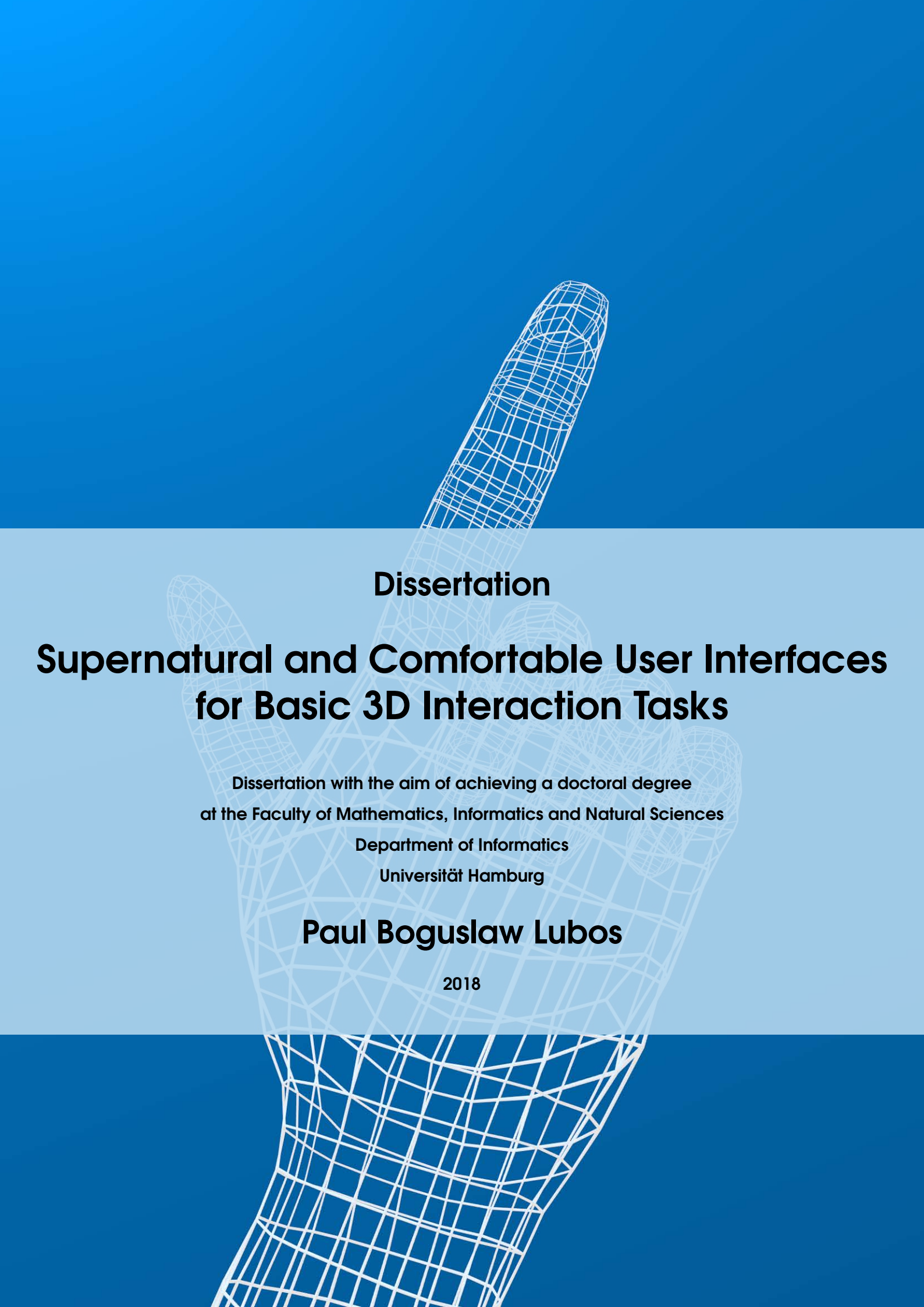


Dissertation

Supernatural and Comfortable User Interfaces for Basic 3D Interaction Tasks

Dissertation with the aim of achieving a doctoral degree
at the Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics
Universität Hamburg

Paul Boguslaw Lubos

2018

Supervisor & Reviewer:	Prof. Dr. Frank Steinicke
Reviewer:	Prof. Dr. Bernd Fröhlich
Head of examination commission:	Prof. Dr. Peter Kling
Deputy Head of examination commission:	Prof. Dr. Eva Bittner

Date of thesis defense: 18.09.2018

Abstract

The renewed interest in virtual reality (VR) in the last 5-10 years led to a necessity of the development of new 3D user interfaces (3DUIs) to interact with three-dimensional content. Mainly due to the vastly superior, affordable hardware, VR is on the way to become a common sight in our daily lives. 3DUIs still pose many challenges, such as distance misperception and the lack of guidelines on the placement or use of UI elements since many guidelines for 2DUIs are not necessarily valid in 3D. Technological advances allowed the reliable tracking of the user, compared to traditional discrete button input. These advances evoked the development of natural user interfaces (NUIs), which enable users to interact with computers using simple gestures, voice and other means previously reserved for inter-human communication. NUIs offer the advantage of being easier to learn for novice users, but a direct mapping of physical movements to VR can be physically tiring after short use. Supernatural user interfaces (SNUIs) are interfaces which are still inspired by the ways humans interact with one another or with their environment, but not limited by it. SNUIs permit actions which are not possible in the physical world. Examples would include teleportation or floating interface elements. Since virtual realities allow developers or users to set the rules within a world, the way users interact with virtual environments (VEs) can also be supernatural. Natural interaction can still inspire these interactions, however, they are less limited by natural constraints.

The goal of this thesis is to develop and evaluate supernatural basic interactions for tasks such as selection and travel within 3DUIs. The contributions in this thesis are split into three main parts. Within the first part, the contributions regarding performance in 3DUIs are shown. First, hover interactions for stereoscopic, head-tracked setups are evaluated. Based on user feedback, a HoverSpace above virtual objects is presented, allowing users to acquire additional information about objects. Afterwards, the interaction space for fully-immersive HMD environments analyzed, showing that distance misperceptions within the line-of-sight are the primary cause of selection errors 3DUIs.

These observations led to the investigation of the effects of comfort on 3DUIs. We show that during prolonged use, users are more efficient at direct 3D selection tasks when they have a more comfortable workspace. Using these results, adaptive 3DUIs are presented, which are always positioned around a user's arm joints, reducing the fatigue during extended tasks. Finally, a supernatural selection technique is evaluated, giving user's more than just two virtual hands, subdividing the interaction space and requiring fewer arm movements.
Within the third part, travel in VR is analyzed. Following the supernatural principle, instead of offering purely natural walking, three different flying setups are presented, one requiring the user to be seated and two where the user is suspended in the air. All three generate a high sense of presence, but all of them also increase cybersickness symptoms in participants. Thus, for further research, walking-based travel techniques were evaluated. Within a study, it was shown that redirected walking, although subconsciously manipulating participants, causes a significantly higher cognitive load. Despite that, a solution for the commonly observed reluctance of users to walk at a natural pace through VE is presented. This Safety-Sphere is using a distinctive, round area within the physical tracking space to provide users with an area without any manipulation. Users had to leave the sphere to travel to other places within the VE. There they were redirected at an optimal rate.

Additionally, this thesis presents the hardware developed throughout this thesis, a stereoscopic multi-touch table, and a low-resolution peripheral vision stimulation to enhance presence and increase the field-of-view (FOV) of head-mounted displays (HMDs).

Zusammenfassung

Das erneute Interesse an Virtual Reality (VR) in den letzten Jahren führte zur Entwicklung von 3D User Interfaces (3DUIs) zur Interaktion mit dreidimensionalen Inhalten. Vor allem wegen der inzwischen überlegenen, erschwinglichen Hardware ist VR nun auf dem Weg, ein alltäglicher Anblick zu werden. Bei der Entwicklung von 3DUIs gibt es weiterhin viele Herausforderungen, wie z. B. die Fehleinschätzung von Distanzen in VR und das Fehlen von Richtlinien für die Platzierung oder Verwendung von UI-Elementen, da viele Richtlinien für 2DUIs nicht unbedingt in 3D gültig sind. Technologische Fortschritte ermöglichten das zuverlässige Tracken des Benutzers im Vergleich zur herkömmlichen diskreten Tasteneingabe. Dies führte zur Entwicklung von natürlichen Benutzerschnittstellen (NUIs), die es Benutzern ermöglichen, mit einfachen Gesten, Sprache und anderen Mitteln, die zuvor für die zwischenmenschliche Kommunikation reserviert waren, mit Computern zu interagieren. NUIs bieten den Vorteil, dass sie für Anfänger einfacher zu erlernen sind. Eine direkte Zuordnung von physischen Bewegungen zu VR kann jedoch nach kurzem Gebrauch anstrengend sein. In Anbetracht dessen, wie körperlich anstrengend NUIs sein können, sollen neue, übernatürliche Interaktionen die langfristige Nutzung von VR ermöglichen. Da VR es Entwicklern oder Benutzern erlaubt, die Regeln in einer Welt festzulegen, kann die Art, wie Benutzer mit virtuellen Umgebungen (VEs) interagieren, auch übernatürlich sein. Diese Interaktionen können immer noch durch natürliche menschlichen Interaktion inspiriert sein, sind jedoch weniger durch Naturgesetze eingeschränkt.

Das Ziel dieser Arbeit ist es, übernatürliche Interaktionen für die Selektion in 3DUIs und das Reisen innerhalb von VR zu entwickeln und zu evaluieren. Im Einzelnen sind die Beiträge dieser Arbeit in drei Hauptteile aufgeteilt. Im ersten Teil werden die Beiträge zur Performanz in 3DUIs gezeigt. Zunächst werden Hover-Interaktionen für stereoskopische Head-Tracking-Setups ausgewertet. Basierend auf Benutzerfeedback wird ein HoverSpace über virtuellen Objekten dargestellt, der es Benutzern ermöglicht, zusätzliche Informationen über Objekte zu erhalten. Anschließend wurde der Interaktionsraum für voll-immersive HMD-Umgebungen analysiert, der zeigt, dass Distanz-Fehlwahrnehmungen innerhalb der Sichtlinie die Hauptursache für 3DUIs sind.

Diese Beobachtungen führten zur Untersuchung der Auswirkungen von physischer Anstrengung auf 3DUIs. Es wird gezeigt, dass Benutzer bei längerem Gebrauch effizienter bei direkten 3D-Selektionsaufgaben sind, wenn sie einen komfortableren Arbeitsbereich haben. Auf Basis dieser Ergebnisse werden adaptive 3DUIs vorgestellt, deren Position auf den Armgelenken eines Benutzers basieren und die Ermüdung bei langen Aufgaben reduzieren. Schließlich wird eine übernatürliche Selektionstechnik evaluiert, die dem Benutzer mehr als nur zwei virtuelle Hände gibt, den Interaktionsraum untergliedert und weniger Armbewegungen erfordert.

Im dritten Teil wird Bewegung durch VEs analysiert. Dem Supernatural UI Prinzip folgend, werden zunächst, statt rein natürliches Gehen, drei verschiedene Flugtechniken präsentiert, eine im Sitzen und zwei in einer Aufhängung von der Decke. Bei allen wird ein hohes Präsenzgefühl erzeugt, aber auch eine Erhöhung der Cyberssickness-Symptome bei den Teilnehmern. Für weitere Studien wurden daher Reisetechniken auf Basis des natürlichen Laufens evaluiert. In einer Studie wurde gezeigt, dass Redirected Walking, obwohl es die Teilnehmer nur unterbewusst manipuliert, eine signifikant höhere kognitive Belastung verursacht. Außerdem wird eine Lösung, für die Angst von Nutzern natürlich in VR zu gehen, präsentiert. Diese Safety-Sphere verwendet einen klar unterscheidbaren runden Bereich innerhalb des physischen Tracking-Bereichs, um Benutzern einen Bereich ohne jegliche Manipulation zu bieten. Um zu anderen Orten in der VE zu reisen, verlassen Benutzer die Sphäre und werden mit einer optimalen Rate umgeleitet.

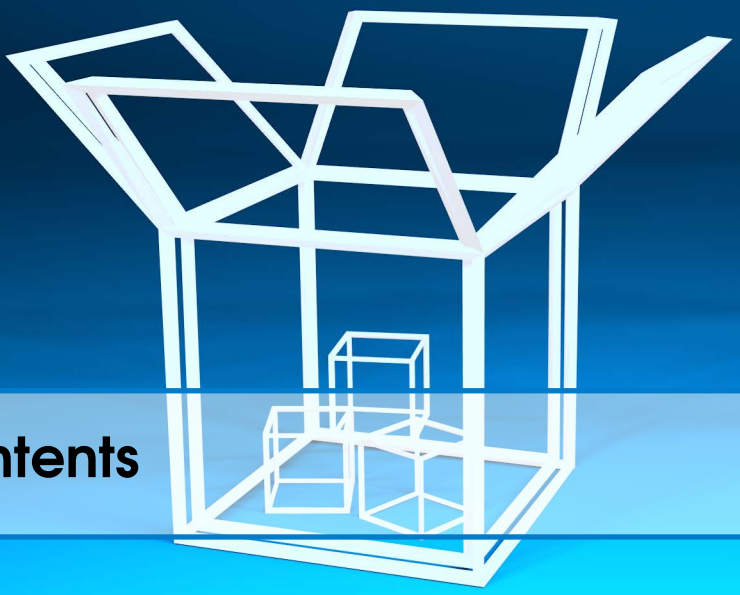
Darüber hinaus präsentiert diese Arbeit die entwickelte Hardware: einen stereoskopischen Multitouch-Tisch und ein periphäres Display mit niedriger Auflösung als Erweiterung für Head-Mounted Displays, um die Präsenz zu erhöhen und das Sichtfeld zu vergrößern.

Acknowledgments

Within this dissertation, my research at the Human-Computer Interaction group at the Department of Informatics of the Universität Hamburg and the work at the Immersive Media Group of the Department of Computer Science of the University of Würzburg are presented. I'm grateful to Prof. Dr. Frank Steinicke, the head of my research group, for his support and guidance during these last few years. He taught me a lot about academia and always had time and valuable input for my research. I would also like to thank Prof. Dr. Gerd Bruder for actively supporting and guiding me especially during the first stages of my academic career, and for introducing me to statistical analysis. Without the support of these two, none of this research would have been possible. I would like to thank the co-authors of my papers for working with me and would like to thank the research group for always providing valuable feedback and active participation in experiments, and the overall friendly, productive work atmosphere. You were always there for me when I needed help with something. I would also like to thank all students and other volunteers who participated in the many experiments done throughout this thesis. Special thanks go to all the student co-authors of some of the research included in this thesis. Thanks also go to everyone who took the time to proofread this thesis.

Thanks go to my friends and family, who helped and supported me throughout these last years, despite me being often busy with work and talking almost exclusively about work. I'm grateful for my parents, who introduced me to computer sciences and always had my back. I would like to thank my best friends, Tobias Lirsch and David Guss for their encouragement and for continuously supporting me.

Finally, I would like to thank my fiancée Sonja Aulfes for her support and love, which undoubtedly helped preserve my sanity during stressful times.



Contents

1 Introduction 11

1.1 Motivation 11

1.2 Roadmap 15

1.3 List of Publications 15

2 Fundamentals and Related Work 19

2.1 Terminology 19

2.2 Mixed Reality Continuum 20

2.3 Roots and History of 3DUIs 21

2.3.1 Multi-Touch Interfaces 23

2.3.2 3D User Interfaces 24

2.4 Human Factors 27

2.4.1 Information processing 27

2.4.2 Perception 29

2.4.3 Cognition 36

2.4.4 Motor Action & Ergonomics 37

2.5 3D Input & Output Devices 40

2.5.1 Input 40

2.5.2 Output 43

2.6 3DUI Basics and Related Work 47

2.6.1 Selection & Manipulation 48

2.6.2 Navigation 56

2.7 Summary 63

3 Increasing 3DUI Selection Performance 65

3.1 Hover Selection 65
3.1.1 Perceptual Experiment 67
3.1.2 The HoverSpace 70
3.1.3 Confirmatory Experiment 72
3.1.4 Results 74
3.1.5 Discussion 76
3.1.6 Conclusion 77

3.2 Direct Selection 78
3.2.1 Experiment 79
3.2.2 Results 82
3.2.3 Discussion 86
3.2.4 Conclusion 87

3.3 Summary 87

4 Improving Comfort for 3D Interaction 89

4.1 Comfort and 3D Task Performance 89
4.1.1 Pre-Study 90
4.1.2 Experiment 93
4.1.3 Results and Discussion 96
4.1.4 Conclusion and Future Work 100

4.2 Joint-Centered User Interfaces 101
4.2.1 Joint-Centered Interaction 102
4.2.2 Experiment 104
4.2.3 Confirmatory Study 110
4.2.4 Conclusion 114

4.3 More Than Two Hands 115
4.3.1 Quadmanual User Interfaces 115
4.3.2 Experiment 116
4.3.3 Multi-manual User Interfaces 119
4.3.4 Experiment 120
4.3.5 Discussion 124
4.3.6 Conclusion 124

4.4 Summary 125

5 Enhancing Traveling and Exploration of Virtual Worlds 127

5.1 Leaning Based Flying 127
5.1.1 Dragon Rider Flight Technique 128
5.1.2 Experiment 129
5.1.3 Conclusion 132

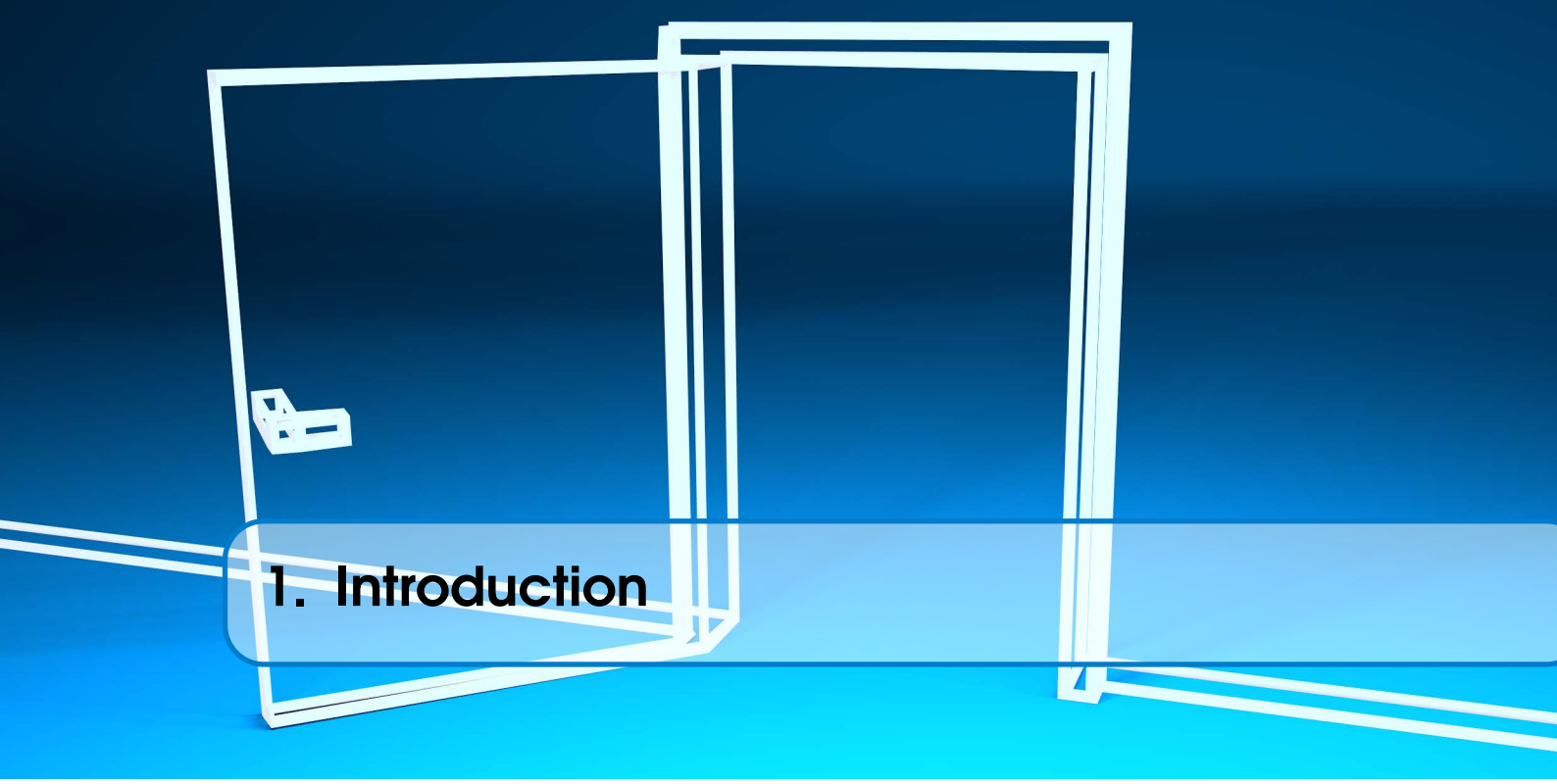
5.2 Suspended Flying 132
5.2.1 System Setup 133
5.2.2 Flight Control 135
5.2.3 Experiment 136
5.2.4 Second Iteration 140
5.2.5 Conclusion 146

5.3 Natural Walking 146
5.3.1 Cognitive Demands . . . 147
5.3.2 Safety Sphere . . . 153
5.4 Presence Through Peripheral Stimulation 160
5.4.1 Related Work . . . 161
5.4.2 Ambiculus Peripheral Display Extension . . . 161
5.4.3 Experiment . . . 162
5.4.4 Conclusion . . . 166
5.5 Summary 166

6 Conclusion . . . 167

Bibliography . . . 171
Books 171
Electronic 188

A Appendices . . . 191
A.1 Hardware 191
A.1.1 iSPACE . . . 191
A.1.2 Hapring . . . 194
A.2 Questionnaires 198
A.2.1 SSQ . . . 198
A.2.2 TLX . . . 199
A.2.3 Borg . . . 200
A.2.4 SUS-PQ . . . 201
A.2.5 Immersive Tendencies Questionnaire . . . 202
A.2.6 WS-PQ . . . 203



1. Introduction

"The ultimate display would, of course, be a room within which the computer can control the existence of matter."

—Ivan Sutherland, 1965

1.1 Motivation

This vision of an immersive computer-generated environment which would be indistinguishable from the real world has been introduced in Ivan Sutherland's essay *The Ultimate Display* in 1965 [Sut65].

Immersive technologies gained importance over the last years, and several new technologies have been presented. These include but are not limited to mixed reality (MR), virtual reality (VR) and augmented reality (AR) devices. For instance, with head-mounted displays combined with tracking systems, it becomes possible to fully immerse a user into a virtual environment and decouple a user from the real world [Ste16].

There is an enormous academic, industrial, economic and social interest in VR. According to DigiCapital, the VR and AR market will create 120 billion revenue USD by 2020 [Dig16]. Throughout the last years, beginning with Palmer Luckey's successful crowdfunding campaign to fund the Oculus Rift head-mounted display (HMD)[1], the public awareness of the fields of VR and AR rose. However, from an academic perspective, VR has been in the interest of many research institutes and companies for a long time.

The first pioneering technological and research contributions towards the ultimate display were made by Ivan Sutherland in 1968, who developed the first AR prototype of its type, the so-called *Sword of Damocles* [Sut68]. While VR has been developed by researchers since then and has been used by research institutions such as the National Aeronautics and Space Administration (NASA) [Fis+87], Massachusetts Institute of Technology (MIT) [Nai06], University of Northern

[1] https://www.kickstarter.com/projects/1523379957/oculus-rift-step-into-the-game

Carolina [Bro+90], University of Connecticut [KGH85] and at the US Air Force [Fur86], it remained a niche field until the 90s. Then, VR has received considerable media attention by public audiences. Companies like W Industries (later called Virtuality), VPL research, Microvision, as well as Nintendo presented research and consumer VR products. Also, different form factors towards the vision of the ultimate display were made many advancements were made during this hype period, such as the development of the CAVE automatic virtual environment (CAVE) [Cru+92], VPL Research's EyePhone [Tei90] and DataGlove [Zim+87].

However partly due to the hardware inadequacies, VR was not successful on the consumer market [Kus14].

VR technology was not able to fulfill the promises as a new interface paradigm for novel human-computer interaction (HCI) in contrast to other technologies. In particular, smartphones dominated the way how humans interacted with computers around the millennium. Interestingly, the advancements in the smartphone sector led to today's technology [Ols+11]. Aspects, such as GPU performance, resolution, tracking accuracy, etc. are all based on advancements in the consumer market, particularly the smartphone market. With the current popularity, commercial companies begin adapting and developing concepts for the user interfaces since users are interested in interactive Virtual Environments (VEs) instead of passive experiences.

In the long history of VR, several researchers and developers have introduced different definitions [Ste16]. In the scope of this thesis, we will follow the definition of Fred Brooks [Bro94], summarized by [Ste16].

Definition 1.1.1 Virtual reality requires three features real-time rendering with **viewport changes** as the head or body moves, **real space** that is either concrete or abstract virtual environments and three real interactions that are possible direct **selection** or **manipulation** of virtual objects.

Due to the immersive nature of VR, which allows users to explore virtual environments naturally, the many two-dimensional design guidelines for user interfaces might not be appropriate anymore. In contrast, novel 3D user interfaces (3DUIs) should be analyzed, designed, developed and evaluated for this new technology. Although there has been a lot of research in this field, there are still many open research challenges. In particular, technological and perceptual issues, such as distance misperceptions estimations [Ste+09], hardware limitations such as the limited field-of-view [Jon+12], the lack of passive haptic feedback [BSS13a; BSS13c; Cha+10] and the lack of guidelines to improve the comfort of such VR applications [LaV+17], which affect the overall VR user experience and performance.

Furthermore, VR requires different input devices, as traditional 2D devices like a mouse and keyboard are often inadequate and burdensome in VR. Novel interaction devices and concepts were developed, such as Zimmermann et al.'s Data Glove [Zim+87], Fröhlich's Cubic Mouse [Frö+00] and more recently the Leap Motion [Lea12], with the goal of providing the user means to interact more naturally with the virtual world and with more degrees of freedom.

In HCI research, there has been increased interest in developing UIs which feel "natural", meaning that the interaction with virtual content should be similar to the interaction with real, physical objects [WW11].

Ortega et al. argue that the term *Natural User Interfaces* (NUIS) is a buzzword and many implementations of NUIs are hardly natural, partly because of the lack of tactile feedback [Ort+16]. Norman argues that NUIs based on gestures might be useful and will find a niche, but classical UI elements will be necessary for certain aspects, for example, to provide the user with feedback after executing a gesture [Nor10]. Finally, Wigdor and Wixon see NUIs are a design principle that is meant to devise interfaces which evoke a feeling of being at home and allow interaction which feels natural [WW11]. Arguably, NUIs are based on intuitive physics [Hv12], meaning the inherent

understanding of real-world phenomena [Ste17].

In our research group we follow the following definition for NUIs:

Definition 1.1.2 A user interface is called **natural**, when it is based on natural human mechanisms, such as grabbing, walking, touching, speaking, looking [Ste17].

While exploiting natural mechanisms when humans interact with technology appears to make sense, the virtual world is not limited by physical world constraints, and thus developers can extend the interaction possibilities in VR. For example, users might interact with distant objects outside their arms' reach, objects might float in space, or users might teleport instantly.

Definition 1.1.3 A user interface is called **supernatural**, when it is based on natural human mechanisms, but not limited to laws of physics or real-world constraints [Ste17].

SNUIs are a promising research field, but there are particular open challenges for these paradigms. The goal of this thesis is to address some of these challenges and contribute in particular to the following aspects.

Selecting Things

Maybe one of the most elementary tasks in VR is the selection of objects or actions [LaV+17]. In VR, an obvious approach for selection would be to grab the intended object with the virtual hand directly. However, this introduces the real-world constraint of only allowing interactions in range. Supernatural interfaces also enable pointing at distant objects, e.g., through gaze or voice input. Additionally, they can enable the user to elongate their arms to reach objects which would usually be inaccessible without traveling, such as the Go-Go technique [Pou+96]. Some research has shown that performance in 2DUIs with a mouse is higher than with 3DUIs [TS08]. Other studies show that bimanual interaction can surpass mouse performance after a training period [Sch+12]. Thanks to advances in the hardware development, spatial gestural interaction is finally commercially available [Lea12]. While direct interaction might feel natural to users and perhaps be easy to learn for users with a similar background, it is necessary to ensure it offers comparable performance to traditional user interfaces. Otherwise, the interface might not be used in workspace environments which aim for the highest productivity. Mine et al. have shown that direct interaction offers performance benefits over indirect manipulation of objects at a distance [MBS97]. These advantages are not negligible, but most interactions are still prone to the unsolved technical problems of VR, such as distance misperception [BSS13a; BSS13c; Cha+10]. Also, the placement of interactive elements within a VE is not thoroughly investigated. Overcoming these performance challenges to allow long-term use of 3DUIs is another goal of this thesis.

