 2022

(Jann Philipp Freiwald)