Exploring Virtual Worlds

Finally, users will have to perform more tasks than selection and manipulation within a virtual environment and, depending on the task, might have to navigate within the environment [LaV+17] to reach objects which they can select and manipulate. While navigation in Virtual Reality has been thoroughly investigated in the past, there are still limitations which are not sufficiently solved. Natural walking is considered the preferred choice for navigation in immersive virtual environments (IVEs) [Uso+99a]. However it is limited by the physical environment, which is why techniques like *Redirected Walking* (RDW) were developed [Raz05], which restrict the user's actual movement to a limited physical space while allowing them to walk infinitely in the virtual environment through subtle camera manipulations. We investigate how cognitively demanding the redirection is and how to prevent the users from losing their feeling of presence due to fear of collisions with physical obstacles [BLS15; LBS14c]. However, walking is physically demanding, and for specific long-term tasks, it may not be the ideal solution. We devised leaning- and suspension-based navigation techniques and evaluated them in user studies [Dib+16; Kru+15b; Sch+16]. Another

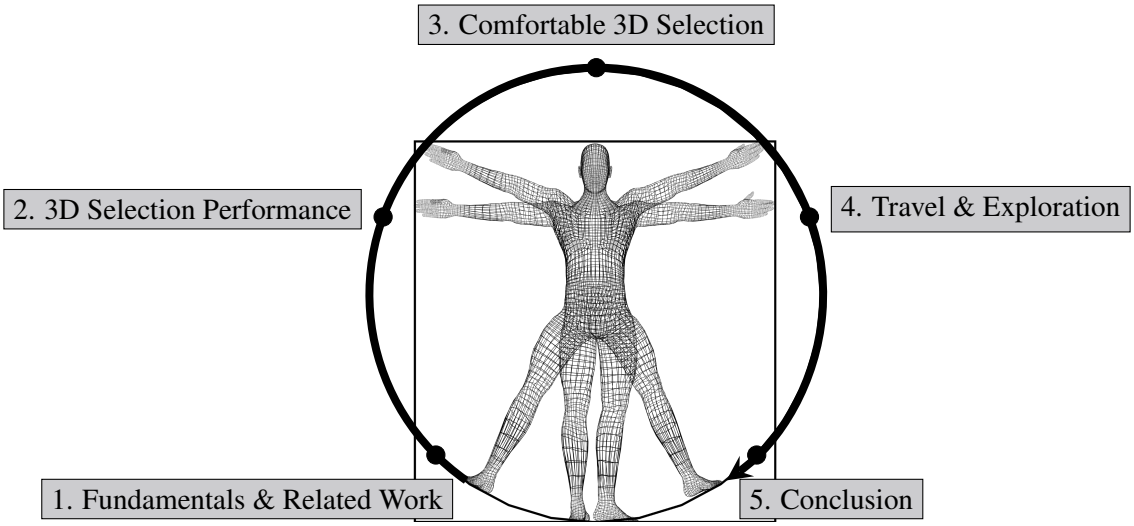


Figure 1.1: Illustration depicting da Vinci's Vitruvian man, showing the road map for the structure of this thesis.

such limitation is the often limited field of view (FOV) of the devices utilized to visualize the virtual environment, mostly head-mounted displays. While the displays are now widely commercially available, they only offer horizontal FOV up to 110 degrees [Ocu13], which is much lower than the human FOV [Pal99]. This hardware limitation can cause misperceptions [Bol+10; Jon+12; JSB13] as well as decrease the sense of presence the users feel [Lin+02]. We present a solution which increases the available FOV in HMDs by including a low-resolution peripheral display [Lub+16a].

Interacting Comfortably

Many of today's VR applications provide users with the ability to perform visually impressive supernatural interactions in 3D mid-air, in which artificially floating objects can be grabbed or manipulated. However, such interactions require exhaustive arm movements which might lead to the well-known *gorilla-arm syndrome* [Com17; Jan+17b].

Despite the renewed interest in the field, many problems remain unsolved, and the VR solutions are not viable for long-term use because they do not factor in the user's comfort, although user comfort has been described as one of the critical issues for 3DUI design [LaV+17]. Current interfaces are often optimized for being easy to use and explain, which is desirable and adequate for demonstrations and short-term research but does not factor in the requirements of long-term use. Especially commercial solutions will be expected to be usable for long time periods, comparable to the traditional desktop setups known in offices nowadays. Thus, one of the primary goals of this thesis is to offer ways to alleviate these restrictions of current 3DUIs.

To summarize, the contributions of this thesis address the following:

- Novel 3DUIs for long-term or comfortable use.
- Selection performance improvements through perceptually-inspired 3DUIs.
- Supernatural travel techniques for exploration of virtual worlds.

Thus, the goal of this is thesis is to investigate these concerns and alleviate the shortcomings of current 3DUIs by through thorough analysis and new interaction concepts. The main contributions of this thesis have been published in conferences, journals and at workshops and this thesis aims to put them in the wide context of the improvement of 3DUIs.

1.2 Roadmap

As illustrated in Figure 1.1, the thesis is structured as follows. In Chapter 2, we discuss the scientific background and related work of this thesis. In Chapter 3, we investigate hover selections for stereoscopically displayed targets on a multi-touch tabletop system [Lub+15; Lub+14b]. Afterward, mid-air selection performance in HMD environments is examined, based on [LBS14a].

In Chapter 4, the topic of comfort and the validity of the long-term use of 3DUIs is discussed. Results of a user study are presented which showed the influence of user comfort on long-term performance during common selection tasks [LBS15]. Afterward a supernatural interaction concept which allows users to utilize more than just two hands in 3DUIs is presented, and its selection performance evaluated [LBS14b]. Finally, Joint-Centered User Interfaces are introduced - a concept of adaptively positioning UI elements comfortably around the user's physically rested or supported joints, on so-called Kinespheres [Lub+16b]. These results allow the design of 3DUIs for comfortable long-term use.

In Chapter 5, an overview of the current state of travel in Virtual Reality is given, and we describe works concerned with the enhancement of redirected walking [BLS15; LBS14c]. We also describe alternative, leaning-based travel metaphors [Dib+16; Kru+15b; Sch+16]. Finally, we try to improve the user experience during navigation by improving upon the sense of presence, by presenting an easy-to-build extension to common HMDs [Lub+16a] that increases the FOV by adding low resolution peripheral optical stimulation. Chapter 6 concludes this thesis.

1.3 List of Publications

For the following publications, most of the implementation, experiment design and conduction and paper writing have been done by myself.

[Lub+15] **Paul Lubos**, Oscar Ariza, Gerd Bruder, Florian Daiber, Frank Steinicke, and Antonio Krüger. "HoverSpace". In: *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT)*. Edited by Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler. Cham: Springer International Publishing, 2015, pages 259–277. ISBN: 978-3-319-22698-9. DOI: 10.1007/978-3-319-22698-9_17. URL: http://dx.doi.org/10.1007/978-3-319-22698-9_17 (cited on pages 15, 65).

[Lub+14a] **Paul Lubos**, Rüdiger Beimler, Markus Lammers, and Frank Steinicke. "Touching the Cloud: Bimanual annotation of immersive point clouds". In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. Mar. 2014, pages 191–192. DOI: 10.1109/3DUI.2014.6798885 (cited on pages 42, 49, 193).

[Lub+16a] **Paul Lubos**, Gerd Bruder, Oscar Ariza, and Frank Steinicke. "Ambiculus: LED-based Low-resolution Peripheral Display Extension for Immersive Head-mounted Displays". In: *Proceedings of the ACM Virtual Reality International Conference (VRIC)*. VRIC '16. Laval, France: ACM, 2016, 13:1–13:4. ISBN: 978-1-4503-4180-6. DOI: 10.1145/2927929.2927939. URL: http://doi.acm.org/10.1145/2927929.2927939 (cited on pages 14, 15, 59, 127).

[Lub+16b] **Paul Lubos**, Gerd Bruder, Oscar Ariza, and Frank Steinicke. "Touching the Sphere: Leveraging Joint-Centered Kinespheres for Spatial User Interaction". In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI '16. Tokyo, Japan: ACM, 2016, pages 13–22. ISBN: 978-1-4503-4068-7. DOI: 10.1145/2983310.2985753. URL: http://doi.acm.org/10.1145/2983310.2985753 (cited on pages 15, 34, 101).

[LBS14a] **Paul Lubos**, Gerd Bruder, and Frank Steinicke. "Analysis of Direct Selection in Head-Mounted Display Environments". In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2014, pages 1–8 (cited on pages 15, 78, 95, 100–102, 114, 118).

[LBS14b] **Paul Lubos**, Gerd Bruder, and Frank Steinicke. "Are 4 Hands Better Than 2?: Bimanual Interaction for Quadmanual User Interfaces". In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI '14. Honolulu, Hawaii, USA: ACM, 2014, pages 123–126. ISBN: 978-1-4503-2820-3. DOI: 10.1145/2659766.2659782. URL: http://doi.acm.org/10.1145/2659766.2659782 (cited on pages 15, 115, 119, 124).

[LBS14c] **Paul Lubos**, Gerd Bruder, and Frank Steinicke. "Safe-&-round: Bringing Redirected Walking to Small Virtual Reality Laboratories". In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI '14. Honolulu, Hawaii, USA: ACM, 2014, pages 154–154. ISBN: 978-1-4503-2820-3. DOI: 10.1145/2659766.2661219. URL: http://doi.acm.org/10.1145/2659766.2661219 (cited on pages 13, 15, 127, 147, 153).

[LBS15] **Paul Lubos**, Gerd Bruder, and Frank Steinicke. "Influence of Comfort on 3D Selection Task Performance in Immersive Desktop Setups". In: *Journal of Virtual Reality and Broadcasting* 12(2015).2 (2015). ISSN: 1860-2037. URL: http://nbn-resolving.de/urn:nbn:de:0009-6-41261 (cited on pages 15, 40, 89, 90, 108, 114).

[Lub+14b] **Paul Lubos**, Carina Garber, Anjuly Hoffert, Ina Reis, and Frank Steinicke. "The Interactive Spatial Surface - Blended Interaction on a Stereoscopic Multi-Touch Surface". In: *Mensch & Computer 2014 - Workshopband*. Edited by Andreas Butz, Michael Koch, and Johann Schlichter. Berlin: De Gruyter Oldenbourg, 2014, pages 343–346 (cited on pages 15, 42, 46, 47, 53, 67, 191, 194).

For the following publications critical parts of the implementation, experiment design and paper writing have been done by myself. However, the first author's contribution is substantial.

[ALS15] Oscar Javier Ariza Nunez, **Paul Lubos**, and Frank Steinicke. "HapRing: A Wearable Haptic Device for 3D Interaction". In: *Proceedings of Mensch und Computer*. Edited by Sarah Diefenbach, Niels Henze, and Martin Pielot. Berlin: De Gruyter Oldenbourg, 2015, pages 421–424 (cited on pages 41, 112, 114, 191, 195).

[BLS15] Gerd Bruder, **Paul Lubos**, and Frank Steinicke. "Cognitive Resource Demands of Redirected Walking". In: *IEEE Transactions on Visualization and Computer Graphics* 21.4 (Apr. 2015), pages 539–544. ISSN: 1077-2626. DOI: 10.1109/TVCG.2015.2391864 (cited on pages 13, 15, 127, 147, 148).

[Kru+15b] Dennis Krupke, **Paul Lubos**, Lena Demski, Jonas Brinkhoff, Gregor Weber, Fabian Willke, and Frank Steinicke. "Evaluation of Control Methods in a Supernatural Zero-Gravity Flight Simulator". In: *Proceedings of the GI-Workshop VR/AR*. 2015 (cited on pages 13, 15, 132).

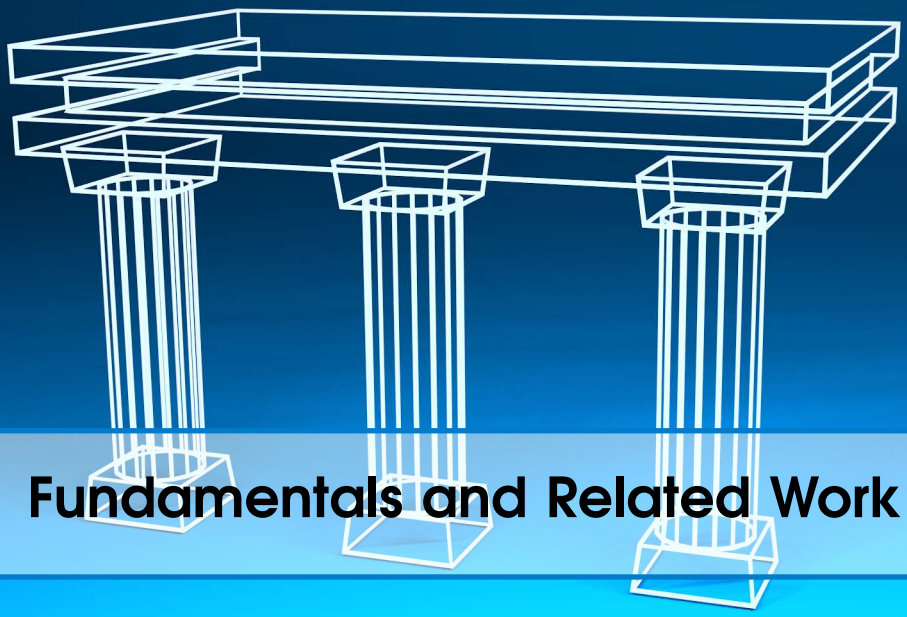
[Lan+17] Eike Langbehn, **Paul Lubos**, Gerd Bruder, and Frank Steinicke. "Bending the Curve: Sensitivity to Bending of Curved Paths and Application in Room-Scale VR". In: *IEEE Transactions on Visualization and Computer Graphics* 23.4 (Apr. 2017), pages 1389–1398. ISSN: 1077-2626. DOI: 10.1109/TVCG.2017.2657220 (cited on page 63).

The following publications are based on student projects which I have supervised. Crucial parts of the contribution, especially concerning the experiment design, implementation and paper writing were done by myself.

[Dib+16] Christian Dibbern, Julian Hettwer, Josephine Höltermann, Daniel Sinn, Sukhpreet Singh, **Paul Lubos**, and Gerd Bruder. "Virtual Reality Flight Interfaces inspired by Iron Man". In: *Proceedings of the GI Workshop on Virtual and Augmented Reality (GI VR/AR)*. Edited by Thies Pfeiffer, Julia Fröhlich, and Rolf Kruse(Hg.) Aachen: Shaker Verlag, 2016, pages 97–108 (cited on pages 13, 15, 127, 132).

[Hae+16] Steffen Haesler, Karen Obernesser, Tino Raupp, Christoph Jahnke, Jonathan Stapf, Julia Bräker, **Paul Lubos**, Gerd Bruder, and Frank Steinicke. "Edutainment & Engagement at Exhibitions: A Case Study of Gamification in the Historic Hammaburg Model". In: *Mensch und Computer 2016-Tagungsband* (2016) (cited on page 59).

[Sch+16] Britta Schulte, Pauline Bimberg, Julia Hertel, Franziska Neu, Philipp Heidenreich, Gerd Bruder, and **Paul Lubos**. "Evaluation of Two Leaning-Based Control Mechanisms in a Dragon Riding Interface". In: *Proceedings of the GI Workshop on Virtual and Augmented Reality (GI VR/AR)*. Edited by Thies Pfeiffer, Julia Fröhlich, and Rolf Kruse(Hg.) Aachen: Shaker Verlag, 2016, pages 109–116 (cited on pages 13, 15, 59, 127).



2. Fundamentals and Related Work

In this chapter, we will establish the required terminology (cf. Section 2.1), discuss the historical background (cf. Section 2.3), summarize aspects of human factors (cf. Section 2.4), as well as previous works regarding 3DUIs (cf. Section 2.6) which are in the context of this thesis.

2.1 Terminology

In this section, we introduce a common language since many terms are used as synonyms in research and media. To avoid confusion and allow using precise language, we define the seven most relevant terms. This terminology is an extended and restructured version of the taxonomy in [LaV+17]. The terminology will be used throughout this thesis.

Human-computer interaction (HCI) An interdisciplinary research field which is concerned with the connection between humans and technology, vice versa. Although the term computer is part of the name, the field is concerned with all kinds of technology, from computers to mobile phones, smart home devices, robot vacuums, car navigation systems, and 3D technology, such as the Kinect. The field is concerned with an understanding of how humans process information and use systems and how to use this knowledge from the field of psychology to improve technological systems [Dix09].

User interface (UI) The medium through which users communicate with technology is called a user interface. The key here is that the user's *input* has to be translated into a format usable by technology and the technology has to compute an *output* in a format understandable for users. The input can have many shapes, whether traditional through mouse and keyboard events or streamed point cloud data recorded through a depth camera from which the computer has to extract the user's skeleton or gestures. Similarly, output devices can stimulate many senses of a user, for example, high-frequency haptics output or just showing a picture on a screen [HH93].

Degrees of freedom (DOF) In this context, DOF refers explicitly to the count of independent dimensions of the motion of a body. A computer mouse, for example, can track its position on a 2D surface, resulting in 2 DOF. Rotating a mouse or lifting it off the surface cannot be tracked via the conventional sensors in a mouse. A controller of an HTC Vive can be tracked

in 6 DOF since the system measures a translation in all directions and a rotation along all axes. Additionally, virtual objects are often defined using a scaling along either a global or a local scaling along one of the three axes, resulting in another 3 DOF. When considering the buttons and additional scroll wheels, sometimes these other input methods are also counted towards the DOF. In the context of this thesis, the term DOF refers to the motion, meaning rotational and translational DOF and to the scaling of an object.

Direct interaction The interaction with virtual objects which does not require the use of body extensions or other metaphors. Instead, it allows the user to interact with objects by touching them. Nowadays, direct interaction is often focused on the interaction by touch. Section 2.3 gives an overview of how the definition of directness changed over the history since mouse input was considered to be direct interaction when it was new [HHN85], compared to abstract keyboard instructions, whereas nowadays it often refers to grabbing virtual objects with hand-tracking.

Natural user interface (NUI) NUIs describe UIs which are based on natural movements or gestures, such as touching an object with the fingers, grabbing it with both hands, or walking to travel [Nor10] (cf. Definition 1.1.2). They have been described as the next step, compared to graphical user interfaces (GUIs) and command line interfaces (CLIs) [Rey08]. NUI has hence been used in marketing although there has been dissent about whether it is correct or not [Nor10].

3D interaction According to Bowman et al., an HCI interaction is a 3D interaction if it is performed directly in a real or virtual 3D spatial context [LaV+17]. Here it is crucial that not exclusively traditional 2D UI elements have been used to change the 3D output. At least one action had to be performed directly in the 3D space. For example, when building a virtual world with a modern tool, usually there is a possibility to display a tree view of the current scene and its contents. Users can interact via this traditional 2D tree view, e.g., to select an object within the scene. Although the contents of the scene may all be 3D objects, this kind of interaction could not be referred to as a 3D interaction. However, in case the user directly clicks on an object in the 3D view to select it, it could be considered as a 3D interaction since the 2D mouse cursor coordinates have to be transformed into scene coordinates, for example, using a kind of raycast technique [LaV+17].

3D user interface (3DUI) A 3D UI is a UI that involves 3D interaction. As such, the utilized interactions need not be exclusively 3D interactions, as long as there are 3D interactions involved [LaV+17].

2.2 Mixed Reality Continuum

To better classify the technology which is utilized throughout the thesis, we consider the Mixed Reality Continuum [MK94]. Milgram and Kushino coined the *Mixed Reality Continuum* as a scale ranging from physical reality to VR, including technologies like *Augmented Reality* (AR) and *Augmented Virtuality* with increasing amounts of virtual content [MK94]. An illustration of this continuum is shown in Figure 2.1. This continuum mainly describes the visual sense or the system output channel towards the user. The term VR was popularized by Jason Lanier since the 1980s, whereas before the term *Virtual Environment* (VE) was also utilized in academia [Bry99; WC15]. We use the VR definition 1.1.1 according to Brooks [Bro94] in Chapter 1. Within this thesis, we refer to VE as any virtual scene. An *immersive virtual environment* (IVE) is a virtual scene which can be viewed at least semi-immersively, e.g. through a Fishtank VR setup. A *fully-immersive VE* is a virtual environment which is viewed fully-immersed in it, e.g. through an HMD.

Brooks also used the term *virtual reality experience* which means that the user is immersed in a responsive, virtual world, requiring real-time rendering, tracking of the user's body, head, and limbs, a display which blocks out environmental distractions and a system to create realistic virtual

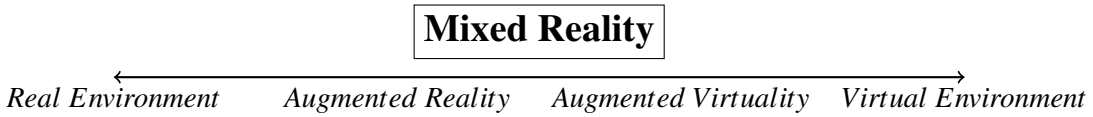


Figure 2.1: The mixed reality continuum according to Milgram and Kushino [MK94].

worlds [Bro99]. However, there are various definitions for VR with slightly different foci. Burdea defines VR as *"(...) a high-end user-computer interface that involves real-time simulation and interactions through multiple sensorial channels (...) visual, auditory, tactile, smell and taste."* and defines the *three Is of VR* as **I**nteractive, **I**mmersive and **I**magination [BC03]. According to Bryson's definition, VR is the abstract notion of using computer technology such as displays, tracking, and other devices to provide users with an interactive, three-dimensional experience invoking a sense of spatial presence [Bry99]. As written in his essay, the terminology was and still is often unclear, as some researchers use the term VE over VR. A VE is a synthetic, simulated environment which is displayed, whether visually or through other or more sensory channels, especially in VR applications. AR describes an application that allows the blending of virtual objects into the real environment. Most of the 3DUIs developed in this thesis have been targeted for VR.

2.3 Roots and History of 3DUIs

Throughout the history of HCI, there are repeating patterns of hardware-software-usability-comfort which will be described in this chapter. Currently, we are at a stage where hardware and software for 3DUIs are sufficiently advanced, so that research has to focus on long-term usability again. The repeating pattern of hardware-software-usability-comfort is shown by giving a historical overview.

Before 1963, UIs were at first line-printers and afterward command-line (CLI) and text-based (TUI) [Nie94]. Output on line-printers can be called 0DUIs, whereas CLIs and TUIs can be called 1DUIs [Nie94]. In 1963, Sutherland's Sketchpad provided the foundations for a new era of 2DUIs UIs [Sut63]. Sketchpad allowed users to create 2D content digitally on a screen by drawing lines and shapes, snapping them to others and enabling cloning and scaling. For the first time, it was possible to interact with the computer in more than just one dimension in a proper application, and while Sketchpad was a research demo, it pioneered the field of HCI [BR03]. A light pen and a physical push button were the primary input devices for Sketchpad. The light pen acquired position data when pointed towards the vector-based cathode ray tube (CRT) screen. In a way, the light pen provided users with a familiar interface inspired by a pen. As part of his research project towards the augmented knowledge [Eng88], Engelbart devised the computer mouse [EEB67]. This allowed users to interact with computers in two dimensions since the potentiometers on this device were orthogonal to one another. However, this was not the first two-dimensional input since before the mouse, trackballs and joysticks could also be used. Around the same time as Engelbart's mouse was presented, the German company Telefunken already sold a mouse-like device with a ball, their *Rollkugel*, developed by Rainer Mallebrein [Ste16]. Considering the success of the computer mouse and it's extensive usage in desktop workplace environments, the mouse can be seen as the central pillar of 2DUIs. In one of the first HCI studies, English et al. compared a joystick, a Grafacon, a mouse, a knee-based controller, and a light pen for text selection, meaning the positioning of a cursor on a screen to select a text or a symbol for further operations [EEB67]. Most of the evaluated devices used *absolute positioning*. This means that the position of the device relative to a specific start position corresponds to a position of the cursor on the screen. For example, the mouse being on the very left corresponds to the cursor being on the very left. In contrast to that, the joystick was implemented in an absolute and a *rate mode*. The rate mode means that the further the joystick

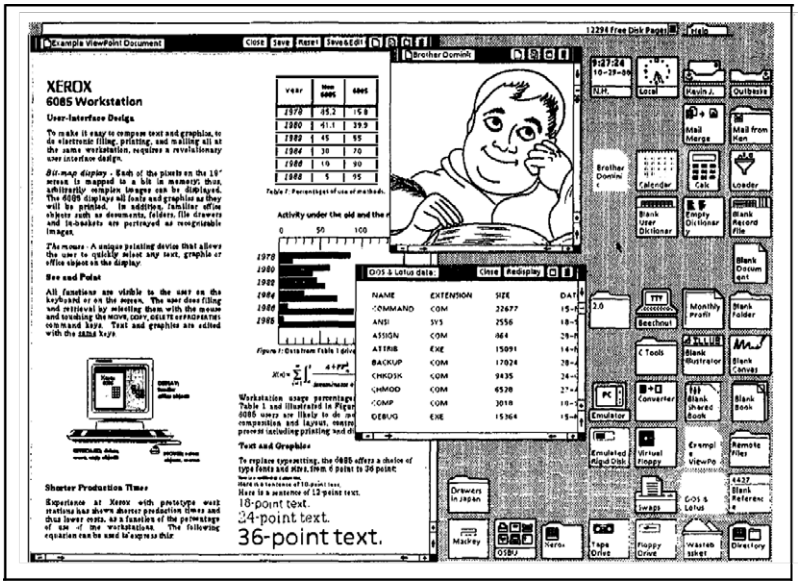


Figure 2.2: Xerox Star User Interface, adapted from [Joh+89].

was tilted, the faster the cursor would move in the corresponding direction. In the paper, a few procedures are used which are still common nowadays such as the evaluation of the *performance* of users during a *selection* task [EEB67].

Along with devices offering 2D input, raster-based graphics were developed, allowing users to visualize pictures and other content instead of just vector graphics as with previous displays [Nol71]. Although the foundations for raster-based displays were made by Farnsworth in 1930 [Far30], the memory required to show digitally generated content on computers was too expensive, and it took until the 1970s for raster displays to overtake direct-view storage tubes [Mac94]. In the 1970s, at Xerox, the next generation of user interfaces was developed, the so-called graphical user interfaces (GUI) or Windows, Icons, Menus, Pointer (WIMP) interfaces. The Xerox Star Information System [Can+82] was one of the first commercialized WIMP systems and the developers designed the system with specific, user-centric requirements in mind. For example, the primary goals were simplicity, consistency, user tailorability, universal commands, seeing and pointing versus remembering and typing as well as a novel interaction concept called *What You See Is What You Get* (WYSIWYG). These goals allow drawing a clear distinction between CLIs and GUIs due to the exploratory way of using and learning a system. Especially the WYSIWYG principle is vastly different from strict CLIs, where remembering the commands is necessary to be able to use a system. The distinguishing factor between fullscreen interfaces and WIMP interfaces is the fact that windows in WIMP Interfaces can overlap, an attribute suggesting three dimensions. Since interaction is only possible with visible content, WIMP interfaces could be classified as 2.5 dimensional [Nie94]. As far as the interaction is concerned, most GUIs are based on *direct manipulation* [Shn81]. This means that users can, for example, drag an object from one position to another using a mouse pointer instead of having to specify the new coordinates with a command. Nielsen argues that this direct manipulation is not necessarily the best solution for all the problems, as some tasks require precise input, e.g., the accurate scaling of an object in a 3D modeling application such as Blender can be done by specifying the exact scaling through keyboard input, while the direct manipulation using the scaling tool can be imprecise.

There have been various comparisons between TUIs or menu-based interfaces and GUIs to determine whether GUIs offer any benefits. Rauterberg compared GUIs and character-based versions of database software on IBM PCs with novice users and expert users. They found that in

both the expert and the novice group the GUI with the mouse as input performed their tasks faster than their Menu UI counterparts with function keys [Rau92]. Staggers et al. found similar results when comparing a prototype GUI and a TUI. Their study was concerned with health care, and they found that the participating nurses were significantly faster, less error-prone and more satisfied than with the TUI [SK00]. While these were direct comparisons offering few variables, other studies have shown that the results can be generalized [Nie94]. Shneiderman describes this type of interaction as direct manipulation, a term that will be used throughout this thesis and describes the possibility that software can directly react to a user's input and the user can directly manipulate things they see. For example, users can manipulate a CAD sketch on the computer screen by touching a pen to it and dragging visual representations of electronic components onto the sketch [Shn81]. Nonetheless, there are indicators that by making the interface more direct or natural, the learnability of software increases, albeit direct interaction and graphical representations do not provide benefits for all tasks. Although GUIs are easier to learn than CLIs and can offer performance benefits even for expert users, direct manipulation can distract users from more efficient direct keypresses such as described in a calculator example in [Nie94].

WIMP interfaces were and are still dominantly used in office environments from which many of the metaphors originate, such as the desktop, folders, files, etc. During this era, the first concerns regarding workspace health rose and spawned a multitude of research regarding how to design a workspace to minimize the long-term negative effects of, for example, sitting in an unhealthy posture for long-term use. The field of human factors provides many guidelines and suggestions for such desktop environments. For instance, work at the computer using a mouse and keyboard was associated with distal arm pain and neck-shoulder pain [DMR12]. In fact, there are guidelines for every small factor of an office environment. For example, there are guidelines considering the seat height, depth, pan angle, eye-to-screen distance, as well as the positioning of the mouse and keyboard or what kind of seat should be used and that a footrest may be necessary to improve posture etc. [DMR12; MN12]. This shows that, with the increasing usage of desktop computers, potential health concerns became drastically more important.

2.3.1 Multi-Touch Interfaces

Touch-based devices caused the next paradigm-shift of UIs. Walker provides an overview and technological background of popular touch technology [Wal12]. Buxton offers a brief history of touch interfaces in [Bux10] where he explains that touch interfaces gained wide popularity, first through touchpads on laptops and later through the increasing prevalence of smartphones. The first touch devices were developed in research institutes in the 1960s [Joh67]. Companies like Casio already developed touch-based character recognition on calculators in 1984 [Bux10]. Touchpads were used as a mouse replacement mainly in laptops and thus rarely provided novel interaction forms since they were still used to move a pointer and for the selection of interactive elements. Touchscreens were often used with traditional user interfaces as a pointer replacement, meaning that instead of moving a cursor, the user could directly touch the interactive elements and thereby induce a system reaction. Although Buxton et al. were often recognized as having developed the first multitouch device, Bob Boie from Bell labs had already built a capacitive multitouch device beforehand [Bux10].

Multitouch technology, which allows the simultaneous tracking of multiple touch points, has evolved as standard in most consumer devices [Wal12] and is also used for several of the research projects presented in this thesis.

A lot of research has been done regarding the user interaction with these new input devices. Wu and Balakrishnan developed multi-finger and whole hand gestures, e.g., the two finger rotations [WB03]. A two-handed technique for simultaneous rotation, scaling and translation (R+S+T) of objects has been presented by [Kur+97]. The R+S+T technique and various other rotation and

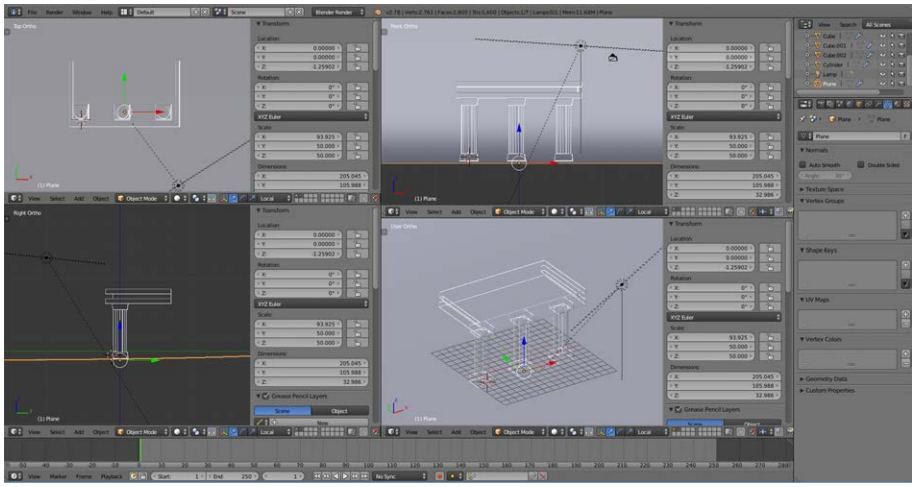


Figure 2.3: The Blender 3D modeling suite with multiple viewports showing content from various sides.

translation techniques were compared in [Han+06], some of which have been widely employed in commercial applications, such as Google Maps. In their work, they also argue for the use of touch-enabled tabletop displays to allow collaborative work. These tabletop displays allow users to sit around the display, facing one another. This helps users to communicate more efficiently in contrast to single user devices such as the above mentioned desktop computer or smartphones. To properly allow collaborative work, users need to be able to see the content simultaneously and also interact simultaneously. Multitouch technology enabled new forms of collaborative interaction since many common single-touch interactions were based on classical WIMP interfaces. These traditional interaction principles did not work in a context with more than one pointer, mainly if these pointers were controlled by multiple people. Nonetheless, these interfaces were mainly based on two dimensions, and besides simple stacking of windows or other elements, they had no depth information, especially no stereoscopic depth cues. A feature that was used, for example, in [WB03] was a correct perspective projection of three-dimensional objects. In their RoomPlanner, the authors use this perspective projection to show a preview of an object.

2.3.2 3D User Interfaces

A comprehensive overview about 3DUIs can be found in [Bow+04] and the second edition [LaV+17]. For 3DUIs to exist, computer graphics had to be developed first. For a more detailed history of computer graphics, the reader should refer to Machover's retrospective [Mac94]. With the pioneering steps taken at the Whirlwind project [Hur+89], many crucial 3D computer graphic algorithms were developed, for example, in connection with the University of Utah [Bli78; Cat74].

Machover describes CAD software as the "killer application" that was necessary to help the field of computer graphics to flourish, and once the software and hardware became more readily available, especially with home computers, the interfaces became increasingly important [Mac94].

In this section, we first discuss monoscopic visualization which is incapable of showing 3D content with actual depth. Thus, multiple views were often used to provide users with different perspectives, usually along the principal world coordinate axes. As Figure 2.3 shows, this is still a popular solution in contemporary design applications. While computer graphics was still a new research field, most of the visible content had to be generated by writing program code. This means that most objects and positions were hard-coded or imported. Numeric typing with a keyboard for object transformations is still one of the most precise and straightforward ways of positioning and rotating items, as Figure 2.4 shows. These interactions were not classical 3DUIs according to

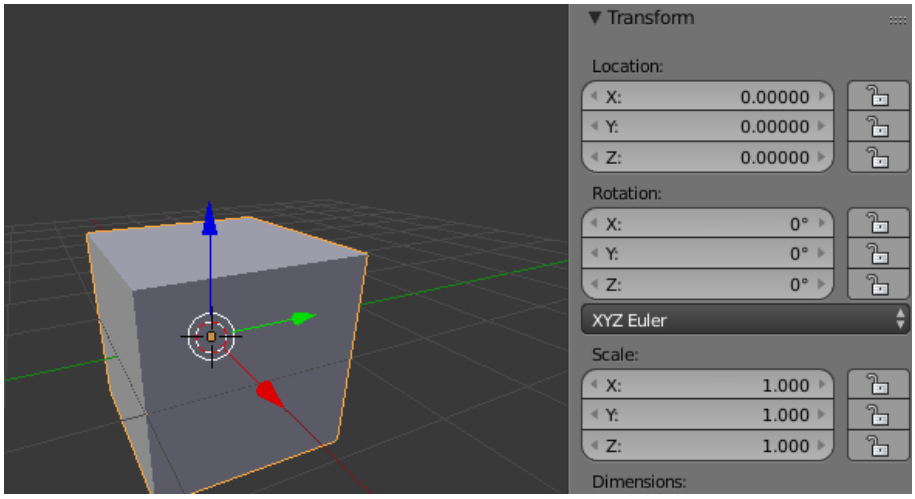


Figure 2.4: 3D translation widget and direct numerical input in Blender.

Bowman et al. [LaV+17].

At first, 3D interaction techniques were often based on 2D input devices, such as the mouse, making it necessary to map 2D control of a mouse onto 3D content. This circumstance often results in the decomposition of complex 6 DOF manipulations into manipulations of lower complexity [LaV+17], for instance, using mouse cursor controlled sliders to move or rotate objects along one specific axis in local or world coordinates at a time [CMS88]. Chen et al. found that for 1 DOF constrained manipulations sliders offer appropriate accuracy. However, for unconstrained, fully 3D manipulations, users had difficulties dividing the necessary rotations into steps that could be executed using sliders. One solution for mathematically correct rotations of objects using a mouse is the ARCBALL [Sho92], which is a Virtual Sphere technique according to [LaV+17]. Virtual Sphere techniques involve a spherical widget allowing users to rotate an arbitrary object by clicking and dragging the mouse cursor over the object and instead of mapping the coordinates directly to the object a proxy sphere is used. The coordinates on the proxy sphere can then be used to calculate the rotation axis, by calculating the axis on the view plane which is perpendicular to the dragging direction. Afterward, various mappings can be used to calculate the turn angle utilizing the distance between the points when the mouse button was pressed and released. It is possible to map these Virtual Sphere interactions to touch interaction, as well. The ARCBALL technique is mathematically more correct, as it does not utilize an arbitrary function to calculate the rotation angle and does not take the perpendicular axis as a rotation axis [LaV+17]. Instead, it calculates a rotation axis using the center of the sphere and the exact positions on the unit sphere to calculate a quaternion describing the rotation between these points.

All of these techniques were based on desktop-based 3DUIs using 2D input devices. Especially for 3DUIs, there are two main transfer functions used to map movement of a physical input device to movement in a VE [Frö+06]. For *rate control*, a physical movement of a device is translated to a change in the rate of movement. A typical example would be flight simulator joysticks, where the offset from the neutral position is often connected to a more rapid turning around the pitch or roll axis. *Position control* is most commonly known from mouse input, where moving the mouse physically also moves it a certain distance in the VE. Herein, the *control-display-ratio* is essential. It defines the ratio between a length moved in the physical space and in the VE. New devices were developed allowing more than 2 DOF interaction, such as the SpaceMouse [3DC14], which is a knob on top of a plate with various buttons. This knob can be pushed in all directions, as well as twisted along all three axes, resulting in 6 DOF. Due to the vast popularity of the mouse, these 3D input devices are primarily used for research purposes and industrial applications. For example, Fröhlich

et al. propose the Cubic Mouse, a hand-held cubic device with three perpendicular rods going through the center of the device, various input buttons and an embedded 6 DOF tracker [Frö+00]. The device can be used to define the origin of the coordinate system, similar to the World-In-Miniature (WIM) metaphor [SCP95]. While the device is held in the user's non-dominant hand, the dominant hand can be used to interact with the rods to manipulate objects or the buttons for system control. This device was used in an industrial context for car manufacturers [Frö+00] to manipulate cut planes through car models, and for the visualization of geological data [Frö+99] for the oil and gas boring industry. Fröhlich et al. also presented the GlobeFish and GlobeMouse, two other 6 DOF input devices [Frö+06]. The GlobeFish is an elastically attached custom 3 DOF trackball that allows rotations to be applied through the trackball and translations by applying lateral force. The GlobeMouse is a GlobeFish attached to a SpaceMouse. Rotations can be done using the trackball and translations by moving the SpaceMouse used as a base. The authors argue that the separation of force-free (trackball) and force-requiring (SpaceMouse or the elastic attachment) movement allows users to quickly differentiate and switch between the rotation and translation. These devices were compared against the SpaceMouse as a baseline and found that both their prototypes are superior to the SpaceMouse. Besides that, the work also provides an overview of other 3D input devices.

There are mixed results regarding the efficiency of 3D input devices, with some researchers arguing that using a traditional 2D input device, such as a mouse, and decomposing the complex 3D task into various 3D tasks can be less stressful and faster than 3D input devices [Bér+09; Sch+12]. There are multiple technologies allowing users to view content in a stereoscopic way, giving them the advantage of being able to use the depth cues from stereopsis in addition to monoscopic depth cues. A couple of milestones were the invention of stereoscopes in 1838 by Wheatstone [Wad87], the first shutter-based cinema [Ham28; Zon14] and Mort Heilig's multisensory Sensorama Simulator [Hei62]. Far ahead of its time, Sensorama not only provided users with wide-angle stereoscopic imagery but also scent and wind simulation. As it was entirely mechanical, using it for digital content was not feasible.

As mentioned in Section 2.3, displays for computers were not as advanced until 1976 when stereoscopic viewing with shutter glasses for computer graphics was developed by Roese and Khalafalla [RK76]. Ivan Sutherland is widely regarded as one of the inventors of interactive stereoscopic 3D graphics, with his Sword of Damocles head-mounted display (HMD), an HMD capable of showing computer-generated 3D imagery in real time [Sut68]. Since the HMD was mounted on the ceiling, the arm that the HMD was attached to could be used to track the user's position. The system provided users not only with stereoscopic visuals but also with motion parallax. Alternatively, an ultrasonic sensor could be used for this purpose. Sutherland even described this functionality as the fundamental idea behind the device. Even with its limitations, such as only being able to show wireframes of objects, the HMD allowed users a believable 3D impression.

The added dimension due to stereopsis raised new questions about the usability of software. The content was not strictly limited to a 2D plane anymore but could be displayed in front of or behind the screen. Hence, due to the different perspectives for each eye, just a 2D cursor could be occluding a 3D object on one eye and not on the other eye. Furthermore, terminology and classification had to be revised. In 1999, following the VR hype of the 90s, Brooks wrote a VR state of the art article [Bro99] in which he categorizes various technologies. The categorization is mainly based on the kind of display technology used and whether the system is static or not.

Significant advantages have been made to improve upon many of the technological challenges described in [Bro99]. Modern HMDs have display frequencies of 90 Hz [HTC17] or 120 Hz [Son17]. These frequencies result in latencies of about 11.1 ms or 8.3 ms respectively. Tracking of the user's head and potentially other body parts, for example, in the consumer-grade HTC Vive, provides low latency at sufficient precision [NLL17]. Research-grade tracking can give even superior results [NLL17]. Considering the systems part of Brook's challenges, a considerable

amount of research has been done in all of those fields, which we will discuss later. However, LaViola et al. provide an overview of the research [LaV+17].

As far as the technology is concerned, a similar quality level has been achieved compared to the desktop computers. While there are still technological improvements to be made, especially concerning non-visual output channels, the main challenges now lie in the development of usable software and guidelines, as well as the matching ergonomic guidelines to make VR usable for long-term use.

The following sections are concerned with the foundations of 3DUIs and ergonomics. These foundations are necessary for the remainder of this thesis.

2.4 Human Factors

This section summarizes the relevant background from human factors. Detailed information can be found in [LaV+17; Sal12]. To fully understand how 3DUIs work, a basic understanding of the user is necessary. This means that it is important to know how users perceive their environment and the interface, how they interpret input and how they decide on actions. Furthermore, understanding classic biomechanics is important since user interfaces are being built for more elaborate software and meant to be used for more extended periods of time.

Within the field of 3DUIs, *behavior* has to be understood in the design of NUIs, to leverage the familiarity with certain actions to ease the accessibility [WW11]. *Skills* are often a requirement for the use of UIs and 3DUIs, and *skill transfer* from VEs to real environments is often a desired result, for instance in training applications [LaV+17].

2.4.1 Information processing

Various models are describing human information processing. A widely accepted one from Wickens and Carswell [Wic+15], is shown in Figure 2.5. Furthermore, Card et al. presented their *Model Human Processor* (MHP), which is a higher-level way of modeling human information processing [CMN86]. We integrated the core ideas of the MHP model to Figure 2.5. According to these models, humans process information in a feedback loop. In the perception stage, humans first gather information with their senses and then filter them in a perception step that is influenced by memories and the available resources. In the cognition stage, selected information from the perception stage is usually processed in the memory. The memory part of the cognition is profoundly affected by the attention resources and the decision phase. A person has to actively decide which content is transferred into working memory and transfers content into the long-term memory, for example, by repetition. As such, working and long-term memory can affect one another. Response actions are in the physical ergonomics stage of the action and are chosen in the cognition stage but also affected by the attention resources.

The attention resources can be distributed in various ways according to LaViola et al. [LaV+17], which is based on [Wic+15].

Selective attention Actively choosing which information from the environment should be processed. It is mainly based on (i) salience, how the environment attracts attention, (ii) expectancy, using the knowledge of when and where to expect new input, (iii) value, the importance of information depending on where it came from, and (iiii) effort, how difficult is it to obtain the information. All of these forces can be represented by the SEEV visual attention model [Wic+09].

Focused attention The desire to maintain processing of a specific source and avoid distractors, for instance, within a 1° visual angle from a visually focused object or within a particular frequency range and intensity of a sound.

Divided attention The simultaneous processing of events at a given time, according to [LaV+17].

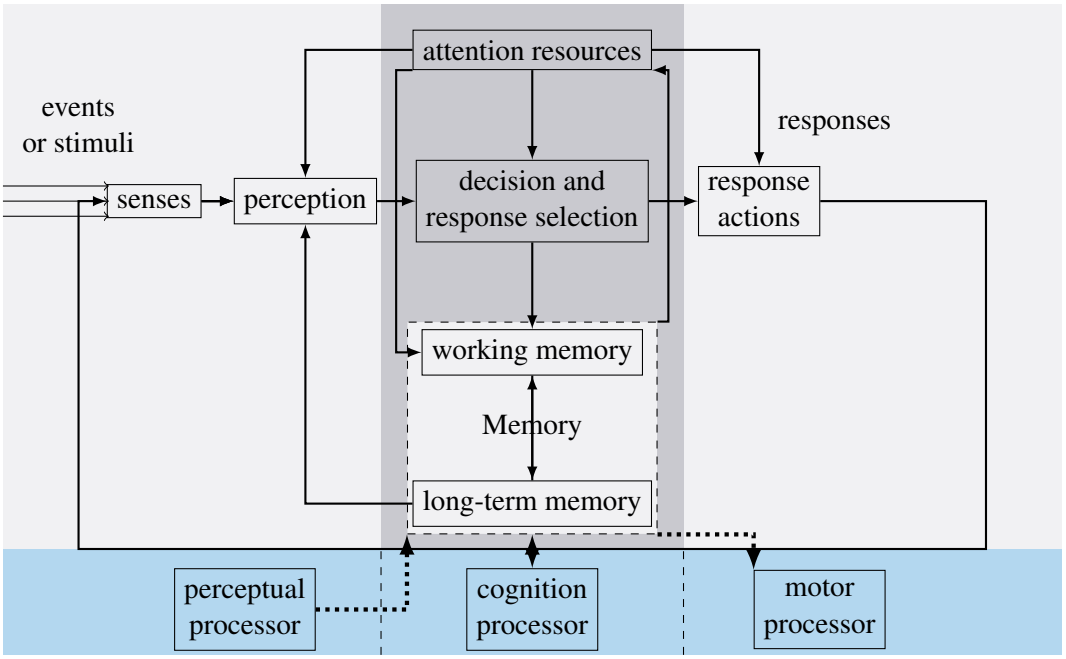


Figure 2.5: Combination of the information processing loop according to Wickens [Wic+15] on top and Card's Model Human Processor [CMN86] in blue at the bottom with dotted connections.

However, the human attention is fallible, as shown in *change blindness* effects that illustrate the human inability to detect changes to an object or a scene [SL97]. Change blindness shows that humans do not transfer a lot of information about their environment from one loop cycle to another, whereas the vast majority is filtered out. This effect is further enhanced under high workload or due to low saliency and can appear despite being safety critical [Wic+09]. In VR, change blindness can be utilized positively, as well, as it allows the dynamic reconstruction of the VE to constrict the user to a safe physical environment without the users noticing as shown by [Sum+11]. In their work users had to enter a room through a hallway and walk to a distractor, for example, a desk, and then turn around again. During the timespan between passing through the door and walking up to the desk, the position of the door and the size of the currently invisible part of the room was changed. This led the participant in another direction upon exiting the virtual room and kept them inside the real room.

The central part of the information processing loop is the decision and response selection [LNY12], the process of choosing between multiple alternatives, given that doing nothing may be a valid choice, with consequences. For example, the decision whether to move single vertices or whole polygons to shape a 3D object during a modeling task. When a user can teleport to a target or walk to their target, they tend to decide for the more comfortable one [LaV+17]. Decision making is highly situational and difficult to model. Especially when multiple parties are involved in a decision-making process, conflict may arise due to the uncertainty of outcome and differing objectives. These conflicts are the reason why conflict resolution is a key aspect of decision-making. Lehto et al. provide a more in-depth look about the three types of decision-making models, such as (i) the rationally and statistically driven normative model, (ii) the behavioral model which takes into account that humans have limitations and are not necessarily rational, e.g., due to fear, and (iii) the naturalistic model that takes into account realistic situations [LNY12].

According to [LaV+17], behavior describes the process of decision-making with regards to previously acquired knowledge. Behavior can also be changed due to habituation and conditioning [FK86]. One of the more well-known examples of habituation in the VR field is Albert Rizzo's

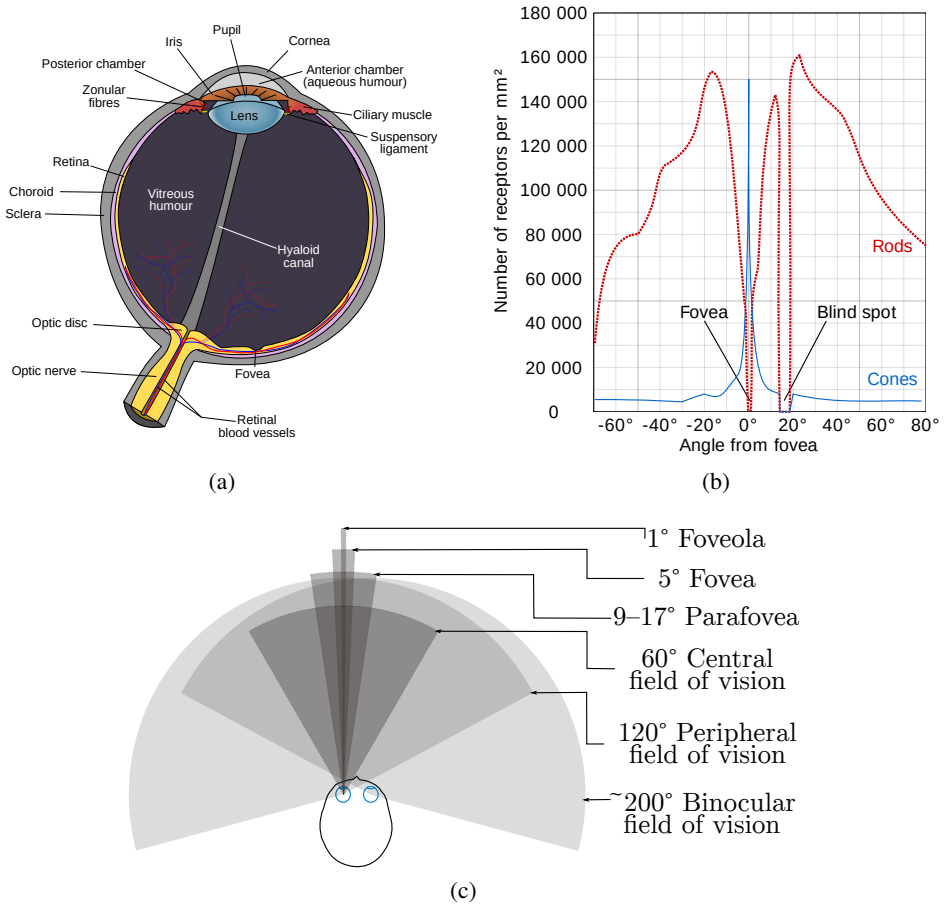


Figure 2.6: (a) Schematic of the human eye. (b) distribution of cones and rods on the human eye [Cmg13] (c) Human field of view schematic.

work on exposure therapy for treating post-traumatic stress disorder (PTSD) using configurable VR reconstructions of field scenarios that led to the soldier's trauma [Riz+14]. Skills are similar to behavior since they are both based on previous experiences. However, skills can be acquired and reinforced over time through repetition and due to this, they have close links to the decision-making process [FP67; LaV+17].

2.4.2 Perception

An essential step in the information processing cycle is the perception stage, and a good understanding of how the human senses work is necessary for our peripheral stimulation extension for HMDs in Chapter 5 and the reasons behind performance differences discussed in Chapter 3.

2.4.2.1 Vision

On a biological level, vision is based on photoreceptor cells on the retina, which send electric pulses through nerves when stimulated by visible light from the electromagnetic spectrum. As shown in Figure 2.6a (a), there are two kinds of photoreceptors on the human retina:

- Rods are responsible for low light vision and are mainly distributed around the periphery of the eye.
- Cones are responsible for higher light level vision and trigger at different wavelengths, allowing color vision.

The distribution is shown in 2.6b which indicates that biologically, the human perception system

only rarely receives information about different wavelengths of incoming light in the periphery. More specifically, the combined, horizontal human FOV approximates on average 200°, with the vertical FOV approximating 120°, with the foveola having a diameter of 1°, the fovea 5.2°, the parafovea about 9° to 17°, the central field of vision being around 60° and peripheral in the remainder [Pal99; SRJ11; Wan95]. The foveola is the area with a high density of cones, providing high resolution. With increasing distance, the density of cones decreases (cf. Figure 2.6).

There are controversies on the importance of FOV for VR. Some studies have found an influence of a wide FOV on various tasks in VR, such as [RB04], who found that users are capable of pointing at objects more easily when they have unrestricted FOV as compared to a smaller display (40°). Jones et al. compared different FOVs in an HMD and found a significant influence of a wide FOV on distance estimations [Jon+12]. However, other studies did not find an impact of a FOV restriction on a search task [LR05].

According to [Dee98], the limit of the resolution of human vision is about 28 arcseconds. This limitation means that some humans can discriminate two stimuli with a distance of 28 arcseconds under ideal conditions when looking at them with their fovea. Thus, the ideal HMD would have a horizontal FOV of 200° and a vertical FOV of about 120 °. Taking into account the maximum acuity of a human at 28 arcseconds, the display resolution should be about 25714 × 15429 pixels. Even when taking the average of one arcminute as human acuity, the resolution (12000 × 7200) is still higher than the resolution of currently available HMDs like the HTC Vive (2160 × 1200 at about 100° FOV).

The distance between both eyes, or more specifically between the pupils in the center of the eyes, is called *interpupillary distance* (IPD) [Mil14]. The IPD is one of the vital calibration factors for HMD calibration. Various ways of measuring the IPD exist, such as the method proposed by [Wil+08]. This method consists of the user facing a mirror while their heads are being held still and then marking their pupils with a marker on the mirror. The eye which is not being marked at the time should be closed. The distance between the marks can then be measured with a ruler. Adjustment for IPD provides a better experience with HMDs since the focal point of the lenses should match with the user's pupils to avoid a distorted image or reduced FOVs [Ras+09]. A proper IPD calibration is also necessary for distance estimations [Kel+12].

In this context, eye dominance denotes the situation in which one eye is preferred for monocular tasks, e.g., when looking through a looking glass [MOB03]. According to Mapp et al., the reasons behind eye dominance are still uncertain, but it manifests in most people [MOB03]. There are several common tests for eye dominance. Foremost, the Porta method, wherein a user has to cover a distant object with the thumb and then close one eye and next opening it again and repeating that with the other eye. The eye which still sees the target is defined as the non-dominant one. The second method is the Dolman method, which requires a card with a hole. First, the user has to extend their arm with the card and look at a distant object through the hole and then pull the card towards themselves. The hole of the card should be in front of the one dominant eye.

To select objects in 3DUIs, proper depth estimations or depth perception are necessary to ensure the user does not touch through or behind objects. In 3D graphics engines, the developer has to decide which depth cues to offer. According to Howard [How12], depth perception is divided into visual depth cues and non-visual depth cues. As our focus is on computer graphics, the focus will be on the visual depth cues. There are various depth cues that Howard [How12] and Palmer [Pal99] categorize into a few categories: oculomotor, monocular and binocular cues (cf. Figure 2.7). Additionally, some depth cues offer absolute depth information while others only provide information about the depth of objects relative to one another [Dör+14] (cf. Figure 2.7)

Monocular cues are depth cues that can be perceived with a single eye and thus a single display without additional equipment (cf. Figures 2.8,2.9).

Binocular cues require both eyes to see a different image and thus specific display hardware.

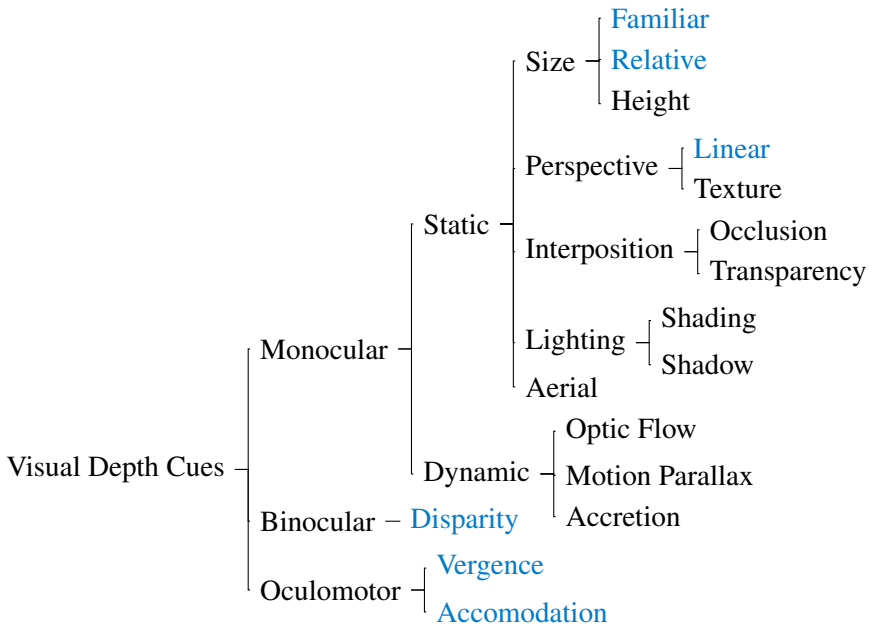


Figure 2.7: Structure of depth cues according to Howard [How12]. Absolute depth cues are marked blue [Dör+14].

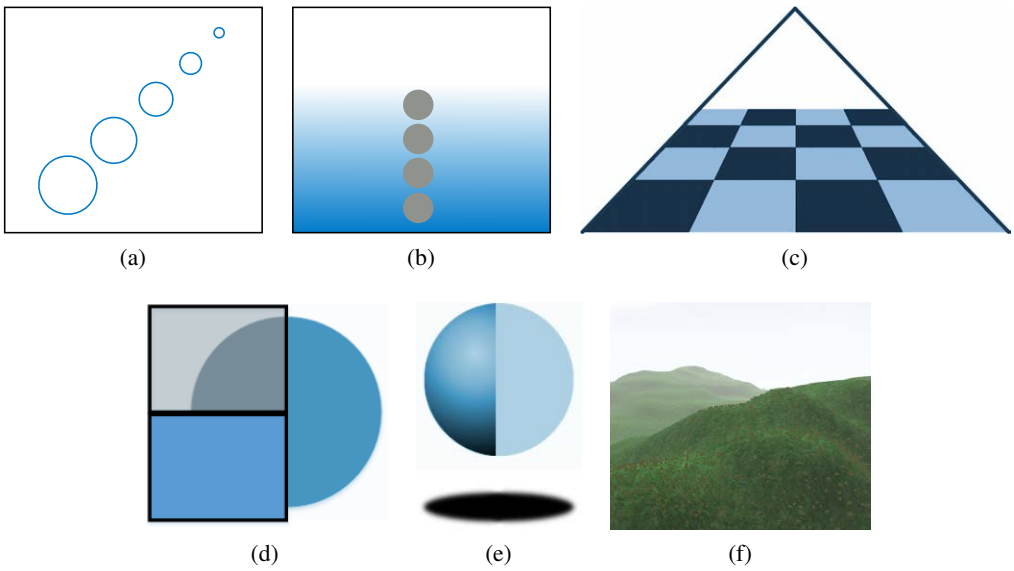


Figure 2.8: Static depth cues: (a) relative size (adapted from [LaV+17]) (b) distance to the horizon (c) perspective and texture gradient (d) transparency and occlusion (e) shading and shadows (f) aerial haze.

They are further divided into vergence and disparity cues. The disparity is caused by the different positions of both eyes. Due to the particular difference and the fact that both eyes look in the same direction, humans can utilize disparity cues for depth estimations [Col96].

All of those depth cues help the estimation of depth to different degrees and the processes in the human brain are still being researched [BR96]. Moreover, there are many possibilities for depth cues to be interpreted falsely, as many optical illusions show.

Oculomotor cues are depth cues that are not sensed through photoreceptors but from the muscle

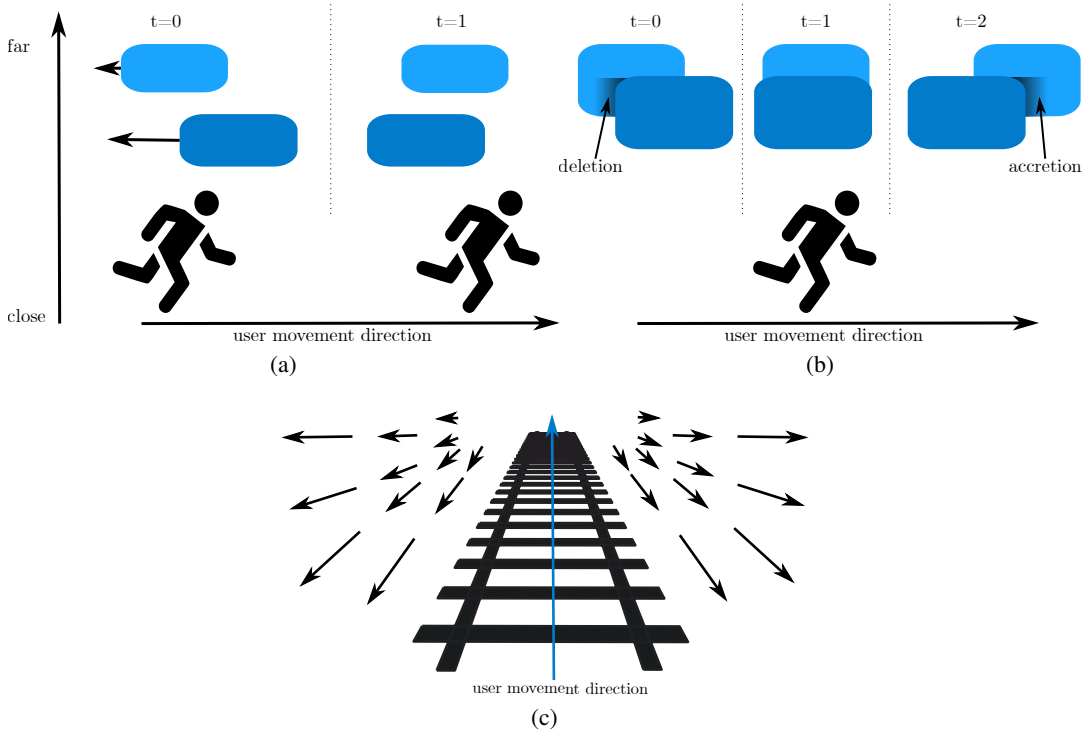


Figure 2.9: Dynamic depth cues: (a) accretion and deletion (b) motion parallax (c) optic flow.

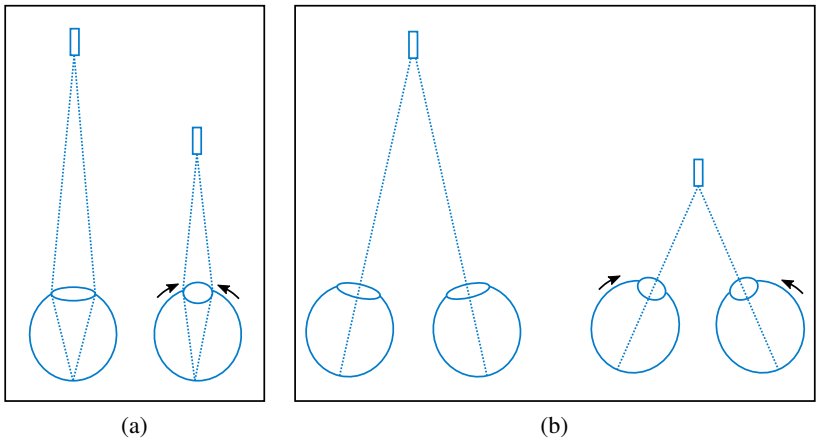


Figure 2.10: Binocular depth cues; (a) accommodation, the arrows show tightening of the ciliary muscles around the lens to focus on the close object (b) vergence, the arrows show the convergence of the eyes to focus on the close object.

action necessary in the visual system to get a crisp image [LaV+17]. Accommodation, as illustrated in Figure 2.10(a), should be classified as a monocular depth cue, as even with a single image or a single eye depth information can be derived by how much the ciliary muscles around the lens have to contract to increase the lens refraction due to diverging light rays from a close object. Vergence is the process of eyes turning inward to focus close objects (convergence) or turning outward to focus objects far away [CJ84], and while the *vergence change* offers depth cues, the *static vergence* of the eyes also allows depth estimations [How12] (cf. Figure 2.10(b)).

Figure 2.11 shows how useful the various depth cues are for distance estimations depending

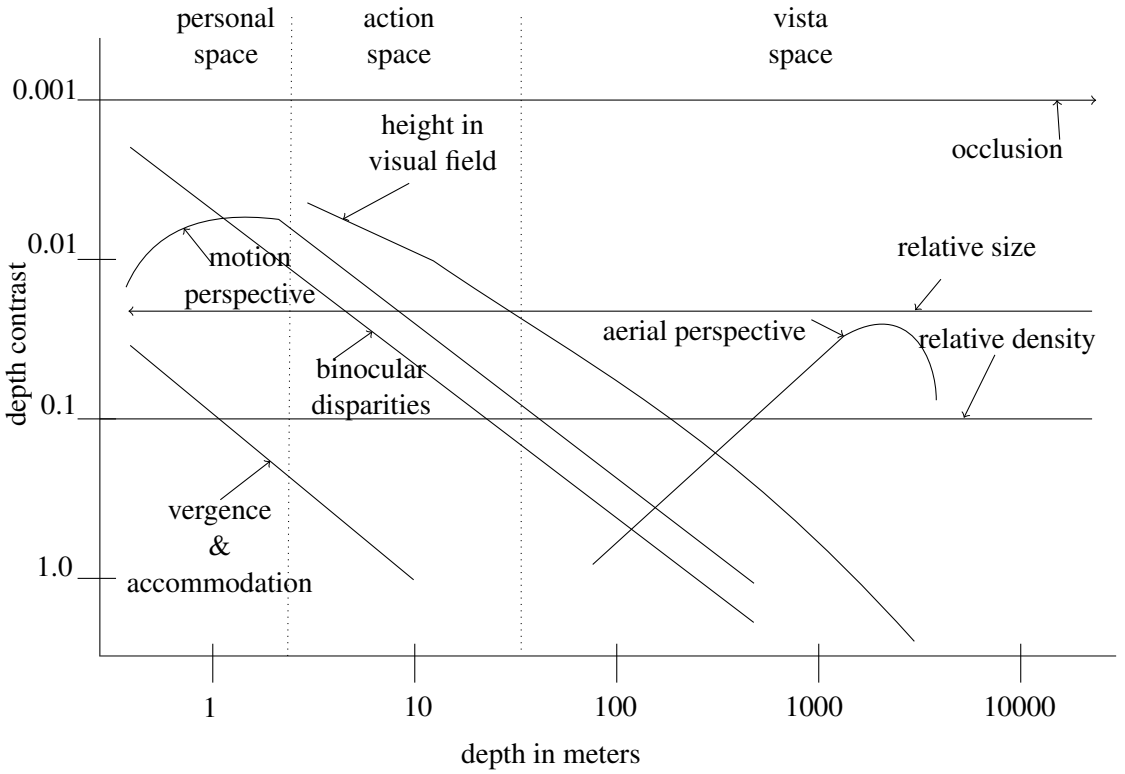


Figure 2.11: Depth cue relevance at different distances. Adapted from [CV95].

on the distance of the observer [CV95]. Occlusion is a dominant depth cue, but it only gives relative depth information. In contrast, vergence, accommodation, stereopsis and motion parallax provide information about the absolute depth [Dör+14; LaV+17] (cf. Figure2.7). The fact that stereopsis and vergence only work on binocular systems, i.e., systems providing a different view for each eye, explains why such systems, such as HMDs, offer a better depth impression than standard 2D displays [Jon+08]. Furthermore, a conflict can be observed when evaluating vergence and accommodation in most VEs. The eyes have to accommodate for a certain distance to the display plane, irrelevant whether it is projected as in a Cave Automatic Virtual Environment (CAVE) [Cru+92] or an HMD, yet the eyes converge normally to focus on an object. This conflict is commonly known as the vergence-accommodation conflict and is believed to be one cause of distance misperceptions in VR, as well as increased fatigue and discomfort [Hof+08]. Hoffman et al. developed a display allowing them to show stereoscopic stimuli with matching vergence and accommodation and conducted four experiments with their setup, comparing various factors, such as the necessary time to identify a stereoscopic stimulus, stereoacuity in a time-limited task, depth misperceptions, viewer fatigue, and discomfort. All these factors were improved when convergence and accommodation were almost or perfectly matching.

Distance estimations in VR are incorrect to some degree, mostly underestimating the distances in VR [Fre+07; LK03; Ste+10]. However, Interrante found no significant difference in distance estimations in a follow-up study where the VE was a replica of the physical environment [IRA06]. Although, in their first study they did find an underestimation of distances. Bruder et al. propose using simulated optic flow fields to compensate the underestimation [BSW11].

2.4.2.2 Haptic Perception

Haptic perception is relevant for our work with stereoscopic multitouch screens (cf. Section A.1.1), as well as the HapRing (cf. Section A.1.2). The commonly known “sense of touch” is divided

into two distinct senses: (i) The *cutaneous* sense, providing information about stimulation of the humans' outer surface through various receptors in the skin, and (ii) *Kinesthesis*, providing the human with information about the relative positioning of parts of their body to one another [LL86]. Thus, tactile perception refers to the kind of stimulation solely through cutaneous information, and kinesthetic perception incorporates almost exclusively kinesthetic information, for example, the task to blindly determine the different lengths of rods. Some works differentiate between *proprioceptive cues*, the cues providing information about the position and angle of a human's joints, and kinesthetic cues, providing information about a human's body and other objects through muscle tension [LaV+17]. Merely through the effort necessary to keep the muscles contracted, the kinesthetic sense can convey enough information to determine whether a weight is heavier [LL86]. Meanwhile, haptic perception means the simultaneous use of cutaneous and kinesthetic information, which matches common tactual perception. Kinesthetic cues work without eye contact and as such are also used for walking and motion. In fact, kinesthetic cues can also be used to interact with things that are out of the human FOV, allowing blind interaction or interaction with objects such as a car radio without actual eye contact. Thus, kinesthetic cues have been investigated in 3DUI research, for example, in Mark Mine's work [MBS97], where proprioceptive cues were being used for widget manipulation in a VE. In their work users were able to complete a task of returning a hand-held object to a previous orientation, which indicates that users were capable of utilizing proprioceptive cues. In the scope of this thesis, we use proprioceptive cues in Chapter 4 by placing UI elements within reach in easily learnable positions around a user's joints, thus allowing them to use UI elements without having to look at them [Lub+16b].

2.4.2.3 Other Modalities

The vestibular system, which can be described as the sense of balance, consists of a series of organs containing fluid and hairs registering the fluid's movement, located behind the ear. Due to their perpendicular arrangement, the organs provide information about angular and linear acceleration in all directions and thus contribute to the sense of balance [LaV+17]. Considering that the vestibular system only reacts to real movements, a conflict may arise when the user perceives visual motion cues in a VE without moving in the real environment. Similarly to the vergence-accommodation conflict mentioned earlier, this vestibular conflict can cause cybersickness [LaV00]. Adding subtle correlating vestibular motion cues, for example, by leaning in the movement direction, may reduce the cybersickness symptoms [Kru+16; Kru+15a].

Within the scope of this thesis, neither the auditory system nor the chemical sensing systems were researched. While the acoustic system is utilized in 3DUIs, the senses of taste and smell are rarely stimulated in VEs, despite the Sensorama already offering an output system for smells. LaViola et al. provide a summary of these sensory systems [LaV+17].

2.4.2.4 Evaluation and Challenges

Research from perceptive psychology suggests that the visual sense is dominating other senses, such as proprioception and the vestibular system, when there is a mismatch between them [Ber00; DB78; PNK76]. With solely optical flow as visual cue, users can reproduce directions of travel but not entire traveled paths [BIL00; LBV99]. The dominance of the visual sense is used in VR travel techniques, especially for redirected walking (RDW) [RKW01] (cf. Section 2.6.2). Inconsistencies between the proprioceptive sense and the visual sense are tolerable up to a certain degree [Azm+16; Bur+05; Jae+02; Jer+08; Ste+08b]. Burns et al. investigated how far they could move the virtual hand position from the user's physical hand position during movements and compared these thresholds to the detection of visual penetration of two objects [Bur+05]. Azmandian et al. also used the dominance of the visual sense to direct a user's grasp onto a real proxy object by adjusting the visual cues in a way to subtly change the grasp direction, similar to RDW [Azm+16]. Jaekl et al. measured when users thought a rotation or translation was considered stable. They found that

rotations had to be exaggerated by 26% and translations by 45% to appear stable to users, again indicating misperceptions in VEs. Jerald et al. measured whether users noticed a moving scene during certain phases of sinusoid head yaw rotations done at a specific frequency and found out that detection thresholds depended on the head-rotation phase [Jer+08]. Participants were less likely to detect movement with the head rotation. This could mean that more manipulations are possible during that phase [Jer+08]. Steinicke et al. measured detection thresholds for rotation and translation gains, as well as curvature adjustments for RDW [Ste+08b]. In their study, participants could not detect rotation gains of 0.88 or curvature gains of 0.12 but consistently had difficulties with the motion compression, where gains below 1.2 were misjudged.

Cybersickness

Cybersickness is still one of the main problems in current VR systems. One of the subjective ways of measuring cybersickness is Kennedy's simulator sickness questionnaire (SSQ) [Ken+93], which is one of the most commonly used questionnaires for this purpose [LaV00] (cf. Section A.2.1). The terms cybersickness and simulator sickness are often used synonymously in literature, although the symptoms differ [SKD97]. In the questionnaire, a multidimensional approach was followed, as various symptoms of cybersickness are being evaluated on 20 step scales. There are multiple theories as to what is the cause of cybersickness, for example, the sensory-conflict theory, mentioned in Section 2.4.2.3, the postular instability, and the poison theory [LaV00]. According to the sensory-conflict theory, cybersickness may be caused if the visual sense provides information about movement, but the vestibular sense does not. Also, if the visual feedback to physical (head-)movements is delayed, the vestibular sense provides movement information that the visual sense does not. This conflict can increase cybersickness [Cob+99]. More in-depth investigations of cybersickness and various ways of evaluating it are analyzed in [Bro+10; LaV00].

Presence

One of the more critical measures in 3DUI research is *presence* or the *sense of being there* [SW95]. Slater and Wilbur try to differentiate between *immersion* and presence by describing immersion as a technological display quality divided into [SW95]:

Inclusive How much the surroundings are blocked.
Extensive How many senses are stimulated.
Surrounding How much of the FOV is encompassed.
Vivid Overall quality of the display.

Presence, on the other hand, is described as a state of consciousness that allows users to feel as if the VE were a place they visited rather than just images they saw [SW95]. There are many factors that have been linked to causing immersion and presence, such as having a virtual body and matching proprioception. Immersion itself is an objective measure that directly influences the sense of presence. There were many investigations about presence in VEs, as the van Baren's compendium shows [VI04]. The dominant presence questionnaire used in this thesis is the Slater-Usoh-Steed presence questionnaire (SUS-PQ) [Uso+99b]. It gives a measure of how much presence users feel when immersed in the VE by asking six questions on a 7-point Likert scale each. The authors describe two ways to evaluate the questionnaire, either by taking the means or by counting the times the user's rated the questions with a 6 or 7. Within this thesis, we use the mean-based approach of evaluating the SUS-PQ (cf. Section A.2.4). Another presence questionnaire used in this thesis was created by Witmer and Singer (WS-PQ), who additionally provide an immersive tendencies questionnaire (ITQ) consisting of 29 questions about the participant's daily activities [WS98]. Their presence questionnaire consists of 32 questions measuring control, sensory, distraction and realism factors. Both questionnaires use a 7 point semantic differential scale with a middle anchor. For reference, the questionnaires are in the appendix (cf. Section A.2.5 and Section A.2.6).

2.4.3 Cognition

Cognition is involved and critical in all aspects of 3DUIs. In this thesis, we analyze cognitive demands of a travel technique in Chapter 5. Cognition depends on *long-term* and *working memory*, both of which require knowledge representation. A conclusive summary of how cognition works can be found in [LaV+17; SK13]. According to Smith and Kosslyn, mental representation is a physical state that conveys information about an object, event or concept and has to be deliberately constructed. Such representations can encode the information they carry in various formats:

Mental images Operating similarly to real images, however often with additional information about sizes.

Feature records Extraction of information that is of importance.

Amodal symbols Symbols that are disconnected from the perceptual and motor systems but can manipulate and relate them.

Statistical patterns Pattern in neuronal nets that mimic brain activity.

These representations are categorized, for instance, all different representations about an HMD can be categorized together, including instructions on how to use it while starting a research demonstrator.

These representations have to be stored in the long-term memory. The long-term memory is divided into declarative and nondeclarative memory. Declarative memory is explicit and can be recollected at will and described to others. It consists of episodic memory and semantic memory. Episodic memory keeps track of events in a context, such as remembering the first time you tried VR. Semantic memory is the knowledge about facts, such as the minimum frame rate of the HMD you were using but not the time and place when you acquired that knowledge. However, not all the information retrieved from long-term memory is perfect since the retrieval often requires stimulation with similar patterns. In the worst case, memory is inaccessible without stimulation. Even with stimulation, the retrieval of information is prone to errors due to conditions such as belief or consistency bias, misattribution or suggestions.

Working memory is temporary storage which is primarily used during activities to keep track of the most current stimuli, and its storage capabilities are highly limited. The popularly known magic number 7 ± 2 by Miller is one attempt to approximate the information amount which can be stored by introducing the term *chunk* as a higher level unit of structure [Mil56]. However, there have been other results in recent research [LaV+17; SK13] which limit the amount of storable information in working memory to 3 ± 1 if participants are not allowed to mentally rehearse the information. Information decay in working memory was investigated, for instance, by Brown and Peterson by asking participants to memorize a string of letters, count backward for varying timespans and recall the string afterward. Therein, longer delays resulted in more lost information [Bro58].

One of the main aspects where cognition is involved is situational awareness, which is an internal mental model of the current state of the user's environment [End12]. This model should contain all the necessary information from the environment, including abstract, derived data readings from instruments and input received from other users involved in the current task, if applicable. All further decisions and planning are based on this model, and as such, it is crucial for tasks such as wayfinding [LaV+17]. In both, real and VR wayfinding, the creation of a spatial *cognitive map* is necessary. A cognitive map is a hierarchical structure based on the spatial information stored in long-term memory and the spatial orientation. Spatial orientation is defined as the current location and rotation within the world and based on the working memory and skills. Both together are parts of situation awareness [LaV+17].

Spatial knowledge can be at least in one of the following categories that are developed over time [SW75]:

Landmarks Remembered in a context due to their importance.

Routes Linked Landmarks.

Minimaps Gradually built by memorizing distances and directions between landmarks.
Survey Knowledge Sum of minimaps and routes into a map-like structure.

Spatial information is gathered from an *egocentric* point of view, meaning that the observer is within the scene and the information is recorded with their body as a reference frame. However, the body is a complex system with many subsystems. This means that the egocentric reference frame stores information not just relative to the observer's eyes but relative to their head, both their eyes or their retinas, the body, and the complete proprioceptive subsystem. A detailed description of various reference frames can be found in [How91; LaV+17]. Meanwhile, advanced spatial knowledge is stored with an *exocentric* reference frame, meaning that the information is stored relative to an object, a map or another point in the world. The investigation of how the transfer process works is one of the areas in psychology research where VR is used extensively, e.g., in [Mei+15].

As such, it is vital that cognitive issues can be identified and evaded during the design of 3DUIs. Within the scope of this thesis, the *NASA TLX* (task load index) was used to gather subjective measures of the mental load in experiments [Har06] (cf. Section A.2.2). The questionnaire is widely used within the field of 3DUI and consists of two parts. The first part is a set of six, 100 point (reduced to 20) scales with descriptions. These scales evaluate the mental, physical, temporal demand, the user's performance, the effort they put in and the frustration they felt. The second part is a set of 15 two-alternative-forced-choice (2AFC) questions where the user has to pick which of two factors mentioned above was more critical. However, the second part of the questionnaire is often omitted as it has been shown that it may reduce the questionnaire's validity [Har06]. Another way of evaluating cognitive load of a task is to force users to perform a cognitively demanding task besides the primary task. One such task is the so-called *N-Back task* by Kirchner [Kir58]. Using a device with 12 lights that could light up in a randomized way and matching buttons, the users had to either press the button at the light (No-back), press the button where the previous light was (One-back), or press the button where the light was three cycles before (Three-back) [Kir58]. An n-back task would then be a task where a user has to remember something n-cycles back and perform a corresponding action, for example, they have to press a button if the shown slide is the same as the antepenultimate one [GC93]. Whereas this is a direct measurement of how well users are capable of remembering this information in the rapidly changing, potentially stressful environment, the task can be used as a secondary task, as well. LaViola et al. present other methods of measuring cognitive load in, such as electroencephalography and objective measures such as blind walking [IRA06], and drawing maps [LaV+17].

2.4.4 Motor Action & Ergonomics

Especially when considering the goal of creating 3DUIs that are adequate for long-term use, an understanding of human physiology is necessary. Taking into account the amount of time human physiology and biomechanics have been studied, there is a vast amount of literature in that field, some of which is summarized in health science textbooks such as [Hal15]. According to Hall, about 40% of the body consists of muscle connected to the skeletal structure of the human body, so-called skeletal muscle and another 10% of the body are smooth and cardiac muscle [Hal15]. Muscle consists of slow and fast muscle fibers. Slow fibers have access to a more extensive blood vessel system, allowing prolonged contractions, and fast fibers are larger and can create more strength during contractions. Smooth muscle is not connected to the bone structure and can be found, for example, in the gut. Skeletal muscles are connected to the bone structure by points of insertion, which is how the muscle lever mechanisms are created. There are various lever mechanisms in the human body, depending on the underlying structure. Since muscles can only contract, to allow back and forth movements, there are agonist and antagonist muscles on opposite sides of joints. Once the signals for a strong, active contraction of a muscle stop arriving from the spinal cord, signals have to arrive at the matching antagonist muscle to contract. Furthermore,

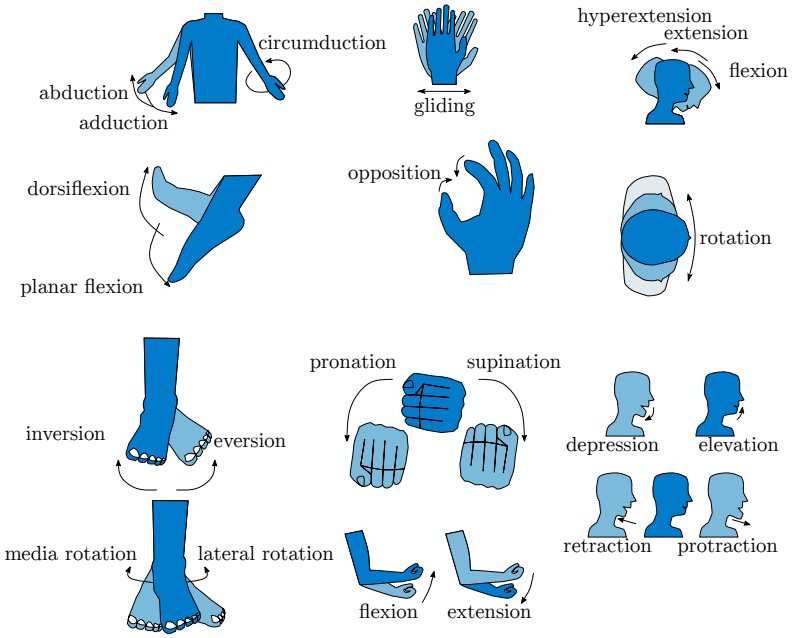


Figure 2.12: Human motion types. Adapted from [LaV+17].

a certain amount of tautness remains in muscles even while at rest, called muscle tautness. The higher their contraction, the less force is produced by a muscle. A fully relaxed muscle can produce most force [GK97]. Muscle contractions block the blood flow, meaning that prolonged muscle contractions cause fatigue rapidly and should be interrupted by short relaxation phases [GK97]. The energy used is in the form of adenosine triphosphate (ATP), which is broken down into adenosine diphosphate (ADP) to release energy, or phosphagen [GK97]. Phosphagen is broken down into phosphoric acid and creatine [GK97]. These broken down compounds are converted back to their high-energy counterparts, requiring energy supplied by the blood flow, namely glucose and oxygen that are only stored in minor amounts in the muscle themselves [GK97]. Muscle fatigue is mainly caused by a lack of glycogen and oxygen, which is a direct result of an inability to supply the work output through restricted blood vessels. If there is a shortage of oxygen, glucose is converted into pyruvic acid, which is turned into lactic acid that caused muscle fatigue. With oxygen, it is converted into water and carbon dioxide, releasing energy to create energy-rich phosphates [GK97]. For 3DUIs, this means that arms should not be raised continuously. Fatigue can also be caused by a reduced amount of signals coming from the spinal cord over prolonged timespans [Hal97]. Isometric contractions are defined as a muscle contraction without shortening of the muscle while isotonic contractions the length of the muscle changes. Either the isotonic contraction is concentric, which means that the muscle shortens and the lever is pulled, or the contraction is not strong enough to support the force acting upon the lever and the muscle lengthens. More detailed insights on fatigue and biomechanics are presented in [C+99; GK97; Mar12].

The human range of motion was investigated by various fields [Kir52] (cf. Figure 2.13). It is possible to utilize this full range of motions and motion types (cf. Figure 2.12) for 3DUIs. We will use these kinespheres which form when tracing the end of the kinematic chain over the whole motion range in Chapter 4. Zhai categorizes 3D input devices by being isometric or isotonic. Isometric devices sense force but do not perceivably move, such as the Spaceball [3DC14]. Isotonic devices have zero or constant resistance, such as a mouse [Zha96]. According to LaViola et al., most tasks in 3DUIs couple the user to the system in a control-body linkage that can be defined using required accuracy, speed, degrees of freedom, direction and duration. These characteristics directly restrict which body parts may be used for its control since various body parts have varying degrees of

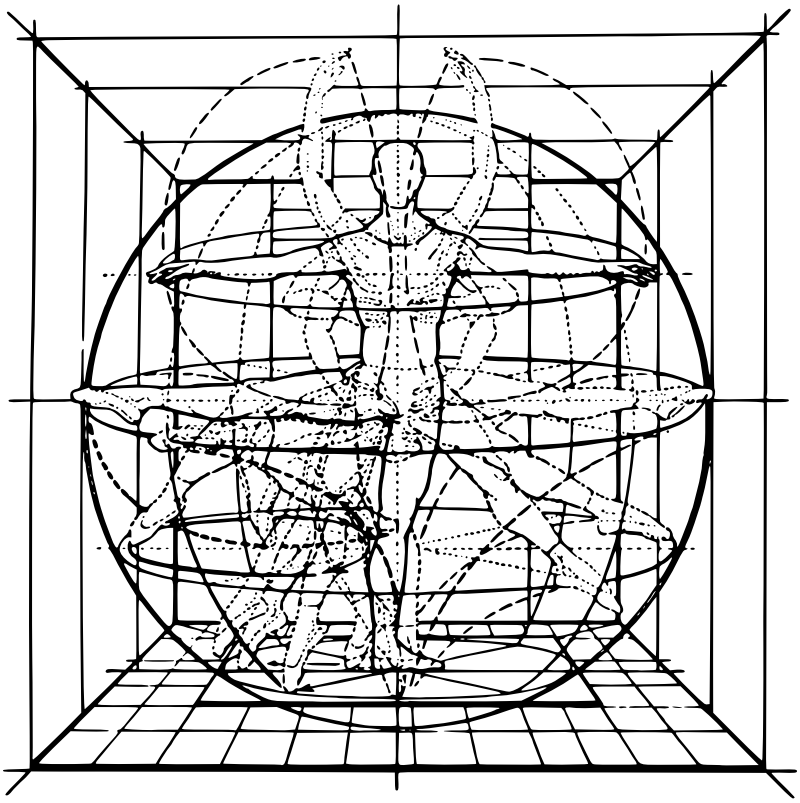


Figure 2.13: Kinespheres showing the human range of motion. Adapted from [Kir52].

expressibility. Especially hands and fingers are widely regarded as having the highest expressibility since we utilize them daily for multiple tasks requiring different amounts of precision, speed, and force. This fact is mirrored by the results of Penfield an Rasmussen and their homunculus [PR50]. It is crucial that the input devices are appropriately designed for their task since a device for tasks requiring high precision has to be designed in a way allowing the user to hold it in a way enabling precise interactions [LaV+17]. Zhai et al. show that using fingers for 6 DOF input results in shorter completion times and is thus preferable [ZMB96] since shorter muscle groups are being used. Finally, control tasks need not be limited to one hand, as many tasks require two hands, such as holding a tablet with one hand while interacting with the interface with the second hand. This kind of interaction is called *bimanual* as opposed to *unimanual* interaction [LaV+17], and the previous example also shows the central task division of such bimanual setups. One hand, typically the non-dominant hand supports the dominant hand during a task. This is supported by Guiard's findings that tasks are divided hierarchically, and the non-dominant hand supports the dominant hand [Gui87]. For example, selections can be done by pressing a button with the non-dominant hand to avoid the *Heisenberg effect* [Bow+01]. In the 3DUI context, this effect describes a jitter or disturbance of the tracked device when a button is pressed.

Besides the hands, various other limbs may be used for input. Most prominently, the legs and feet are used for movement within a VE. But, especially when considering weight shifts, works exist which use feet as an input device [Sch+09a], although feet are seen as only suitable for coarse actions [BKB97]. Nonetheless, especially when movement is desired, such as in many games, feet can be used as an input device, especially when coupled with pressure sensitive or discreet buttons.

One of the often neglected issues in 3DUIs is *fatigue*, although it has been well-known for a long time [LaV+17]. Fatigue is often called exertion or *Gorilla-arm-Syndrome*, when related to arm pain, in the 3DUI community and is mostly caused by repetitive or prolonged contractions

of muscles, whether due to uncomfortable pose or due to heavy instrumentation. The time a user can use a system depends on their physical endurance, their pain tolerance but also their bones and ligaments [LaV+17]. With increasing fatigue, the performance during tasks can decrease [Bar09; LBS15].

Signs of fatigue can be a change of posture, sagging of the shoulders, actively trying to readjust the position ease the physical exertion, such as supporting the acting arm with the other arm. Within this thesis, we often refer to the opposite of fatigue as *comfort*, and one of the goals of this thesis is to offer ways to create comfortable 3DUIs. Comfort was mainly measured using subjective questionnaires and through user observations. As previously mentioned, the NASA TLX measures the physical demand, the mental and physical effort and has been used to measure decreasing comfort, as well. Borg developed two scaled to measure the *rate of perceived exertion* (RPE), the Borg15 and Borg10 scales. These correlate with objective measures for exertion, such as heavy breathing, heart rate, and sweating, and have become one standard way of determining exertion [Bor82; Bor98] (cf. Section A.2.3). The Borg scale is one of the suggested standard measures for exertion, according to [Sta12]. Hincapié-Ramos et al. propose consumed endurance, a metric to measure fatigue during mid-air interactions which can be automatically recorded using devices necessary for tracking. They validate their metric using the Borg scale and NASA TLX and find a high correlation between their metric and the questionnaires [Hin+14]. Thus, if the necessary joints are tracked by the system, it can be used to either validate the questionnaire results or shorten the experiment by avoiding the questionnaires. A modified version of the Borg questionnaire for more specific target regions in the human body has been proposed by [Lin+12].

2.5 3D Input & Output Devices

In this section, we will present a summary of the technology used in this thesis, divided into input and output devices. For a more complete summary, we refer the reader to [BC03; LaV+17].

2.5.1 Input

Within this thesis, depending on the experiments, various kinds of input devices were necessary. For simple questionnaires and traditional input, traditional input devices were used, such as keyboards and mice. For a more detailed explanation about input in 3DUIs, cf. [LaV+17]. Unlike the output systems, we did not develop new tracking technology throughout this thesis and merely used commercially available solutions. In most cases, tracking of some part of the user's body was necessary. For example, in selection experiments, at least the finger position has to be tracked, meaning that a device has to be used that is capable of delivering reliable 3 DOF (x, y, z coordinates) of the user's finger. For most experiments, we used the Worldwiz Precision Position Tracking (PPT) system, a camera-based system that uses active infrared (IR) markers. Using at least two cameras with IR filters, the system needs a calibration phase with an active rig that has to be placed visibly for all the cameras. This rig has a specific shape, and the IR LEDs light up in a sequence allowing the software to calculate the extrinsic calibration. Once the calibration phase is done, the system can track IR LED markers in 3 DOF with submillimeter precision. One challenge with this is differentiating the various markers, especially if the system loses more than one due to occlusion. It is difficult to estimate which of the reappearing markers corresponds to the previously lost ones. The software doing the calculations has multiple plugins. For example, special IR markers pulsing with specific patterns can be used with a plugin to allow the system to differentiate between the markers and give them an ID. An important feature is the *Virtual Reality Peripheral Network* (VRPN) streaming, which is a protocol for streaming various types of data for VR applications [Tay+01]. Rigid-body tracking in 6 DOF (position and rotation) is also supported by the WorldViz software.

Due to the high precision and the broad support of the VRPN protocol in other software,

the PPT system was the preferred choice for most our experiments, albeit the requirement for instrumentation was there, and this instrumentation had to come with a power supply for the IR markers.

To acquire the rotation of an object in addition to the position data, an inertial measurement unit (IMU) can be used that measures the orientation of an object in space [Tha17]. The WorldViz software also includes a plugin allowing the merging data and then streaming 6 DOF tracking information. Such setups were used to track an HMD, as it is required to track position and rotation of a user reliably to avoid cybersickness (cf. Section 2.4.2.4). The HapRing project (cf. Chapter A) also supports an IR LED which can be tracked by the PPT system [ALS15]. Additionally, it included an IMU for rotational data. Nowadays, modern HMDs such as the Oculus Rift [Ocu13], HTC Vive [HTC17] include IMUs within the device, as well as external tracking for positional tracking.

The Oculus Rift set contains a factory-calibrated camera (Constellation camera) that detects the arrangement of IR LEDs hidden behind the front cover that then be used to reconstruct the HMD's position. The same type of camera, with the second one for more reliable tracking, can also be used to detect the Oculus Touch controllers. It offers submillimeter precision [Cue17].

The HTC Vive includes the so-called Lighthouse tracking system, within which two emitters have to be placed in a room. These emitters flash IR light and send fans of laser-light through the room that are detected by photosensors embedded in the HMD. The time difference, between the flash of light and the time when the light fans hit the device, can be used to calculate the device position in the tracking space. Additionally to tracking the HMD, the same approach and sensors are used to track the hand controllers for the Vive. The controllers are connected wirelessly to the HMD, meaning that only the connection from the HMD is necessary to the computer. Regarding the precision of these consumer-grade tracking systems there is one study claiming that, while the system latency is low, the Lighthouse tracking is prone to errors, such as a tilted ground plane that can change whenever tracking was lost temporarily. The comparison was made with a WorldViz tracking system. The tracking precision was stable with little jitter but always suffered from an offset by the tilted ground plane. Thus the authors conclude that for self-motion experiments requiring accurate visual stimulation the HTC Vive tracking is not suitable [NLL17].

The second type of dedicated tracking system we used was NaturalPoint OptiTrack systems [Nat17]. These systems work similarly to the PPT systems described above, but instead of having active markers, they have IR LEDs illuminating the tracking space mounted at or around the cameras while the markers are retroreflective. The calibration phase requires generating a point cloud by moving a calibration wand with three reflective markers and afterward setting the ground plane with a ground calibration rig. The software calculates the extrinsic matrices and is capable of tracking rigid bodies with at least three markers in 6 DOF with submillimeter precision. These 6 DOF rigid bodies can be differentiated, as the reflector count or their distances towards one another have to be unique [Ste+07]. These rigid body positions and rotations can be streamed to client computers with the software, for example, using the VRPN protocol. Additionally it is possible to stream single marker positions directly, however without a way to persistently differentiate the markers, similarly to the PPT. The advantage of the OptiTrack compared to the PPT system is that the markers do not need a power supply, allowing more extensive experiments without exchanging batteries. Also, the distinguishability of rigid bodies and the rotation tracking allows easier setup, as accidental switching of marker IDs is impossible. However, the rigid bodies are larger than PPT markers, which can be bothersome in studies concerned with touches, especially when only tracking of the position is necessary.

Finally, the recent Windows Mixed Reality HMDs offer inside-out tracking. The tracking works with two IR cameras and an IMU built into the device, constant feature detection from the camera, SLAM algorithms, and sensor fusion with the IMU to correct errors [Mic17]. While the devices

are tethered, they can also work outdoors as long as a mobile computer is included in a backpack or similar. However, currently no performance measurements are available, yet the prospects of a portable, wearable VR system which does not require a complicated tracking system and calibration are high.

The Windows Mixed Reality HMDs are based on the technological advancements made through depth based sensors such as the Kinect [Mic12]. We used the Kinect 2 sensor within our leaning-based projects (cf. Section 5.1). The Kinect 1 works by projecting structured IR light onto a scene in a specific pattern and by capturing the scene reflections, it can calculate the depth by comparing the captured image to the projected one [Sho+11]. The Kinect 2 uses a time-of-flight sensor data to acquire depth images [Kni13; MZC12]. Skeletal tracking can be done with both Kinects [Sho+11]. Alternatively, if leaning needs to be tracked, the Wii Balance Board can be used [Nin18]. Weight shifts and the center-of-mass can be detected using four pressure sensors in the feet of the device.

2.5.1.1 Hand Tracking

For some experiments only tracking of fingers or hands is necessary. There are commercial solutions for hand tracking, such as the Leap Motion [Lea12] and the system we used in [Lub+14a]: 3Gear, a company now bought by Oculus [3Ge14]. Both solutions are camera-based, but they used computer vision algorithms to detect hands in front of the camera instead of using markers. While the Leap Motion is a device with IR emitters and a camera, 3Gear was based on OpenNI and an RGBD camera, the PrimeSense Carmine.

Both solutions were able to track hands and provide the user with extracted positions of the various hand joints in 6 DOF. However, they only work in the small FOV of the camera and, especially for the Leap Motion, just up to a limited distance. Furthermore, while the Leap Motion provides high precision, the tracking volume is limited [Gun+14]. Also, both systems suffer from self-occlusion, meaning that the hand-tracking can get lost if the hand is not visible, either through a rotation occluding parts of the hand or through occlusion by another object. The occlusion problems are more severe than with previously mentioned tracking solutions, considering both systems rely on a single camera.

2.5.1.2 Touch

Touch interfaces are based on technology that allows the hardware to send a signal as soon as the user physically touches a device. The hardware then usually passes the touch location, as well as sometimes its size and strength of touch press. This input is generally updated at a particular frequency, and by taking the correlations of various touch points into account, a touch can stay on the device as long as the user keeps their finger on the sensors.

For multi-touch, various technologies are on the market, with projected capacitive having the highest market share as the dominant technology, especially in the smartphone market. Even amongst these, there are various technologies, but the dominant one is mutual capacitive, which works through electrodes in a two-dimensional grid. The capacitance at each intersection in the grid is measured individually, resulting in precise detection of multiple touch points, as a touch reduces the capacitance at an intersection.

While it is possible to use hand tracking to detect a touch, we used a dedicated touch detection solution within this thesis for more reliable results [Lub+14b]. The touch device used was a PQ Labs frame based on IR emitters and sensors built into a frame. While technical details on the patented system are not available, by taking information from [Wal12], and looking at the system through IR cameras we could determine they work through infrared (IR) emitters and receivers embedded in a frame on top of a display. Touch points are detected by reconstructing convex hulls of occluding objects. These are detected using the timing of the LEDs and a factory calibration of the component placement relative to another. These hulls can even be used to differentiate between shapes, assuming they have different convex hulls.

2.5.2 Output

Research done within this thesis was mainly conducted on HMDs and stereoscopic displays. Additionally, we developed a FOV extension for contemporary HMDs, which is why the output devices are discussed in more detail than tracking devices (cf. Section 5.4.2). For visual output devices a few specifications are essential, as they affect, for example, on distance perception [LaV+17]:

Field of view (FOV) and field of regard (FOR) FOV in the context of output technology is the maximum degree wherein the system can visualize content, or the amount of space a display covers. The field of regard is defined by the maximum degrees where content can be visualized around the user. A tabletop may have only 90° FOR, a three-wall CAVE may have a 270 ° FOR, and HMDs typically have a 360° FOR.

Spatial resolution Typically, displays have a fixed resolution, defined by how many pixels they have. The resolution can be specified in various ways, but usually, the absolute resolution is given in pixels. However, the screen size and the distance of a user to the display are also essential factors to determine the quality of output. A more viable measurement is dots per inch (DPI), which is a measurement of how many pixels are on one inch of space. Considering the limit of human vision is at about 28 arcseconds [Dee98], ideal displays should have this resolution.

Screen geometry The display geometry. While many different shapes are possible [LaV+17], within this scope, we focus on rectangular displays.

Light transfer mechanism Where the light originates. Computer displays typically emit light. If they are LCDs, they have some background light from the sides or in a grid from behind, whereas OLED displays emit light on their own. Projector based setups typically emit light onto a projection plane, either from the front or back, which in turn influences on whether interactions in front of the screen are desirable. Front projection setups will show shadows when objects, such as the user's hands, get in between the projector and the projection plane [LaV+17]. In that case, the projection will end on the occluding object that can also be used as an output for interactions.

Refresh rate Displays typically replace the show image a set count of times a second and are measured in Hz. While CRT screens needed a specific amount of time to move the electron ray over the whole screen, LCD and OLED screens can refresh the entire screen. Nonetheless, the refresh rate is often an upper limit to the latency of the system, as even if the rest of the system refreshes at higher frequencies, such as sensors measuring at 1000 Hz and the rendering offering 200 frames per second (FPS), the display refresh rate is often the limit. Higher refresh rates are preferable as higher latencies can cause side effects such as cybersickness [LaV00].

Ergonomics The weight, size and overall comfort of the device, for example, inadequate cushioning in a heavy HMD.

Effect on depth cues The depth cues (cf. Section 2.4.2.1) which the display can adequately represent. Monocular depth cues can be adequately represented by most displays, whereas some displays do not offer stereoscopic depth cues or require additional instrumentation, such as stereo-shutter-glasses. The main shortcoming is the accommodation-vergence conflict (cf. Section 2.4.2.1), which may be solved by Light Field Displays [HCW15].

2.5.2.1 Head-Mounted Displays

The term HMD is commonly used in the literature although there are researchers suggesting to replace it by head-worn-display to differentiate between displays which were mounted to pilot's heads [LaV+17]. To ensure consistency with the older literature [Sut68], we will continue using the term HMD. VR-HMDs allow fully-immersive VR (cf. Section 2.2). Over the course of this thesis, various HMDs have been released and used, starting with the Oculus DK1, DK2, Oculus

Figure 2.14: Brewster Stereoscope. Image adapted from [Nas14].

HMD	Resolution in Pixel	Resolution in Pixel per Degree	Diagonal FOV in pixel per arcminute	Refresh Rate in Hz	Lens	Display
Oculus DK1	1280 × 800	110	0.229	60	Aspheric	LCD
Oculus DK2	1920 × 1080	100	0.367	75	Aspheric	OLED
Oculus CV	2160 × 1200	110	0.374	90	Aspheric / Fresnel Hybrid	OLED
HTC Vive	2160 × 1200	110	0.374	90	Fresnel	OLED

Table 2.1: Comparison of the specifications of utilized HMDs.

CV and finally HTC Vive. These HMDs are based on so-called Brewster stereoscopes, which were invented in 1849 [Far94] (cf. Figure 2.14a). Brewster stereoscopes were hand-held devices with two lenses. Through a slot on the side, images showing two different perspectives could be inserted. When looking through the lenses, the user would get stereoscopic depth cues. The shape, size, and functionality of modern HMDs mirror the Brewster stereoscopes. In modern HMDs, a lens for each eye is used and images are displayed on one divided displays or two separate ones.

Table 2.1 shows a comparison of HMDs used today with the relevant specifications. The HMD quality influences distance misperceptions as demonstrated by [Cre+15] who compared an Oculus DK1 with an nVisor SX60 and found that distances were more accurately matched with the Oculus, although distance estimations were worse in the outdoor VE compared to the indoor VE they tested. A recent study has found that distance misperceptions are less severe with the HTC Vive compared to older HMDs (nVisor SX111 and ST50) [KCS17]. They also found a difference between outdoor and indoor VEs but also report that the distance estimations with the HTC Vive were close to the predetermined distances. Moreover, in an experiment Jones et al. found that a 150 degrees visual field significantly improved distance judgments in a VE compared to a reduced FOV of 60 degrees [Jon+12]. They used a Wide5 HMD and emulated different FOVs. Even

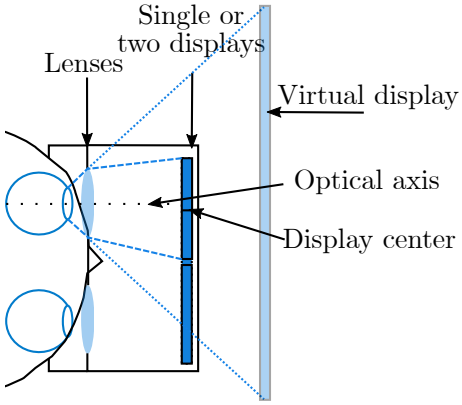


Figure 2.15: A typical HMD setup.

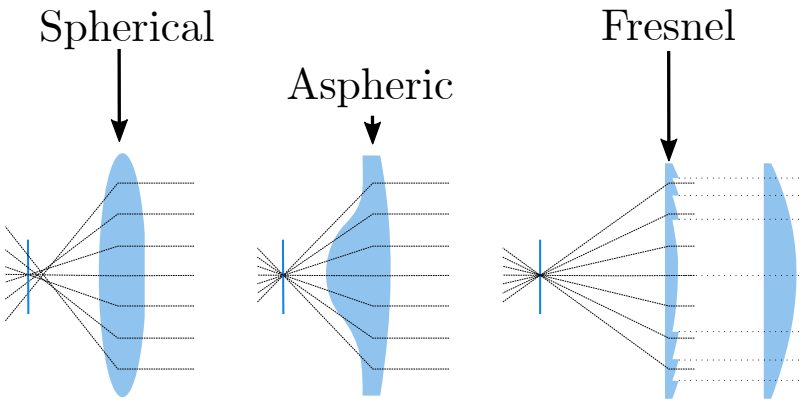


Figure 2.16: Lens types used in HMDs.

just static peripheral stimulation with white light in the far periphery of HMDs resulted in more accurate distance judgments and size judgments [JSB13]. These results lead to the assumption that technological improvements influence distance perception. Research from immersive [Lin+02] and semi-immersive [SVK07] setups agrees that an increased FOV can increase presence, although at the cost of potentially worse cybersickness.

As mentioned above, the HMDs used in this thesis all use lenses to create the impression of a wider FOV, similarly to some of the first HMDs [Fis+87] and the works that inspired the development of the Oculus Rift [Ols+11]. Figure 2.15 shows a typical setup for modern VR HMDs. The lenses refract the light in a way which makes the display appear larger and further away than without, enabling eyes to focus on the distance. Lenses are either aspherical, fresnel, or hybrids, as shown in Figure 2.16. The most simple lenses are spherical. However, they are heavy and suffer from spherical aberration, as shown in Figure 2.16. Aspheric lenses are lighter in weight and have a clear focal point. Fresnel lenses are even lighter than aspheric lenses, but they suffer from a glare effect due to the ridges in the lens. These round lenses refract the corners more than the outer edges of rectangular displays. These different refractions make a distortion of the image on the screen necessary to ensure it is perceived correctly for the user. This process of pre-distorting the image to ensure its correct visibility when viewed through a lens was devices by Fisher [Fis+87]. This distortion is called a barrel distortion and is natively supported by the drivers of these HMDs (cf. Figure 2.17). Formulas for the calculation of these distortions can be found in [RH95]. However, their parameters heavily rely on the lenses used in the display. Additionally to the barrel distortion, chromatic aberration has to be accounted for due to the different light wavelengths refracting differently. Chromatic aberration are caused by light of different wavelengths refracting differently

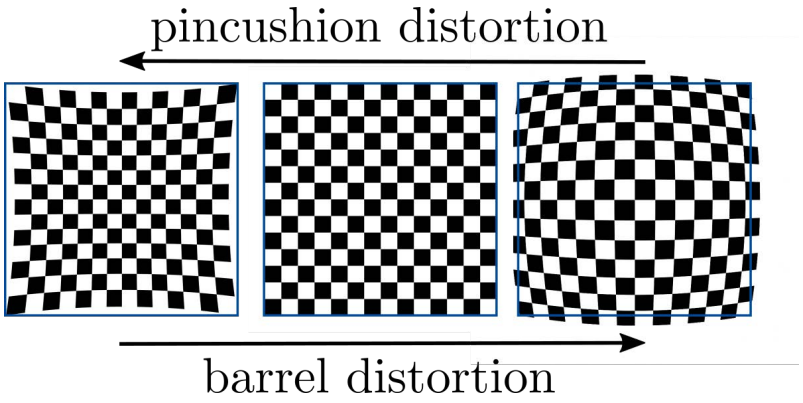


Figure 2.17: Barrel and pincushion distortion.

at lenses [Jav02]. This chromatic aberration splits the color channels of a display and can be corrected with a shader.

While the Oculus DK1 and DK2 did not have a physical setting for the IPD, both the HTC Vive and the Oculus CV offer a setting to match the lens distance to the user's IPD (see Section 2.4.2.1. For the effects of IPD on the VR refer to Section 2.4.2.1. It is necessary to render the scene twice, once for each eye, for the correct position and rotation of each eye. As Figure 2.15 shows, the optical axis is not necessarily in the center of the display, which means that it may be necessary to use additionally an asymmetric frustum. The Oculus SDK, for example, accounts for this[1]. One reference for the maths behind asymmetric frustums would be [Bou99]. Besides the IPD, the height of a user's eyes should also be taken into account during the calibration, as it affects distance perception, too [Dix+00].

2.5.2.2 Workbenches

The second type of devices we used in this thesis is a workbench or tabletop setup with the capabilities to show 3D imagery stereoscopically with correct head-tracked perspective. Workbenches allow semi-immersive VR (cf. Section 2.2). Workbenches and tabletops are often used synonymously in this context [LaV+17]. One of the first workbench setups for 3DUIs is the responsive workbench [Frö+95; Krü+95]. In addition to stereoscopic rendering, it offered 6 DOF head-tracking, thus allowing the user to walk around the virtual objects to view them from a different perspective, although restricted by the limited field of regard. For interactions, it also tracked the user's hand position using a 6 DOF tracker on a dataglove, which tracked the hand gestures. Our setup is a similar to the responsive workbench, and it is described in more detail in Section A.1.1[Lub+14b]. According to Brooks, workbenches can be seen as VR devices since they offer a large visual angle compared to desktop setups [Bro99]. According to Akka, such fixed setups provide a different kind of VR than HMDs, mainly since they do not always stay in front of the user's eyes [Akk93]. The user has less of an impression of being in the VE than looking at the VE through a window. However, to get the proper impression of a window, head-tracking is necessary. They argue that although sometimes the accurate geometry approach where the virtual camera position is moved precisely according to the user's position may offer the best results. However, they conducted studies letting users judge the preferred distances and found that users prefer exaggerated movements. A distinction has to be made between monoscopic and stereoscopic displays since the shown virtual objects can have varying depth. This means they can be in front of the screen (*negative parallax*), at (*zero parallax*), or behind the screen (*positive parallax*). This distinction results in images being displayed at different positions on the display depending on which for eye the image is rendered (cf.

[1] Oculus SDK `https://developer.oculus.com/documentation/pcsdk/latest/concepts/dg-render-advanced/#dg_render_advanced` retrieved 17.12.2017

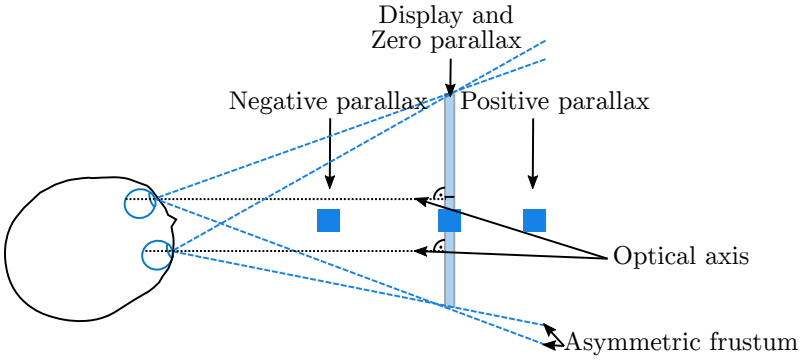


Figure 2.18: A schematic of a tabletop display showing the asymmetric view frustum.

Figure 2.18). Many workbenches are based on the concept of Fish-tank VR [WAB93], in which Ware et al. constructed a setup with head tracking through a mechanical rig and stereo rendering. They evaluated the system and found that users prefer the head-coupled perspective over just the stereo condition and that users made fewer errors in a tracing task.

The direct predecessor of our setup is the smARTbox, a setup using commercial products to create a VR workbench [Fis+12]. The smARTbox was built following the instructions in [Sch+10]. For the touch input, a rear-diffuse-illumination setup was used. Head-tracking for motion parallax was done using a Kinect or OptiTrack system. With a stereo-capable projector and the matching shutter glasses, smARTbox was able to provide the user with a fish-tank VR experience.

The iSPACE [Lub+14b] offers a similar setup, using commercial touch solutions, a stereo-3D capable TV screen and a PQ labs touch frame (cf. Section A.1.1). Head-tracking was done with a TrackIR consumer grade head-tracker and hand were tracked by a PrimeSense Carmine sensor and the 3Gear SDK. Both were later replaced by an OptiTrack system to acquire more precise measurements. Hand tracking allowed the use of hover interactions, the tracking state that is available in mouse setups but was not used in touch setups.

2.6 3DUI Basics and Related Work

Since in this thesis the focus is on 3DUIs, a more complex taxonomy is necessary. LaViola et al. divide 3D interaction into three categories based on Bowman and Hodges' work on categorizing interaction techniques for VEs [BH99; LaV+17]:

Selection and manipulation It is first necessary to indicate which object the user wants to interact with, to enable further interactions with virtual objects, such as changing their material or shape. Afterward, manipulations can be done.

Navigation The task of moving within and around an environment. It consists of *travel* and *wayfinding*. Herein, travel refers to the low-level motor tasks of moving in the environment, or more specifically, translating and rotating the viewport by actions such as simple walking, pushing a button to move forward and so on. Some travel techniques also allow the user to control the acceleration and other factors. Wayfinding refers to the higher level tasks, such as planning a path to avoid obstacles and wayfinding techniques are described as helpers which support the user in this cognitive task.

System control Changes the system state, for example, from color to shape manipulation. These tasks are also often on a higher level and can consist of selection and manipulation tasks. The symbolic input of numbers and letters also falls into this category.

It is noteworthy that these taxonomies [BH99] were defined for interactions in VEs before the definition for 3DUIs was made [LaV+17](cf. Section 2.1). Thus some parts of the taxonomy seem in direct conflict with the definition of 3D interaction. For example, according to the definition, the

indirect selection of an object from a list during a selection task is not a 3D interaction, unless a spatial interaction is involved. Hence, if the selection is made on a purely 2D UI with a 2D input device, the target selection through a list would not be considered a 3DUI. Nonetheless, Bowman's taxonomies provide a framework that can be used to categorize the contributions of this thesis.

2.6.1 Selection & Manipulation

LaViola et al. define manipulation tasks as acts of handling rigid spatial objects with one or two hands [LaV+17]. They intentionally leave out the deformation of objects and focus on rigid objects although this still leaves many variables open. They define canonical manipulation tasks that can be analyzed separately. The results thereof can be extrapolated to more specific tasks.

LaViola et al. suggest the following tasks as canonical manipulation tasks with parameters relevant to them:

Selection Designating one or many objects for further actions. Also called *target acquisition task* [ZBM94]. An analogy from real space would be picking up a pen to be able to write with it [LaV+17]. Selection depends on the *distance* and *direction* to a target since a selection within arm's reach can differ from selections outside. While a direct selection can be possible within arm's reach, a travel task or a different interaction technique may be necessary to reach other objects. The *target size* and the *cluttering* in the target area may also require various changes. Whereas a rough volumetric selection may be adequate for selecting one single object in an otherwise empty room, a more precise technique may be necessary for a cluttered environment. The number of targets and whether they are occluded or not may also require adjustments, for instance, either a travel task or a selection technique capable of selections behind occluders [SRH04].

Positioning Changing the position of the previously selected object(s). It is influenced by the initial and *target position* and direction of the target and thus the necessary translation to move the object there. Positioning may require a travel task or a more elaborate technique capable of moving the object over longer distances. Especially when considering high-precision placement, an ad-hoc ray based solution may place items inadequately over longer distances, considering the relationship of ray orientation and potential ray intersection points.

Rotation Changing the orientation of the previously selected object(s). It is often done in conjugation with positioning. Similarly to positioning, the distance of the object, the initial and target orientation of the object, as well as the required precision are essential factors.

Scaling Changing the size of the previously selected objects(s). This operation is typical in 3D interfaces, however, finding analogies to real space would be difficult. Stretching a non-rigid substance could show some similarities to scaling, however usually at the cost of rigidity, which was defined as a prerequisite [LaV+17]. The parameters are similar to the positioning task. Given a high distance to the target, finding an adequate scale could be difficult, especially in high-precision tasks.

The selection within VEs is a basic task of Bowman's taxonomy, as objects or locations within a VE need to be selected before a manipulation- or travel task can be done. The taxonomy is shown in Figure 2.19. The selection process is divided into indication which object is meant to be selected and an indication that a selection should take place. The first step can be done through touching, pointing, occlusion or framing or indirect selection. Pointing can be done in 2D, e.g., with a mouse cursor, or 3D by using a hand or a device or the user's gaze in 3D. Occlusion or framing is a technique to select by framing or occluding desired objects, whereas framing is, e.g., the box selection in common 3D design tools. Afterward, the item selection has to be confirmed, meaning that the user needs to do an additional step to indicate they want a selection to take place or, more specifically, they want the indicated item to be selected. Finally, the user has to receive some feedback, stimulating one sense. Within the scope of this thesis, this feedback is usually

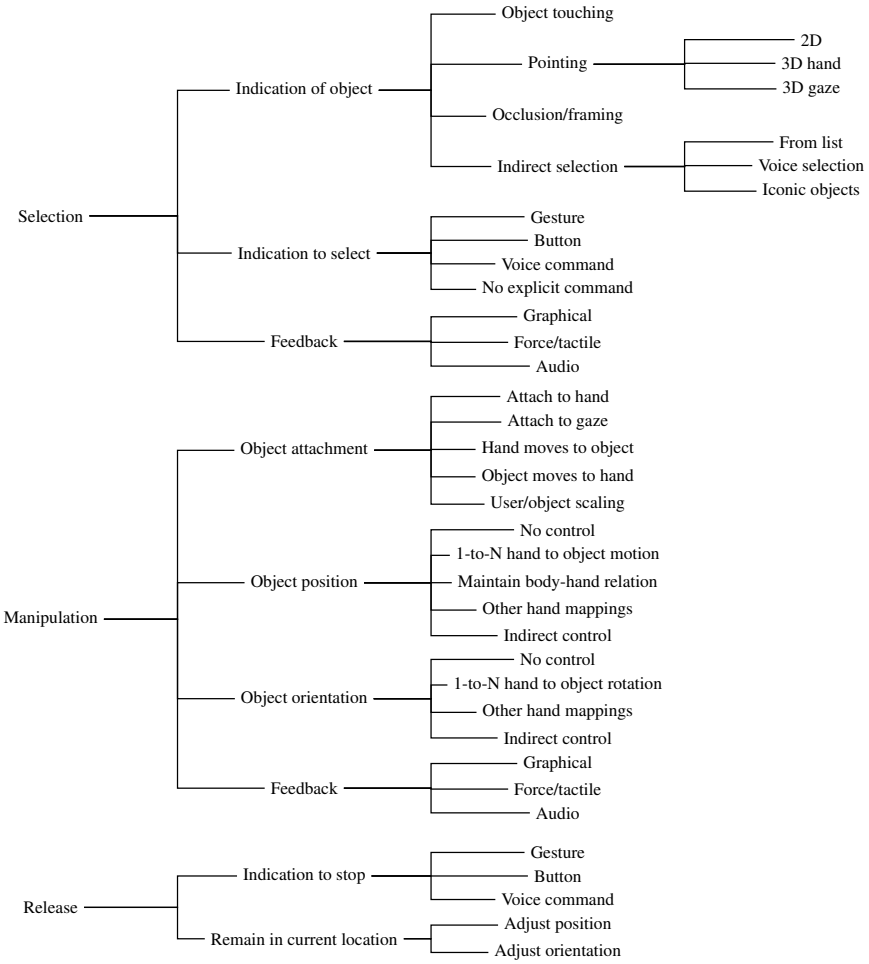


Figure 2.19: Selection/manipulation taxonomy. Adapted from [BH99].

visual, acoustic or vibration feedback.

Bowman and Hodges argue that it is possible to compare various techniques using performance measures but also develop new interactions by mixing them using their taxonomy [BH99]. For example, we mixed object touching to select an object with voice commands to indicate a selection and graphical output by color changing in [Lub+14a]. For these techniques, various performance measures can be conducted, such as error rate, speed, accuracy, throughput and subjective measures such as ease of use or comfort, as discussed in Section 2.4.3.

2.6.1.1 Fitts' Law and Selection

Within this thesis, performance is usually evaluated with a Fitts' law task or a task inspired by the traditional and ISO Fitts' law tasks. Fitts' Law is a commonly used performance measure for selection and describes the trade-off between speed and accuracy in selection tasks [Fit54]. It has been implemented into an international standard, describing a selection task, its evaluation and how to use it to design interfaces [Int00]. Selections by touching or grasping objects with a user's hands can be split up into two phases, the *ballistic* and the *correction* phase [LCE08]. The ballistic phase consists of focusing on the target object and bringing the hand in the proximity of the goal by using proprioceptive motor control. After that, visual feedback is used in the correction phase to reduce the distance from the hand to the goal incrementally.

Fitts' Law predicts the movement time MT for a given target distance D and size W [Fit54].

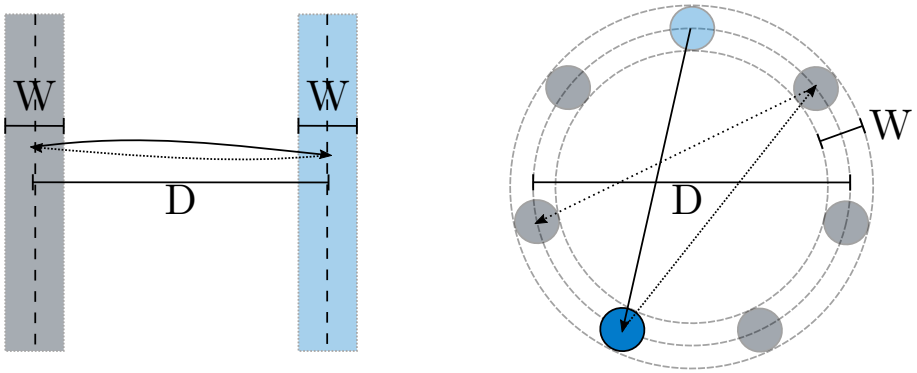


Figure 2.20: A classic Fitts' law task [Fit54] and a modern task according to the ISO standard [Int00].

They are brought together in a log term that describes the difficulty of the task overall with:

$$MT = a + b \cdot log_2(\frac{D}{W} + 1) \tag{2.1}$$

The values a and b can be empirically derived for different setups. The *index of difficulty* (ID) is given by the log term and indicates overall task difficulty; smaller or/and farther targets result in increased difficulty.

$$ID = log_2(\frac{D}{W} + 1) \tag{2.2}$$

The formula has been extended to get effective measures. The error rate is adjusted to 4% by resizing targets to their effective width W_e. This is supported by an international standard [Int00], wherein a task is described that controls the effect of direction. Both the old task, as well as the new task according to the ISO are shown in Figure 2.20. By calculating the average of the measured movement distances, D_e can be determined. With that, the effective throughput can be computed as a useful combination of speed and accuracy:

$$TP = \frac{log_2\,(\frac{D_e}{W_e} + 1)}{MT} \tag{2.3}$$

These formulas that are used in HCI research, are not the exact version described in [Fit54], but instead a variation proposed by MacKenzie, which closely resembles Shannon's information theorem [Mac92].

The validity of Fitts' Law for 3D interaction has been researched in the last years. Results from studies of several research groups imply that Fitts' Law is indeed valid for the kinematics of arm movements in a 3D interaction space [DKK07; Mac+87; MI08].

Also, Wingrave and Bowman [WB05] showed that Fitts' Law still holds when pointing in VEs. They observed that D was related to the amplitude of the movement, and W to the visual size of the target. Poupyrev et al. [Pou+97] defined the size of an object W according to the vertical and horizontal angles that the object occupies in the user's FOV.

Similarly, Kopper et al. [Kop+10] propose a modification of Fitts' Law as a model for human performance in distal pointing tasks. Their model is based on the angular amplitude and angular width. They argue that, contrary to classic 2D Fitts' Law tasks, the objects are floating in 3D space and the utilized sizes and distances depend on the user's position, which can be solved by using angular measurements.

Ha and Woo adopt Fitts' Law for 3D tangible augmented reality environments with virtual hand metaphors [HW10] by using the model established by Grossman et al. [GB04], which was based on a 3D objects arranged on a 2D plane.

Murata and Iwase [MI01] proposed an extension of Fitts' Law for 3D pointing tasks introducing a directional parameter into the model. Further studies investigating whether Fitts' Law can model object selection in 3D mid-air are described by Ware and Balakrishnan [WB94]. Teather and Stuerzlinger [TS11] found that using the simple Euclidean distance is viable enough for their system using a fish-tank VR system. Pfeiffer and Stuerzlinger compared electric muscle stimulation (EMS), vibration and visual feedback mechanisms for 3D selection with a Fitts' law task and found that EMS and vibration are reasonable additions to visual feedback [PS15].

Li et al. [Li+05] have shown an increase in performance when using the dominant hand for the selection, while the non-dominant hand should be used for the confirmation. Dividing the selection tasks reduces inaccuracies induced by hand tremors when pressing a button with the dominant hand while targeting with it.

2.6.1.2 Indirect Interaction

3D manipulation can be classified into *isomorphic* (direct) and *nonisomorphic* (indirect) techniques. The terms isomorphic and direct interactions/techniques will be used synonymously in this thesis. Isomorphic techniques adhere to a strict 1-to-1 mapping of hand position and orientation in the real world to the virtual world, which is often seen as more natural. Nonisomorphic techniques offer users sometimes supernatural abilities, such as elongated arms [Pou+96]. The 1-to-1 mapping does not have to be followed here, and performance and usability requirements can still be met, of improved upon strictly isomorphic techniques [LaV+17].

Different indirect interaction techniques have been proposed that allow users to interact with virtual objects in vista space by nonlinear scaling of hand positions within arm's reach, such as the Go-Go [Pou+96] and HOMER technique [BH97]. The Go-Go technique is a combination of isomorphic and nonisomorphic interaction. Up to a certain distance ($\frac{2}{3}$ of the user's arm length), the mapping of the virtual hand (R_v) and the real hand (R_r) was strictly 1-to-1 ($R_v = R_r$), and afterward a nonlinear function ($R_v = R_r + k(R_r - D)^2$) was chosen to elongate the arm. HOMER (**H**and-centered **O**bject **M**anipulation **E**xtending **R**ay-casting) uses a raycast to designate a target and then moves the virtual hand to the designated manipulable object. Afterward, the object can be manipulated directly with the hand until released. In particular, these techniques make use of the entire reachable space of a user's arms during interaction with distant objects. Using the entire reachable space may become exhausting when interacting at a distance, and may thus result in degraded performance over time [AA13]. However, it is often observed that indirect interaction techniques result in reduced performance during interaction with virtual objects located within arm's reach compared to direct interaction [MBS97].

One interaction technique that can be used for various purposes and is a combination of indirect and direct selection is the *World in Miniature* (WIM) [SCP95]. The concept here is that the user can to see a miniature version of the environment in front of them. This avoids the distance problems since the miniature can be mostly scaled in a way which constrains its bounds to the user's direct interaction space. However, as the user interacts with proxies of the objects in their VE, it should not be called direct interaction, although the user's interaction with the miniature proxies may be direct. The miniature world was attached to a tracked clipboard held by the user's non-dominant hand whereas the dominant handheld a tracked two button ball to interact with the proxies. Besides giving the user an overview of the VE, the WIM can be rotated rapidly to look at objects from various sides. Using these mechanics, they can select and interact with objects out of reach, and if the user is also represented as a virtual object with a virtual body, they can pick their representation up and place themselves somewhere else, thus enabling traveling (cf. Section 2.6.2).

2.6.1.3 Pointing

As an alternative to virtual hand approaches, pointing techniques have been proposed that do not utilize direct interaction with virtual objects but instead cast rays [Min95]. Each approach offers solutions to different problems, such as occlusion selection by manipulation of the ray's curvature with the off-hand [OF03]. Other pointing techniques utilize heuristics to bend rays towards targets [SRH04] or cast volumes instead of shapes to increase the selection performance [LG94]. One of the earliest techniques was Bolt's *Put-that-there* [Bol80] wherein users had to point at objects and the location they wanted to move the object to and express their commands using voice input to accomplish the task. A basic version of a pointing interface has reduced expressiveness compared to hand metaphors since they usually consist of a user holding a device, or a 6 DOF tracked device being attached to the user's hand or finger. However, the position is used as the origin of the ray, and two axes of the rotation are necessary to point the ray at a position. This only leaves one DOF to express the desired rotation, resulting in the object only rotating around the axis specified by the ray. The real world analogy would be a laser pointer. Pointing can also be done with the head orientation, especially in cases where hand or controller tracking are not available. This procedure is often called gaze-based ray-casting [LaV+17][2]. Such ray-based selections are often visualized using a line, emitted from the point of origin, such as the finger or hand. The lines are going in the pointing direction for an infinite distance since that allows users to merely intersect the ray with the target object without the additional cognitive demand of calculating the approximate intersection points of the ray [BH97]. There are many variations of pointing techniques [LaV+17]:

- Fishing reels, allowing users to pull an object closer, e.g., by pressing buttons [BH97].
- Volume-Based Pointing that requires less precision than rays, as they have a more extensive volume and need not touch the target object directly, such as flashlights [LG94]. Flashlights and their improvement, the aperture selection [FHZ96] are techniques currently implemented in 3D modeling software such as Blender, allowing quick selection of multiple objects, as shown in Figure 2.21.
- Automatic bending to the closest object to the ray [Rie+06; SRH04].
- An adaption of the mapping function allowing users to use a normal raycast until further precision is necessary. Then the control-display-ratio is changed by a factor of 10. This slight change maps a user's wrist rotation to a tenth of that, thus also increasing the possible precision tenfold [Kop+10].

However, due to human physiology limitations, such as hand tremors, there are also precision limitations in pointing interaction, especially for small objects or distant objects [Pou+98]. The abstraction necessary to use a pointer for selection and manipulation can be a hindrance depending on the application context. Within the scope of this thesis, we focus on the isometric or direct interaction.

2.6.1.4 Direct Interaction

According to Mine et al. [MBS97], direct interaction leads to significantly higher performance than manipulation of objects at a distance from the user's hand. Most results from similar studies agree on the point that optimal performance may be achieved when visual and motor spaces are superimposed or coupled closely [Dja98; WM99].

However, the influence of the position of virtual target objects located within arm's reach may affect interaction performance is still unclear [Hof+13]. As described in Section 2.4.2.1, direct interaction is subject to perceptual limitations, e.g., the vergence-accommodation mismatch, ghosting or double vision. These limitations can result in strong misperception effects [BSS13a; BSS13c; Cha+10]. Depending on the location of virtual objects, users may be unable to discriminate

[2] Microsoft Hololens Gaze `https://developer.microsoft.com/en-us/windows/mixed-reality/gaze` retrieved 09.12.2017

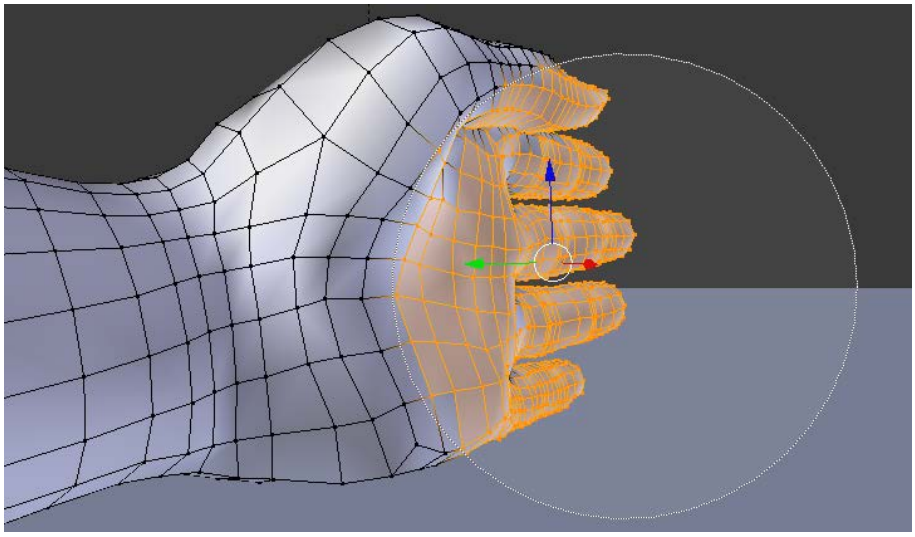


Figure 2.21: Circle select tool in Blender, which is an implementation of aperture selection [FHZ96].

interrelations or perceive distances to objects to be smaller or larger than they are displayed [Hof+13; LK03]. Furthermore, Hofmann et al. [Hof+13] show that some regions within the user's arm's reach are considered to be more comfortable than others by users without finding any differences regarding performance. Such distortions do not appear in the real world and may be related to limitations of current technology to correctly reproduce natural cues from the real world in a perfect way [Wil+09].

Moreover, internal representations of the 3D space are influenced and updated by both visual as well as motor input, which may affect interaction performance [Tho83; Wil+08]. Due to varying energy expenditure between users based on differences in strength and endurance of arm muscles, interaction performance in mid-air within arm's reach in IVEs may be affected by different factors related to the ergonomics of direct interaction. In particular, contributing factors may include interaction duration, hand and arm postures, the frequency of movements, and comfort [AA13].

The space above interactive multi-touch surfaces has been considered for different user interfaces with stereoscopic displays, in particular using direct touch for objects displayed on the surface as well as 3D mid-air touch for objects that are displayed with negative parallax above the surface. Bruder et al. [BSS13c] found that users tend to incorrectly perceive the 3D position of stereoscopically displayed objects with negative parallax when touching these objects by moving their finger inside their perceived 3D shape.

2.6.1.5 Tabletop

There are tabletop setups and other devices capable of touch sensing that can also display at least a monoscopic or even a stereoscopic 3D image [Lub+14b]. The interaction with content shown on these devices can be seen as a specific subclass of 3DUIs While some of the typical 2D interaction metaphors were explained in 2.3, here we will focus on 3DUIs. Many of these 2D interaction techniques can be used in monoscopic 3D, such as the two finger translation and rotation described in [HCC07; Han+06].

Within this thesis, we focus on the stereoscopic displays. Considering the different positions on the screen, touch-interaction with these stereoscopically displayed objects is challenging, as only for objects at zero parallax, the projections for both eyes match. Only for zero parallax it is thus apparent to a user where to touch. For other objects, the user has to touch the surface at a point either between the projections for both eyes, at the projection for the dominant eye, or at the projection for the non-dominant eye [Val+11]. Negative parallax allows the use of mid-air

tracking to allow the user to simulate a touch with objects once their finger is in contact with the virtual object, depriving them of the passive haptic feedback a touch surface offers. However, the accommodation-vergence conflict is still a factor, as the user has to focus on the screen whereas their finger may be at a different height and should the user focus on their finger instead the rendered image will not be focused. This behavior was investigated by Bruder et al. who did a Fitts' law task and compared various depths on a stereoscopic tabletop display for selection on the touch surface and 3D mid-air selection [BSS13c]. While they found a significant influence of a target depth on the precision in the mid-air condition, they also found a significant influence of the height on the touch area in the touch condition. Different groups of users tend to touch to different projections, either to the dominant, non-dominant or right in between. They suggest using the center between both projections for objects of a depth below 10cm and also that 2D touch is preferable to mid-air interaction for these distances. For higher parallaxes, a more precise calibration would be required, but the disadvantages suggest that using mid-air interaction offers higher precision in these cases. In line with guidelines proposed by Schöning et al. [Sch+09b] this underlines that multi-touch interaction with stereoscopically rendered objects is mainly limited to a distance of about 10 cm from the plane of the interactive surface.

In addition to the analysis of touch regions, various 3D interaction techniques were developed on tabletops. The pinch gesture, which is used to enlarge or shrink the touched object in classic 2D cases [WMW09], can be used to change the depth of objects in 3D environments, which has almost the same effect of visually increasing and decreasing in size [GVH14]. In their void shadows technique, Giesler et al. use a projection of objects behind the display plane to generate a visible, interactive representation of the object on the touch-sensing display surface. These projections, called void shadows, all converge at one shadow plane above the display, resulting in projections that are smaller than the object. Additionally, this results in closer objects of the same size having a larger shadow than distant objects. This allows users to differentiate between these objects and touch the desired object, manipulating it with the standard rotation and translation gestures but also changing its depth by pinching. This technique was created for objects below the touch surface or at positive parallax. To select objects above the touch surface, Benko and Feiner suggested a balloon selection technique, based on two touch points [BF07]. The first touch point works as an anchor for a spherical cursor, the height of which is adjusted by the second touch point, using the metaphor of a balloon on a string. By moving the second touch closer to the anchor, the spherical cursor moved up and vice versa. Thereby, the 3 DOF selection task was reduced to two 2 DOF tasks, offering a higher selection speed than a keyboard technique and fewer errors than a simple virtual hand metaphor. Daiber et al. presented a single-handed corkscrew widget, which is a circular 2D widget that can be placed below an object [DFK12]. After placing it, the outer edge can be rotated around its middle to adjust the selection height. They compared their widget to the balloon selection and found that their technique offered higher selection speed at lower precision.

Strothoff et al. developed a single-handed gesture called a triangle cursor and compared it to the balloon selection [SVH11]. By placing two fingers on the touch surface, the user could define two points. A third point right in between both fingers was shown and used for the selection. Based on the distance of the fingers from one another, the height of the selection cursor above the center point was changed, up to the maximum selection height in the given 3D environment. Their solution offered lower selection times, higher precision and higher rotational precision, possibly due to smaller muscle groups being utilized in the single-handed gesture.

Thus, much research was concerned with the interaction with virtual objects shown on stereoscopic displays. However, as Buxton describes in his three-state model of graphical input, for traditional input devices, like a mouse, there is a so-called tracking state, as in the state where the cursor can be moved without pressing a button. The position and movement of the cursor can be directly transferred to input in user interfaces [Bux90]. One of the most common uses of such a

tracking state is for so-called hover effects in classic 2D user interfaces [Bux90]. The tracking state is often used for highlighting or tooltips and can declutter interfaces by providing context-sensitive information. In most touch interfaces, for example, using capacitive sensing technology, this tracking state is missing [Gro+06; PHH11]. While dragging is possible, when pressing down on the screen, the hovering usually cannot be detected. Especially during the design of touch-enabled tabletop and mobile user interfaces, such as mobile versions of websites, the missing tracking state becomes evident and many design principles, such as flat design, become complicated [Gro+06]. To compensate for the missing hover capability in 2D multi-touch setups Benko et al. [BWB06] simulated a hover state with techniques that make use of a second finger. The second finger adjusts the control-display ratio while the first finger controls the movement of the cursor, thus resulting in more precise selection.

Beyond hover interaction in multi-touch setups, other input modalities have been investigated as well. Grossman et al. [Gro+06] presented Hover Widgets that extend the expressiveness of pen-operated touch surfaces by using the tracking state of the pen as hover input. In particular, they proposed a particular hover technique that activates a widget by a small discrete gesture that is followed by a pen-down action. The Hover Widget technique shows that the space between the hover state and the touch state can be effectively used. With recent advances in 3D sensing technologies, it becomes possible to track a user's fingers above touch surfaces [Ber+16; Hil+09a]. This allows using a tracking state to be leveraged for user interfaces. For example, Wilson proposed using a depth camera as a touch sensor [Wil10].

Different interaction techniques have been proposed for monoscopic and stereoscopic display environments which make use of this capability. Han and Park [HP12] explored hover-based zoom interactions in monoscopic display environments. They proposed a technique that relies on a magnifying lens metaphor. This approach allows users to quickly zoom in and out in a restricted range of multiple zoom levels that are defined by layers above a multi-touch display. Initial evaluation results revealed that their technique outperforms the common pinch-to-zoom technique in both speed and user preference. However, in their implementation zoom layers were discrete and no continuous zooming in a 3D hover volume above the multi-touch display was possible.

Echtler et al. [EHK08] presented a multi-touch tabletop that was extended with a ceiling-mounted light source to create shadows of hands and arms. By tracking these shadows with the rear-mounted camera of their frustrated total internal reflection (FTIR) setup, they augmented the multi-touch tabletop with mouse-like hover behavior. With this setup, users can control multiple cursors by hovering above the tabletop and trigger a "click" event when touching the surface. They evaluated their system concerning tracking accuracy. The evaluation indicated that users were aware of their hand position above the display and tried to avoid occlusion by orienting their hand in an unnatural pose parallel to the edge of the tabletop so that the cursor pointed perpendicular to the user's viewing direction.

Annett et al. [Ann+11] presented Medusa, a proximity-aware multi-touch tabletop that is capable of tracking multiple users and differentiate between their hands. Besides supporting collaborative multi-user settings, they proposed different hand-dependent hover techniques. Hovering with the right hand above the display triggers a marker below the hand that turns into a component-specific marking menu when it is touched. Hovering with the left hand displays an icon that deletes a component when it is touched. The prototype by Pyryeskin et al. [PHH11] uses light reflected from a person's palm to estimate its position in 3D space above the table based on the diffused surface illumination vision-based principle.

2.6.2 Navigation

Throughout this thesis, various travel techniques are utilized and investigated, especially in Chapter 5. According to LaViola et al., Navigation is a process consisting of wayfinding, which was discussed in Section 2.4.3, and travel, the motor task of moving around in a VE [LaV+17].

Additionally, LaViola et al. provide other classifications of travel tasks, for instance, whether they are exploration, search or maneuvering tasks [LaV+17].

Exploration A task where the user has no goal except to explore the environment. This mirrors the first cognitive steps to create a mental map. It is essential that the travel technique does not distract the user during this stage nor force them to go along a specific task unnecessarily.

Search A user actively searches for something specific within the VE. A distinction has to be made between naive search tasks where the user does not know the path or location of the target and primed search tasks where the user has prior knowledge of the target position.

Maneuvering Local and precise measurements, for example, to avoid small obstacles during the more extensive search or exploration task or overall any task that requires the user to move precisely within the VE, for instance during a docking task.

They also describe factors that are essential for travel tasks. First of all, the distance to be traveled plays an important role since if the user only has to look through a virtual window, head-tracking may be enough. If they have to travel extreme distances, for example, in city simulations or applications such as Google Earth VR [3], at least a way to control the movement velocity is necessary, for example, through scaling of either the avatar or the environment [LaV+01]. The amount of curvature is the second factor. Environments with many turns could be more easily traveled through by using hand gestures for turns compared to whole-body gestures, which are more tiring.

The visibility of the target is crucial for direct selection based travel metaphors, such as teleportation. A gaze-based technique may hinder the traveling during a search task since the user cannot look around at the same time as they are moving in a specific direction because the travel direction depends on the gaze direction. The available DOF for movement in the VE are another limiting factor. A simple gaze-based flying technique can make traveling in an approximately flat environment unnecessarily tricky since the user has to ensure they do not fall through the ground. Thus, restrictions such as collisions might make traveling more accessible, and a technique that completely restricts movement in the unused axis would be even easier to use [LaV+17]. As mentioned in the maneuvering task, the necessary precision is another crucial factor. A quick teleportation technique may be inadequate due to the imprecision of the pointing metaphor and world in miniature techniques may have insufficient precision due to the small scale of the object [SCP95]. Finally, distraction can play a role, and traveling can be a secondary task that should not distract the user from their primary task, e.g., evaluating the light and proportions in an architectural visualization [LaV+17].

Bowman et al. proposed a taxonomy of travel techniques [BKH97] that divided traveling into subtasks, similarly to selection techniques discussed in 2.6.1. The goals were to find the single steps which have to be done in a technique. This allows replacing these steps and evaluating different ones against one another. According to Bowman et al., the quality factors for immersive travel techniques are [BKH97]:

Speed Whether the velocity is appropriate

Accuracy How close it is possible to get to the target using the technique

Spatial Awareness How precisely the user knows their position and orientation in the VE during and after travel

Ease of Learning How easily novice users can learn the technique. Can be quantifiable by measuring the time until a specific performance or confidence is reached.

[3] Google Earth VR `https://vr.google.com/earth/`

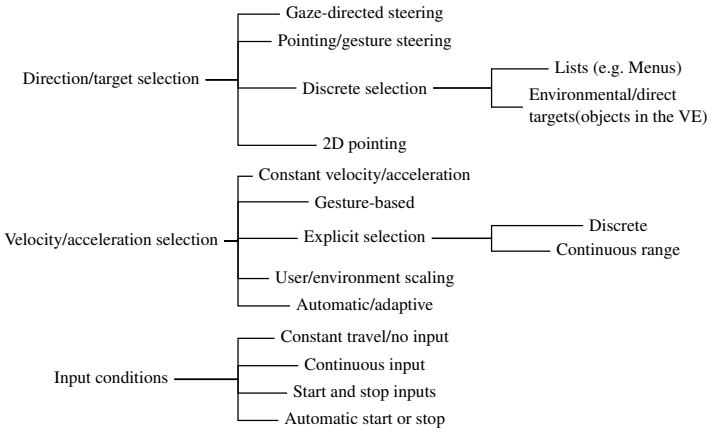


Figure 2.22: Taxonomy of VE travel techniques. Adapted from [BKH97].

Ease of Use How difficult the technique is from a user's point of view. Can be measured with subjectively, e.g., with the NASA TLX or by measuring performance using an N-back task (cf. Section 2.4.3).

Information Gathering How well the user can gather information about the environment during travel

Presence The sense of presence (cf. Section 2.4.2.4)

User Comfort How tiring the technique is and how much cybersickness it induces (cf. Section 2.4.4)

Instead of decomposing the travel task, LaViola et al. provide another classification of travel techniques into walking, steering, selection-based and manipulation-based travel [LaV+17]. We will use this classification in this thesis, focusing on walking and steering travel, which we evaluated and used in this thesis. To begin with, we will describe selection- and manipulation-based travel since they are partly based on the selection and manipulation techniques explained in Section 2.6.1

2.6.2.1 Selection-Based Travel

Various methods of selection can be used to determine the travel location in this metaphor. The most straightforward way is to specify a single location, optionally through a menu or by directly interacting with the scene. Arguably the most popular supernatural travel technique employed in the commercial sector is the teleportation metaphor [4]. Often implemented using a straight or curved pointer for the target position selection, the teleport metaphor does not require the user to actively control the whole moving process to a new position in the VE. According to LaViola et al., this is especially useful when the movement to the objective may not be the primary task but only a means to reach a destination where the higher-order task has to be fulfilled [LaV+17]. In work describing the task-based taxonomy for travel techniques, Bowman et al. found that instant teleportation results in a loss of spatial orientation [BKH97]. Within the third experiment they describe in [BKH97], they evaluate the acceleration selection and also include infinite velocity (i.e., teleportation), which they call *jumping*. As a result of this, the movement is body-centric and not head-centric, meaning that the user's point-of-view will not end up at the exact spot they selected on the ground, but their virtual body (i.e., their feet) will. To evaluate the spatial presence, they devised a task where the user was moved within a VE wherein a cube with a letter was shown and asked the users to press one of two buttons corresponding the letter. The time the users needed to find the cube and press the button was statistically evaluated between the tested techniques, with

[4] Oculus Sample Framework `https://developer.oculus.com/blog/introducing-the-oculus-sample-framework-for-unity-5/`, accessed 02.01.2018

standard velocity, ten times velocity, infinite velocity and a slow-in-slow-out velocity. The results showed that the infinite velocity was significantly worse. These results may be one of the reasons why, for example, Microsoft suggests using a rapid motion towards the target instead of a direct jump for immersive mixed reality applications [5].

A more recent and subjective evaluation was done in [Boz+16], wherein an HMD setup and instant teleportation with target selection by pointing was compared to a joystick and walking in place. They measured the time to reach a given destination and found that the teleport was the fastest technique without obstacles and the slowest when obstacles were present, possibly due to the fixed selection time they implemented. The teleport technique also resulted in the least amount of collisions, but they asked many subjective questions on Likert scales to further evaluate the different techniques. They did not find significant differences, except that the joystick metaphor was easier to understand than the walking in place and the walking in place resulted in the highest tiredness ratings. In a second experiment, they implemented an extension to the teleport, allowing the user to set the direction they would be facing after the teleport. However, users found it more difficult to understand, and it caused disorientation without offering performance benefits.

Zeleznik extended upon this simple teleport by developing a device with so-called pop through buttons to allow a user to travel to a location by pressing a button lightly and immediately return upon releasing, or press harder to trigger a second button press (the pop through button) to lock the position [Zel+02].

According to LaViola et al., users can also specify a whole path instead of a single target location and travel along this path after creating it, either by drawing it or specifying key-points [LaV+17].

2.6.2.2 Manipulation-Based Travel

Instead of moving the position of the user's virtual body within the VE, the user can also change the whole VE relative to their body position, or a representation of their body within the VE [LaV+17]. Overall, many manipulation techniques can be used for travel and may tidy up the interface, as the action used to manipulate the environment can be reused [LaV+17].

Ware and Osborne suggest various travel techniques, most prominently the *eyeball in hand* technique. The eyeball in hand technique provides the user with a mental model of moving a hand-held camera that they can use to position the virtual camera and thus their viewpoint [WO90]. They also evaluated a flying vehicle control that mostly fits into the field of steering metaphors. Furthermore they evaluated a *scene in hand* metaphor that closely corresponds to the world in miniature model discussed in Section 2.6.1.4 [SCP95]. The evaluation was done using a 3D scene shown on a 2D screen and not in an immersive setup. The mainly subjective experiment did not yield a single best strategy but instead seems to emphasize that different users have an easier time to adjust to specific metaphors and the task profoundly influences how well a metaphor was perceived.

LaViola et al. also introduce fixed-object manipulation allowing the movement of the user's viewport constrained to a movement around a specific object [LaV+17]. The difference here is that the object will remain at a fixed position and only the viewport moves around the object [Pie+97]. Chung describes this navigation as orbital mode [Chu93], and they tracked the user's head orientation to orbit the point of view around the object, always keeping it in focus in the center of the viewport. In their study, the orbital mode compared the best compared to other input techniques, such as joysticks and mice. It yielded significantly better scores than what they called immersive mode, a mode where the user's point of view was within the target object, and they were looking outwards, using only the orientation. For the task completion rates, the mouse resulted in the best rates, followed by orientation. The orientation was only significantly better than the walkaround technique, the classical mode allowing users to walk around the target object.

[5] Comfort, Microsoft Development Center `https://developer.microsoft.com/en-us/windows/mixed-reality/comfort#self-motion_and_user_locomotion`, accessed 02.01.2018

Such navigation around objects is an integral part of navigation within desktop 3D development software, such as Unity [Uni13]. Whereas the result may look similar when the viewport is moved compared to the scene being moved, we found that moving the camera may have practical performance advantages. Since many current 3D engines are built on a hierarchical system, moving world may require a traversal of all objects below the moving object to update the scene graph [Hae+16]. This traversal is not necessary when just moving a single camera object.

2.6.2.3 Steering

Another common movement technique for traditional desktop 3DUIs is steering, meaning that the user specifies a direction in which the viewport moves within the VE [LaV+17]. This technique is especially prevalent in desktop 3DUIs and 3D video games since 1993, usually allowing the user to turn the viewport using the mouse by mapping the mouse axes to the pitch and roll axes of the camera. Keyboard buttons are mapped to movement straight forward and back, up and down or left and right [6] [Min95]. However, this gaze-directed steering can also be employed in immersive environments since the head orientation can be directly used and another input device can be used to the translation. The translation can be either unrestricted, i.e., flying, or restricted to a virtual body, for example, by employing terrain collisions to ensure the user's virtual body stays above ground and cannot walk through walls. Strictly gaze-directed steering only allows translation of the camera in the view direction, and the extension of strafing and moving up and down was developed by [CH02][7]. Using the IMUs included in modern HMDs it is easy to implement gaze-directed steering [Lub+16a], but some movements may feel uncomfortable for users, such as directly moving up or down, turning multiple times or trying to move straight forward without drifting higher or lower [LaV+17].

Very similar to the extended gaze-directed steering is the pointing or hand-directed steering [Min95], wherein the user defines their movement direction by changing their hand or finger pointing direction. Alternatively, an input device can be tracked and used. Whereas gaze-directed steering is egocentric, meaning that the strafe movement would be perpendicular to the view direction, pointing allows a decoupling of the user's view direction and their movement direction and a change of view direction does not have to be compensated to ensure the final travel direction remains the same.

Mine et al. propose using two hands to indicate the direction, whereas the distance between hands indicates the speed and a minimum distance of 10 cm indicates no motion [MBS97]. However, no in-depth evaluation of this technique was included [MBS97]. The Rocket-Belt metaphor is a bimanual steering metaphor, used at a large stereoscopic display, where the controller in one hand controls the flight direction, and the other controller can be used to climb or descend [SBH07]. LaViola et al. propose using the torso instead of the head for the steering direction since it would decouple the gaze from the walking direction [LaV+17].

Finally, steering metaphors can be thought of as controlling a vehicle in a VE, and such vehicles often are controlled by an input device, such as a steering wheel or a joystick [LaV+17]. Such input devices can be used as physical props for a VR application to provide additional haptic feedback and utilize the user's prior knowledge of how to use these input devices. LaViola et al. provide several examples of controls based on what they call cockpits, meaning that the VE is a representation of a real vehicle and specific input components for that vehicle are used as physical props within the application [LaV+17]. Even without actual physical props, gestures known from real vehicles can be utilized for VR applications, as we show in [Sch+16] where we utilize a gesture derived from slapping or shaking reins of a horse (cf. Chapter 5). However, previous studies have

[6] Navigation in the Unity SceneView `https://docs.unity3d.com/Manual/SceneViewNavigation.html`

[7] Arguably, some of these steering techniques were existed in non-immersive 3D video games before that. See System Shock, a Looking Glass game from 1994 that allowed looking around including up and down, in contrast to games like DOOM (1993) and environment-constrained movement including strafing.

shown that steering metaphors cause higher cybersickness than more natural travel metaphors, such as walking [Uso+99a].

2.6.2.4 Leaning

Leaning-directed steering [LaV+17] requires users to lean in the direction they want to travel in. These techniques are based on the fact that users often shift their weight to steady their stance against opposing forces, for instance when standing in public transport. The Joyman interface used a device for users to stand in, with a rail to hold onto when they were leaning, allowing them to lean their whole body in a direction they wanted to travel in and using the leaning angle to manipulate the speed [MPL11]. Krujiff et al. have found in experiments that leaning forward or backward statically, meaning keeping a certain incline, has a positive influence while dynamic leaning shows that users move further than intended, showing mixed results [Kru+15a]. In another study, they investigated the influence of sensory substitution on the sense of self-motion and their sense of presence during travel tasks [Kru+16]. Users were either standing on a Wii balance board (cf. Section 2.5) to travel through the VE by leaning in a direction, as well as turning while moving forward.

Several interfaces have already used Wii Balance Boards in the field of virtual traveling, such as the Silver Surfer [WL11] interface that treats shifts in weight as on a surfboard in the real world and extended it to 3D traveling. Other approaches such as an interface proposed by Hilsendeger et al. [Hil+09b] use a Wii Balance Board to approximate the user's leaning direction. Valkov et al. demonstrate ways to use the weight distribution provided by the Wii Balance Board to recognize foot gestures for an implementation of a human-transporter [Val+10a]. Williams et al. [Wil+11] use it to track the user's steps in a walking-in-place implementation. Alternatively, a gamepad with a joystick was used in a seated condition. The authors investigated adding footstep sounds, vibrotactile foot-sole stimulation matching the steps and visually adding head-bobbing that matches the movement speed and found that adding these cues significantly increases the user's sense of presence.

2.6.2.5 Walking

Within Chapter 5 we present a walking travel technique that meant to give users a sense of security within a VE to allow them to walk more naturally. Walking is considered the most natural travel technique, as it is the way humans travel through real environments and thus offers various sensory cues, such as proprioception and vestibular cues (cf. Section 2.4.2). Still, it is not always feasible in VEs due to real obstacles, limited tracking spaces, cable lengths and other real constraints [LaV+17]. This limitation is only problematic if the VE's size is bigger than the available real tracking space. For many applications the VE can be designed in a way that completely fits the tracking space, naturally restraining the user. For example, a virtual assembly line for training purposes can have the same size as the tracking space. First prototypes exist that allow users to see virtual content in the real world while walking through it [Höl+99] and thus naturally exploring an environment. To enable walking as input in VEs tracking the user is necessary [Ste+13] (c.f Section 2.5). Besides the sensory advantages and the naturalness of the interface, practical advantages include having free hands for other interactions [LaV+17]. Even walking-in-place was found to offer a higher sense of presence (cf. Section 2.4.2.4) than steering metaphors [SUS95]. Usoh et al. found that walking in VR offers a greater sense of presence compared to walking-in-place (WIP) and gaze-directed flying [Uso+99a]. Whitton et al. compared real walking, walking with restricted FOV, virtual walking, joystick and WIP in a task requiring the participants to walk up to a wall [Whi+05]. They found a significant effect on various factors, such as peak velocity, peak deceleration and distance from the target, resulting in the virtual walking technique being the best after the real walking and real walking with restricted FOV conditions [Whi+05]. Langbehn et al. used to better self-motion perception of leaning to improve upon WIP techniques, using the leaning angle to manipulate the

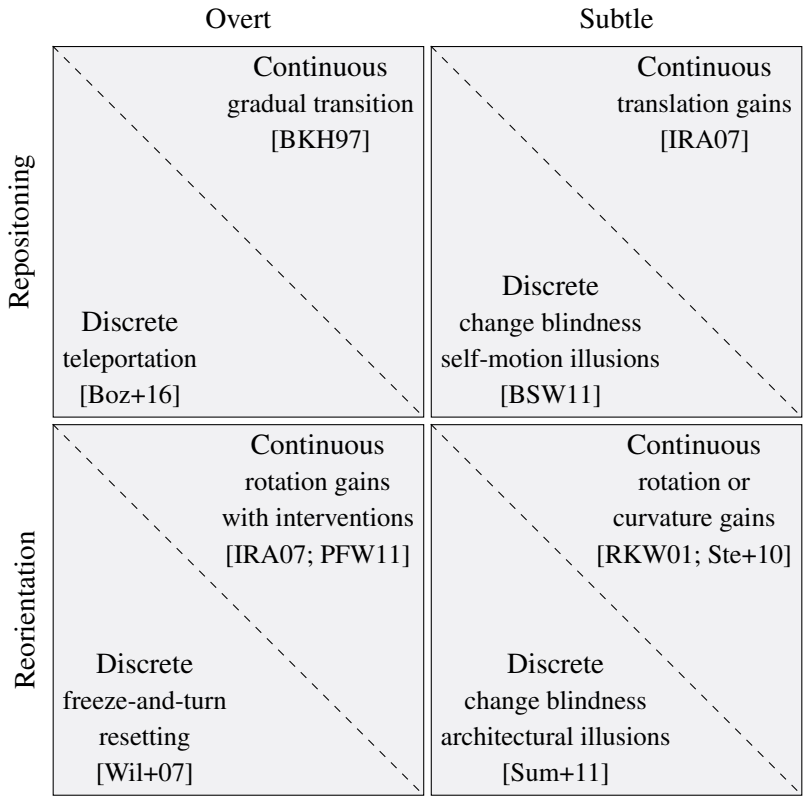


Figure 2.23: Taxonomy of redirection techniques [Sum+12].

speed while walking in place [Lan+15].

A study has shown that using a walking interface also improves performance during search tasks compared to tethered HMDs or desktop displays [RL09]. They concluded that only the walking interface provided performance comparable to a real-world scenario since only visual cues are not enough. They also compared the visual fidelity and found no benefit from using a rich visual fidelity compared to a impoverished one for their search task [RL06]. Williams et al. found a significant influence of using walking compared to joystick traveling on spatial orientation through the means of a blind pointing task [Wil+06]. These findings are in contrast to results from [LR05] where a higher visual fidelity had a significant beneficial influence on desktop VR setups.

Various hardware solutions to the problem of limited physical space have been proposed, such as omnidirectional treadmills [Bou+02; Sou+11]. However, these solutions either only offer sliding feet motions or are prohibitively expensive and thus not likely to be available to a broad public in the foreseeable future [Ste+10]. Another way to alleviate the problem of a limited physical space is scaled walking, meaning that a user's tracked natural walking is scaled in a way. The scale factor is called gain [Wil+06]. For example, Williams et al. evaluated a uniform translation gain of 10 against no scaling and walking versus joystick input [Wil+06]. This scaling means that during when the participants walked 1 m in the physical environment, they walked 10 m in the VE. In their experiments, this gain was applied to all translations. Besides the previously mentioned result that walking resulted in better results, applying the gain also resulted in worse results in their first experiment. In their second experiment, more scalings ($1:1, 2:1, 10:1$) were compared, and the expertise with games was evaluated, resulting in no significant influence of the gain. Interrante et al. devised the *Seven League Boots*, an improved version of scaled walking that also scales the movement of a user, however not uniformly but only in the direction of travel [IRA07]. The direction of travel was defined using a mix of gaze direction when there is little to no movement,

and movement direction, if the user is moving. By only scaling the translation in the direction of travel, the user's head sway while walking was not, or only marginally increased. When compared to other techniques, such as ordinary walking with 180° real turns scaled to 360° at physical boundaries, flying using a wand and uniform scaling, walking gave participants the best sense of distance and gain the worst and furthermore gain was most cybersickness-inducing. Similarly to the Seven League Boots, Steinicke et al. conducted an experiment to find the detection thresholds for a non-uniform gain in the movement direction [Ste+10]. They let people walk straightforward in their second experiment, using gains of 0.6 to 1.4 in 0.1 steps, each with 8 repetitions in a randomized order and after 5 virtual meters, the participants had to answer whether the physical distance they walked was shorter or longer than the 5 virtual meters. They found that the detection thresholds are at 0.86 and 1.26, meaning that users were unable to discriminate within the range of 4.3 m to 6.3 m. This distance underestimation is in line with previous research about wrong distance estimations in VR (cf. Section 2.4.2.1 and Section 2.5).

A second approach overcoming the physical space size limitation of walking in a VE is RDW [Raz05]. The concept of RDW is applying slight rotations of the virtual camera that the user then compensates by turning in the opposite direction. These rotation gains are mainly applied to the yaw axis [Ste+10]. The rotation gain is defined as $g_R := \frac{R_{virtual}}{R_{real}}$, whereas $R_{virtual}$ is the virtual rotation a user should take and R_{real} the real physical rotation a tracked person did [Ste+10]. The gain can be applied by tracking the user's change in rotation around a specific axis between two tie points and applying the new rotation $R_{virtual} = R_{real} * g_R$ to the user representation. The same concept can be applied to translations, wherein translation gains are defined as $g_T := \frac{T_{virtual}}{T_{real}}$ with $T_{virtual}$ being the virtual distance moved and T_{real} being the physical distance a participant moved [Ste+10]. Additionally, Steinicke et al. propose using curvature gains g_C that are a combination of the gains mentioned above. With curvature gains, a user, who tries to stay on a straight path in a virtual world, would be directed to walk along a curve in the physical world. This behavior can be invoked by turning the camera around the yaw axis depending on the distance the user walked. A gain $g_C = 1$ would result in a 180° rotation of the user after walking π m [Ste+10]. Steinicke et al. investigated the detection thresholds of these three types of gains and found that (i) users can be redirected by using rotation gains since participants had difficulties discriminating a real 90° rotation from virtual rotations ranging from 72.6° to 134.3°, (ii) participants could not differentiate real distances between physical 4.3 m to 6.3 m walking in a straight line when walking 5 m in the VE and (iii) when users can walk in a real circle of about 22 m radius, the curvature gain can give them the impression of walking in a straight line indefinitely [Ste+10]. Motion compression techniques, such as [IRA07], can also be considered a redirection technique. Redirection techniques are explained and analyzed in the *taxonomy of redirection techniques* [Sum+12] (cf. Figure 2.23). Their taxonomy is three dimensional between overt or subtle, translation or orientation, and continuous or discrete techniques. Detection thresholds are essential to determine whether techniques are overt or subtle. Langbehn and Steinicke provide an overview of various detection thresholds measurements in research [LS18].

For these techniques, it is necessary to know the user's target to calculate the relevant gains. However, it is difficult to estimate where users might want to travel within VEs. This is why estimations have to be made based on the torso, head, eye gaze directions [Ste+08a; ZK16]. Additionally, within confined tracking spaces redirection techniques can be used to steer the user away from hazards, such as tracking space boundaries [Raz05]. Razzaque proposes three algorithms [Raz05],

Steer-to-center The redirection gains are chosen to direct the user to a specific point (the center) defined within the tracking space.

Steer-to-multiple-targets Multiple target points are defined within the tracking space. One thereof is chosen as a target for the redirection.

Steer-to-orbit A circle is defined within the tracking space which the user is led to. The circle allows the user to walk in a straight virtual line.

Larger tracking spaces make these techniques more straightforward to use, with 6×6 m shown as the minimum viable size [Azm+15]. When constraining the user to curved paths in a VE, a subtle redirection is possible within an even smaller space of 6×6 m, allowing users to walk endlessly on predefined paths [Lan+17].

In the case that users get close to the edge of the tracking area or other hazardous areas, a reset may be necessary [Wil+07]. Williams et al. propose three:

Freeze-and-backup The user's position within the VE is frozen, to ensure it does not move when the user moves. This allows the user to walk backward until they have enough free physical space around them to move, after which the tracking is activated again.

Freeze-and-turn The user's orientation in the VE is frozen, and they can turn around until they have free space in front of them again. Afterward, the tracking resumes.

2:1 turn When the user reaches the tracking area's bounds, they have to turn around with an orientation gain of 2, resulting in the user facing the tracking area again and allowing them to move in the direction they were walking towards virtually.

Peck et al. devised ways to make these resets less overt using distractors [PFW11].

Despite walking being a natural way of traveling through a VE, our informal observations during demonstrations and lab visits show that users often prefer other techniques when both walking and the other technique are available. We assume that this preference can be explained by walking requiring more physical effort, compared to techniques like teleportation. These observations are confirmed by [LaV+17].

One negative aspect of walking through VR is closely related to the limited tracking space because the full occlusion that VR HMDs offer, users may feel detached from the physical environment and may behave differently. For example, we observed users walking more slowly and carefully when immersed in a VE as compared to real walking because they felt afraid of colliding with real objects. In Section 5.3.2 we describe an overt way to use RDW only within the outer ring of a tracking area, whereas the middle has no redirection gains to promote more natural exploring.

Some travel techniques aimed to combine the advantages of these multiple techniques, for example, *Arch-Explore*. In Arch-Explore users were able to walk within a virtual replica of the lab around a miniature of an architectural model and select a room within the miniature to create a portal towards that location at a wall of the virtual lab [BSH09]. Users can walk to the portal without hitting a physical lab wall through motion compression. Within the VE users are redirected, partly to direct them close to physical walls as they approach virtual walls, to allow passive haptic sensation. The jumper metaphor is another combined technique, allowing natural, unmanipulated walking for short distances and teleportation triggered by rapid forward movements, real jumps, for long distances [BSB11].

2.7 Summary

In this chapter, we present the fundamentals of this thesis. In Section 2.1 we establish the terminology, introducing the crucial terms. We explain where we place the research done in the scope of this thesis within Section 2.2. The history of 3DUIs and the essential literature in the field is discussed in Section 2.3. The basics of human factors that are necessary for this thesis are explained in Section 2.4. For example, we explain the basics of biomechanics and fatigue, presence, and cybersickess. Also, the methods to evaluate these are presented in the section. Afterward, in Section 2.5, we describe the hardware utilized in this thesis, divided in input and output devices. Within the output section, we focus on HMDs and their important technical aspects. We use these technical aspects for the research we present in Section 5.4. To avoid duplicates throughout the thesis, we summarize the related work for 3DUIs in Section 2.6. In the first part, we summarize

related work concerning selection and manipulation in 3DUIs. These are important interactions for Chapters 3 and 4. Thereafter, we explain relevant related work for navigation and travel through VEs that is especially important for Chapter 5. We cross-reference this chapter throughout the thesis.

3. Increasing 3DUI Selection Performance

Abstract

According to Bowman's taxonomy of 3DUIs, selection is a fundamental interaction in 3DUIs (cf. Section 2.6) [BH99]. Within this chapter, we evaluate current 3DUIs for semi-immersive tabletops and HMDs in order to identify performance problems. We provide perceptually inspired solutions to improve the performance during selections that will make the 3DUIs more efficient.

This chapter consists of two main sections. In the first one, Section 3.1, we investigate hover interactions in semi-immersive tabletop setups using a perceptual experiment, letting users define areas where they expect hover effects, which they know from 2DUIs, to appear. Afterward, in Section 3.2, we compare this *HoverSpace* to a naive approach in a performance evaluation. The results show that the HoverSpace offers better performance than the naive approach.

In the second part of this chapter, we investigate selection performance in various regions in the interaction volume of a stationary user in fully-immersive HMD environments. The results indicate that most errors are caused by distance misperception and propose using elliptical selection volumes instead of spherical ones.

Thus, we show the importance of including user perception during supernatural 3D interactions to improve their performance and efficiency.

3.1 Hover Selection

This section is based on the works published in [Lub+15]. Recently, the combination of two different technologies has attracted enormous attention. Several setups have been released combining touch-sensitive surfaces with 3D mid-air finger tracking [BSS13b]. These technologies provide direct interaction with 2D or 3D data sets, respectively. Direct interaction is primarily leveraged in the fields of natural interaction for spatial application domains such as geospatial applications, architectural design, games or entertainment [BSS13c]. While multi-touch technology is available for several years, recently multiple hardware solutions from the professional and consumer domains have been released. These provide the means to sense hand and finger poses as well as gestures on 2D surfaces or in 3D space without the requirement to wear gloves or use other encumbering

instrumentation (e.g., Leap Motion [Lea12], Microsoft Kinect [Mic12]). The combination of these technologies and the resulting expanded interaction space consisting of 2D touch input and 3D mid-air sensing provides enormous potential for novel interaction techniques. Until recently, research on interaction techniques in the scope of tabletops and interactive surfaces have mainly been focused on (multi-)touch 2D input with monoscopically displayed data (cf. Section 2.3.1). The direct nature of multi-touch gestures and interaction including haptic feedback has excellent potential for natural and intuitive interaction for novice and expert users. The matching perceptual and motor space during direct touch interaction proved beneficial over less direct interaction techniques [Zil+13]. Spatial interaction above tabletop surfaces has received much attention over the last years, in particular since Hilliges et al. [Hil+09a] discussed the limitation of 2D input on surfaces for natural 3D interaction and proposed interactions above the tabletop.

Since 3D-capable displays are available for the use in interactive tabletops, the interaction space has to be extended to the third dimension to facilitate a coherent space for input and output of such interactive systems. (cf. Section 2.6.1.5). With stereoscopic display, objects can appear detached from the display surface, i.e., in front of or behind the display surface. Such situations induce challenges for natural touch interaction due to missing haptic feedback when interacting with stereoscopically displayed floating objects [BSS13c]. Schöning et al. [Sch+09b] considered challenges of multi-touch interaction with stereoscopically rendered projections. They conclude that most of the existing interaction techniques have in common that the interaction and visualization are limited to a region close to zero parallax (i.e., the interactive surface) [BSS13b; VGH14].

While the described setups provide interesting challenges to the interaction with stereoscopically displayed 3D objects on a touch surface, it is often not evident for users which objects are interactable, i.e., 3D stereoscopically objects often miss the affordance of touch [BSS13b; BSS13c]. In mouse-based interaction setups, such affordances are often presented by hover effects [Bux90] and transferring these concepts to tabletop setups is reasonable. With purely touch-based interaction such hover effects are challenging to implement since few sensors distinguished between touches and proximity. However, new technical developments allow detecting proximity [Wal12]. Hover interaction has been successfully applied for monoscopic displays to support multi-touch tabletops with contextual information [Hil+09a]. However, we are not aware of existing solutions considering hover interaction for stereoscopic, head-tracked multi-touch environments. For a more in-depth description of the related word, regarding tabletops and 3D selection refer to Section 2.6.1 and 2.6.1.5.

In the following, we focus on hover interaction that does not require users to touch an object in 2D or 3D spaces, e.g., by moving a finger inside the object or on the surface. It is instead based on hovering "over" the object with a finger or hand relative to the considered object. Hover interactions are also supernatural since interaction with objects just by hovering over them is not possible in the physical world. However, so far it is not clear how users perceive affordances of hover spaces above the interactive surfaces, especially, if objects are displayed stereoscopically. In particular, it is not clear which shapes and sizes of volumes match the perceived affordances of 3D hover interaction. For these reasons, we determine a perceptually-inspired model for hover interaction volumes, which we call the HoverSpace. We evaluate the model and compare it with a naive approach in a confirmatory study. The results of these experiments provide guidelines for interactive applications using hover gestures in tabletop setups. To summarize, our contributions are:

- An analysis of above-surface volumes for hover interaction in tabletop setups
- A usability comparison of perceptually-inspired and naive hover volumes
- Guidelines for designing hover interaction in touch-sensitive tabletop setups.

3.1.1 Perceptual Experiment

In this section, we describe the experiment in that we analyzed the perceived affordances of hover interaction concerning the 3D volume above rendered objects of different sizes and shapes in an interactive tabletop setup with a stereoscopic display. Based on the previous work described in Section 2.6.1, we explored the following expectations in this experiment: The hover volume where users expect hovering effects to occur may not necessarily be oriented vertically above the object, but rather influenced by a user's head position relative to the target. Moreover, we assumed the hover volume shape to be influenced by the target object shape.

3.1.1.1 Participants

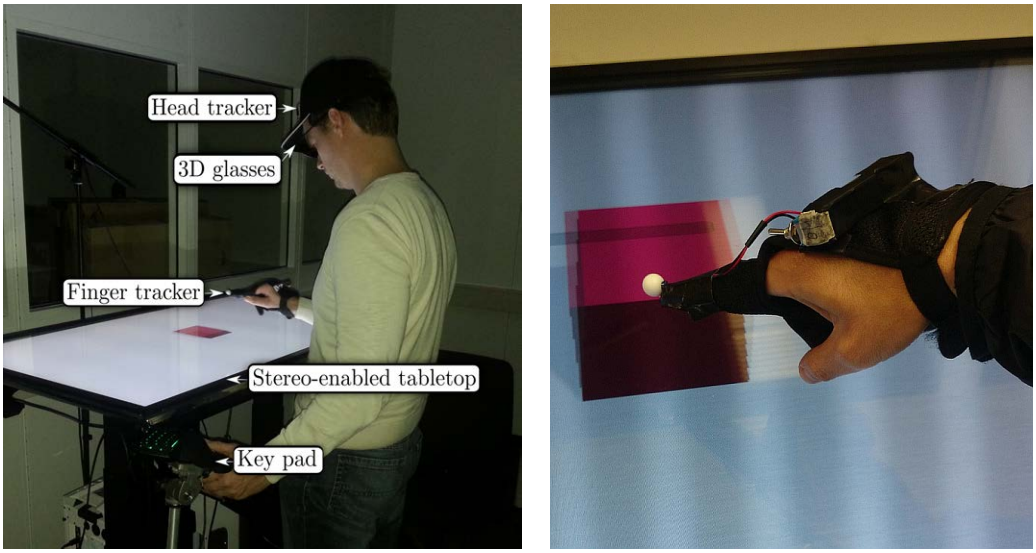
We recruited 15 participants for our experiment (11 male, 4 female), all of whom were students or professionals from the field of human-computer interaction or computer science (ages 24–54, M = 34.7, SD = 8.58, heights 1.55 m to 1.92 m, M = 1.76 m, SD = 0.12 m). The students received class credit for their participation. Two of the participants were left-handed, the remaining 13 participants were right-handed. All participants had normal or corrected vision. Using the technique proposed by Willemsen et al. [Wil+08], we measured the IPD of each participant before the experiment (M = 6.54 cm, SD = 0.35 cm) and used it to calibrate the rendering for each participant (cf. Section 2.4.2.1). Only one participant reported no experience with stereoscopic display and ten participants reported high or very high experience (rating scale 0 = no experience, 4 = very high experience, Median = 3, Counts = (1,2,2,3,7)). Only one participant reported no experience with 3D computer games while ten participants reported high or very high levels of experience (rating scale 0 = no experience, 4 = very high experience, Median = 3, Counts = (1,1,2,5,5)). The mean total time per participant, including questionnaires and instructions, was about 35 minutes. The mean time for performing the actual experiment was about 25 minutes. Participants were allowed to take breaks at any time.

3.1.1.2 Material

As illustrated in Figure 3.1, participants were instructed to stand at a stereoscopic multi-touch table in an upright position facing the table. A Razer Tartarus keypad was adjusted to a comfortable height for the non-dominant hand of the participant. Participants were instructed to keep their hand in that position during the experiment to confirm their selections. The experiment was conducted with the participant wearing Samsung SSG-P51002 radio frequency active shutter glasses, a hat with an infrared (IR) marker and a glove with an IR marker at the fingertip of their dominant hand. The markers were tracked with an optical WorldViz Precision Position Tracking (PPT X4) system with submillimeter precision for view-dependent rendering and finger tracking (cf. Section 2.5). The visual stimulus displayed during the experiment showed a 3D scene that was rendered with the Unity3D Pro engine [Uni13] with an Intel computer with a Core i7 3.4 GHz CPU and an NVIDIA GeForce GTX780TI. The scene was displayed stereoscopically on the iSPACE setup (cf. Sections 2.5 and A.1.1) [Lub+14b]. The scene showed a gray brushed metal surface at the zero parallax plane and targets in a red color. For each trial, a single target was visible, either a sphere or a cube. Those shapes were chosen as they are approximations of objects typically found in user interfaces and compound objects consisting of these shapes could approximate almost any other shape.

3.1.1.3 Methods

We used a $2 \times 2 \times 6 \times 4$ design with the method of constant stimuli for the experiment trials. The two target shapes (cube, sphere), two target sizes (5 cm,10 cm), six target positions (P0 = (-0.2, 0, 0), P1 = (0, 0, 0), P2 = (0.2, 0, 0), P3 = (-0.2, 0, -0.2), P4 = (0, 0, -0.2) and P5 = (0.2, 0, -0.2)) and four repetitions were uniformly and randomly distributed between all 96 trials for each participant. Each trial consisted of a single shape of one size at one of the positions being shown to the participant (cf.



(a) (b)

Figure 3.1: (a) Participant during the experiment and (b) close-up of the participant's hand while indicating the hover volume with the tip of the index finger. The scene was displayed stereoscopically. IR markers on the head and index finger of the dominant hand were tracked. The non-dominant hand rested on a keypad.

Figure 3.2). The participants were instructed to think of hovering in traditional 2D user interfaces and indicate the volume where they would expect a hover effect to be triggered by moving the index finger of their dominant hand in that volume, i.e., by "drawing" the volume. During each trial, the participants indicated the hover volume for ten seconds while pressing a button on the keypad with their non-dominant hand. The non-dominant hand was chosen for this task to avoid the Heisenberg effect while indicating the volume, which may be induced by pressing buttons attached to a glove on the dominant hand (cf. Section 2.4.4). After ten seconds the next trial started. We recorded tracking data at 30 Hz while the participant pressed the button. Each recording consisted of the participant's head and finger position.

The participants completed training trials before the main experimental phase to ensure that they understood the task correctly. Training trials differed from the main trials by showing the participants the volume that they drew to help them understand the task, while this visual feedback was excluded in the main trials. I.e., participants only saw their real hand and the virtual 3D object so as not to bias the results due to cluttering the virtual scene over time. The training trials were excluded from the analysis.

3.1.1.4 Results

We had to exclude three participants from the analysis, as they misunderstood the task and touched the 2D surface throughout the experiment instead of indicating a 3D hover volume. This false behavior was confirmed during the debriefing after the experiment. Since we had four repetitions for each condition of the experiment, we pooled the results over the repetitions. We normalized the tracking data to account for the varying head positions within and between participants by normalizing head positions in target-centered coordinates (cf. Figure 3.3). We visually analyzed the resulting coordinates and observed two main behavior patterns:

Orthogonal Hovering The first behavior pattern was characterized by seven participants indicating the hover-space at the surface of the object or above it. As shown in Figures 3.3(a) and (b), the horizontal width and depth of this volume increased with increasing distance to

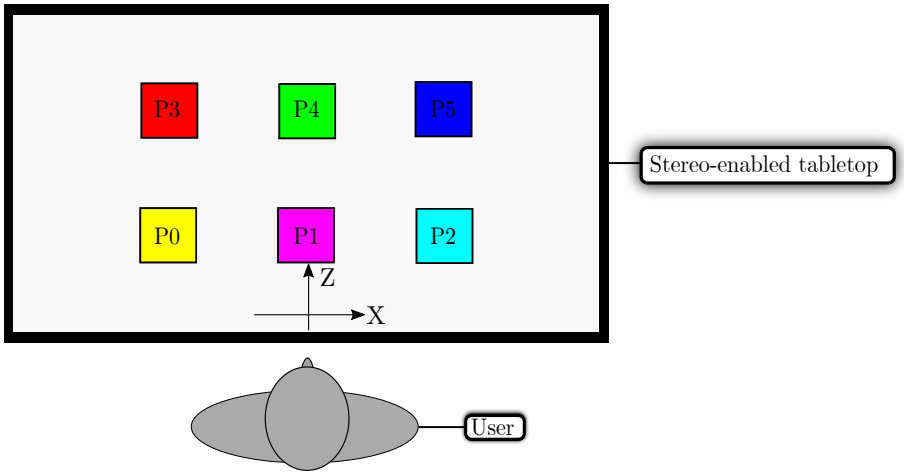


Figure 3.2: Positions used to locate the objects on the tabletop surface. P0 = (-0.2, 0, 0) (yellow), P1 = (0, 0, 0) (magenta), P2 = (0.2, 0, 0) (cyan), P3 = (-0.2, 0, -0.2) (red), P4 = (0, 0, -0.2) (green) and P5 = (0.2, 0, -0.2) (blue). These colors are only used in the following result plots.

the tabletop in the vertical direction.

Line-of-Sight Hovering The second behavior pattern was shown by five participants. Their hover-space was tilted towards line-of-sight, i.e., the volume extended in the direction of the participant's head position instead of orthogonally from the tabletop surface as shown in Figures 3.3(c) and (d). The observed behavior patterns were consistent for each participant throughout the experiment, i.e., we did not observe any participant changing the behavior during the experiment.

We observed a difference between the cube target shape and the sphere target shape. For the sphere, all participants indicated a round hover volume, often drawing circles at various distances from the object (cf. Figure 3.4(a)). Conversely, for the cube shape, all participants indicated a rectangular hover volume, often drawing a rectangular outline and then used a zigzag pattern to fill the area (cf. Figure 3.4(b)). The comments of the participants during debriefing also reflect this behavior.

3.1.1.5 Discussion

We instructed our participants to think of hovering in traditional 2D user interfaces and indicate the 3D-volume with their finger where they would expect a hover effect to be triggered. Considering this instruction, these behaviors could be explained as follows. When hovering in 2D Desktop environments, hover effects are triggered when the mouse cursor is (a) above a target object and (b) occludes the target object. Depending on the participant's understanding of 2D hovering, this has different effects on the behavior with a third, height dimension. Indeed, the results show two groups of participants with distinct behavior patterns (Orthogonal Hovering vs. Line-of-Sight Hovering). We observed no changes in behavior for any participant during the experiment, which suggests that there are two mental models of where users expect a hover volume to be above interactive tabletops. While seven of our participants would expect a hover volume to be located above a virtual object orthogonally to the display surface, five of our participants would expect the hover volume to be located along the line-of-sight from the object to their head position, suggesting that they interpreted hovering as occlusion of the target object. Additionally, we found two different hover volumes for the sphere and cube target shapes. These differences show that the object's shape determines the hover volume shape as well. Round target shapes imply round hover volumes and angular shapes imply angular hover volumes. Based on our data and the results presented in [BSS13b], we suggest a hover height of 10 cm.

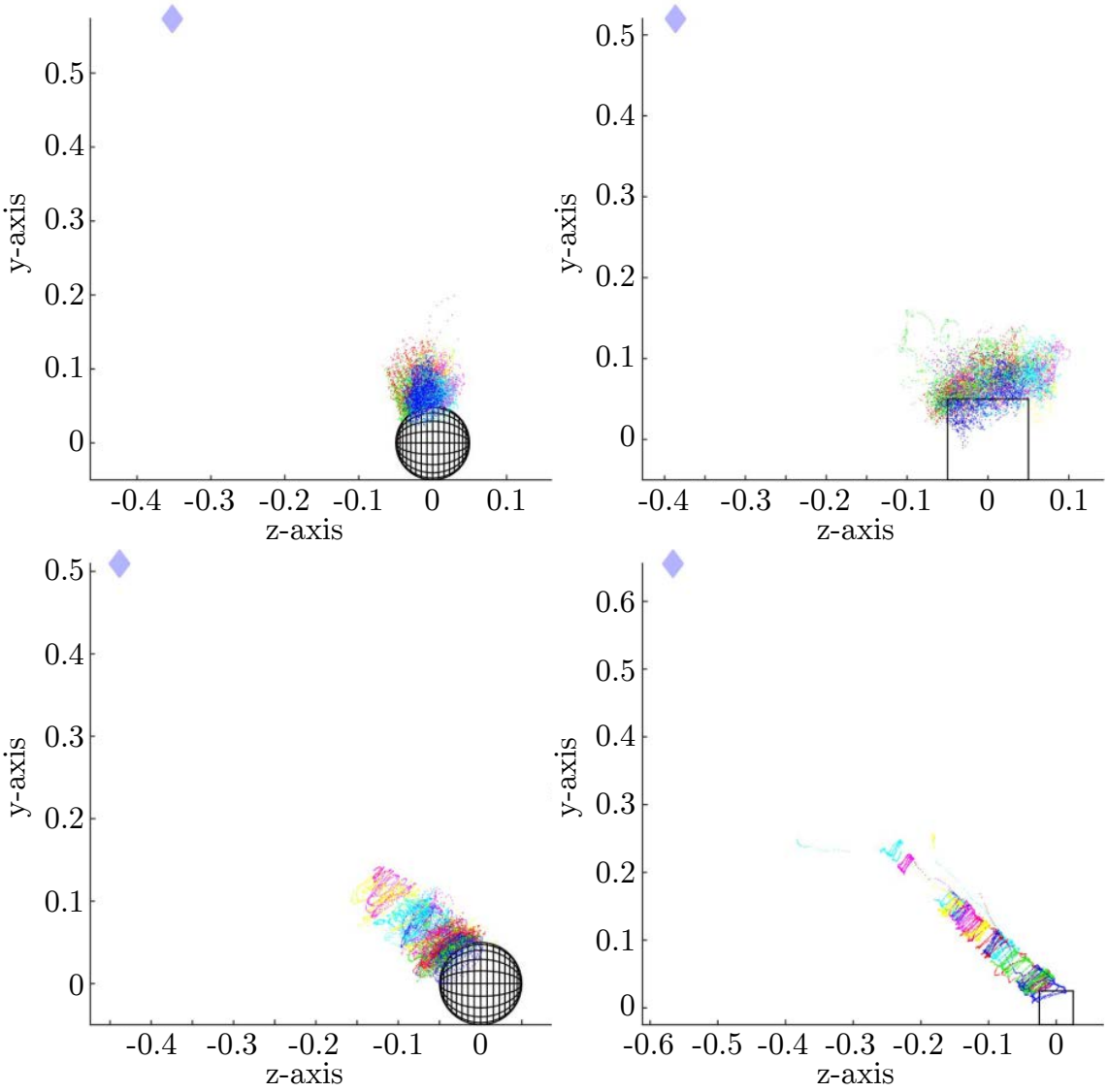


Figure 3.3: Examples of the two behavior patterns in 2D coordinates with the y-axis indicating the up-direction from the tabletop at y = 0 and the z-axis indicating the direction from the participant towards the opposite side of the tabletop. (a, b) Orthogonal Hovering indicates a hover volume at the surface of the object or above it. (c, d) Line-of-Sight Hovering shows a hover volume within line-of-sight that converges towards the participant's head position. The colors represent the different tested positions of the objects on the tabletop (cf. Figure3.2).

3.1.2 The HoverSpace

With the results from the perceptual experiment, we defined two main volumes, where participants expect hovering effects. In the following, we define a combined hover volume that we call *HoverSpace*. Since the volumes depended on whether the target shape was rectangular or rounded, we defined two formulas for the HoverSpace. These formulas allow easy testing whether the tracked input object, such as the user's fingertip, is within the hover volume. The formulas are written for a left-handed Cartesian coordinate system, where the y-axis corresponds to the up direction. Let $(x,y,z) \in \mathbb{R}^3$ be the finger position in 3D coordinates centered around the target object. Let $a \in \mathbb{R}^+$ be the scale of the target object on the x-axis, $b \in \mathbb{R}^+$ the z-axis scale, $c \in \mathbb{R}^+$ the y-axis scale and $d \in \mathbb{R}^+$ an empirically determined value defining the spread of the hover region.

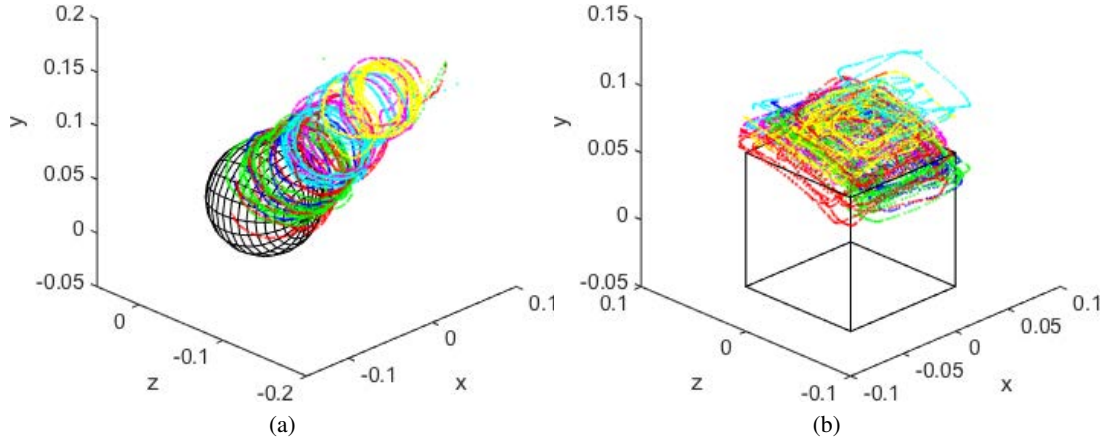


Figure 3.4: Illustrations of the drawn hover volumes of participants for the (a) sphere target shape and the (b) cube target shape. The y-axis indicates the direction orthogonally to the tabletop surface at y = 0, the z-axis increases towards the opposite end of the tabletop, and the x-axis is oriented laterally to the right of the tabletop. The drawing patterns of outlining the circular or rectangular regions and then filling the regions with zigzag patterns was observed for many participants.

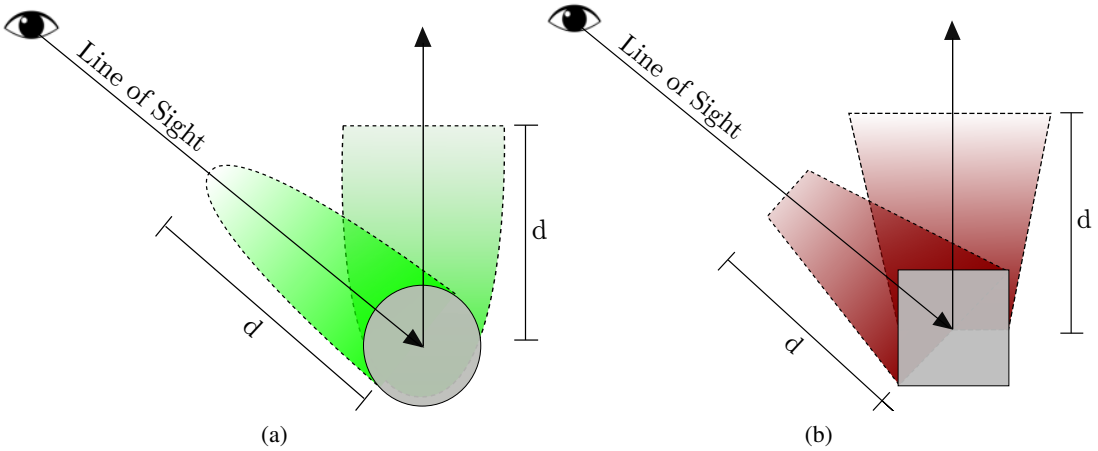


Figure 3.5: Illustrations of the HoverSpace volumes in vertical direction and along line-of-sight for the two shapes: (a) For round shapes the volumes were approximated with paraboloids, and (b) truncated pyramids were used for rectangular shapes.

Given the height of 10 cm, let $0 \leq y \leq 10$cm. The HoverSpace is based on two formulas:

1. For round shapes, a paraboloid can be used that can be approximated by the following formula:

$$\frac{x^2}{a^2} + \frac{z^2}{b^2} - \frac{y}{d} \leq 0 \tag{3.1}$$

2. For rectangular shapes, the results can be approximated by a truncated pyramid.
Let $hf = \frac{y}{c}, x_{max} = lerp(a, d*a, hf), y_{max} = lerp(a, d*b, hf)$
Points are in the HoverSpace if:

$$x \leq x_{max} \wedge y \leq y_{max} \tag{3.2}$$

3.1.2.1 Orthogonal Hovering

From the results of our participants who expected the hover volume to be located above the target object, we derived a volume enclosing 95% of the finger positions. The volume is oriented upwards from the target object, but the volume also expands in width and depth the higher the participant's finger was from the tabletop surface. Depending on whether the target shape was rectangular or round, we found the 95% volume to follow a mathematical function. The region above the object is illustrated in Figure 3.5. For the orthogonal region, the origin of the coordinate system to transform the tracking coordinate of the finger position into for equations 3.1 and 3.2 is given by the center of the object and the up axis along the display normal.

3.1.2.2 Line-of-Sight Hovering

From our participants expecting the hovering effects to occur when occluding the object along their line-of-sight, we found that the width and depth of the volume increase in size as it gets closer to the object until it covers the size of the object. The volume enclosing 95% of finger positions is illustrated in Figure 3.5. In contrast to the hover volume defined in Section 3.1.2.1 the size of this volume depends not only on the size of the object but also on the distance of the head from the object. For rectangular and round shapes we found different volumes. For line-of-sight hovering, the same formulas can be used as for the orthogonal hovering. Here, the origin of the coordinate system is on the line-of-sight between the object center and the head position. The line-of-sight gives the up axis. After transformation of the finger position from tracking coordinates into the line-of-sight coordinate system, the formulas can be applied.

3.1.3 Confirmatory Experiment

In this section, we describe the experiment in which we compared the HoverSpace with a simple straight-up infinitely extruded outline (called *Extruded* in the following) approach of 3D hover volumes.

3.1.3.1 Participants

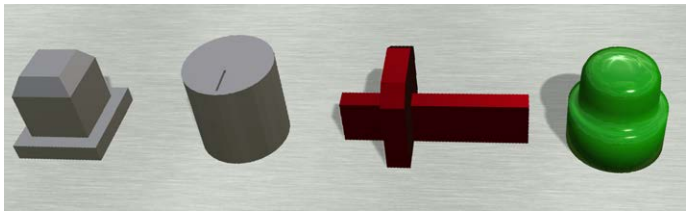
We recruited 16 participants for our experiment (11 male, 5 female), all of whom were students or professionals from the field of human-computer interaction or computer science (ages 19–36, M = 27.37, SD = 4.72, heights 1.60 m to 1.93 m, M = 1.78 m, SD = 0.1 m). Six participants already participated in the first experiment. The students received class credit for their participation. Two participants were left-handed, the remaining 14 participants were right-handed. All participants had normal or corrected vision. We measured the IPD of each participant before the experiment started (M = 6.61 cm, SD = 0.29 cm). We calibrated the system accordingly for each participant.

All participants reported at least some experience with stereoscopic display and twelve participants reported high or very high experience (rating scale 0 = no experience, 4 = very high experience, Median = 3, Counts = (0,3,1,7,5)). Three participants reported no experience with 3D computer games while ten participants reported high or very high levels of experience (rating scale 0 = no experience, 4 = very high experience, Median = 3, Counts = (3,1,2,3,7))).

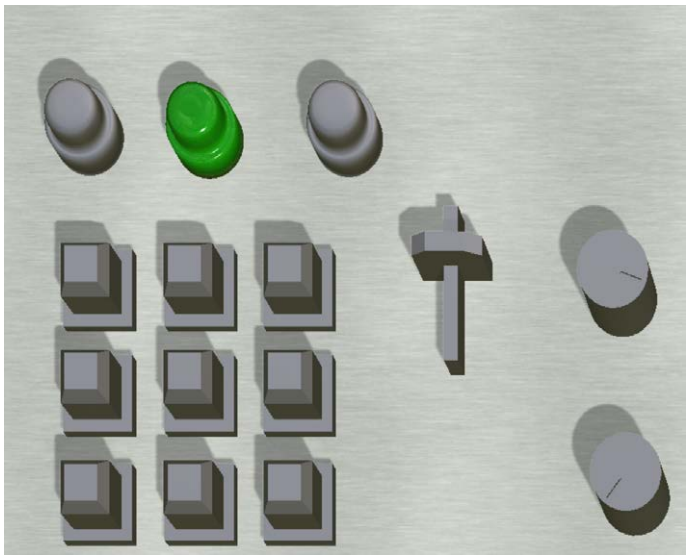
The mean total time per participant, including questionnaires and instructions, was 20 minutes. The mean time for performing the actual experiment was about 15 minutes. Participants were allowed to take breaks between the conditions.

3.1.3.2 Material

The setup in the confirmatory experiment was the same as in the experiment reported in Section 3.1.1. The setups differed only in the visual representation. The scene showed a gray brushed metal surface at the zero parallax plane and targets were shown in a gray color. For each trial, six target objects were visible, either a round button, a round knob, a slider or a rectangular button, as illustrated in Figure 3.6. Those shapes were chosen as they represent objects that we often find



(a)



(b)

Figure 3.6: (a) Illustration of the used target shapes and the colors in the confirmatory experiment. From left to right: rectangular button and round knob in grey, red rectangular slider, green round button. (b) Illustration of an example interface.

in practical applications for tangible user interfaces. When a participant reached with the index finger within the hover volume of an object during the experiment, the object either highlighted red (interpreted as the wrong target) or green (interpreted as the correct target).

3.1.3.3 Methods

We used a $2 \times 4 \times 2 \times 6$ design with the method of constant stimuli for the experiment. We considered two hover volumes: HoverSpace vs. Extruded. The four target shapes (round button or knob, rectangular slider or button), two target sizes (2.5 cm or 5 cm), and six repetitions were uniformly and randomly distributed between all 48 trials in each hover condition for each participant. The hover volume condition was counterbalanced between participants, i.e., half the participants started with the Extruded condition and the other half with the HoverSpace condition. Each trial consisted of six instances of the same shape of one size at different positions being shown to the participant. The positions were arranged in a grid representing the participant's interaction space, such that they were able to reach them comfortably. The sizes of the objects, at 2.5 cm diameter and 5 cm diameter resemble the typical button size for a display of that size, and double the size, to allow a comparison between different sizes, respectively. The grid was spaced so that the objects did not overlap at any time. The participants were instructed to find the green object by hovering over it with the index finger of their dominant hand and press a button on the keypad with their non-dominant hand when they were within the hover volume. During the experiment, the participants saw their real hand and the six virtual objects. In the HoverSpace condition, we used the functions described in Section 3.1.2 and determined whether the participant's finger was

in an object's hover volume. In the Extruded condition, we set the y-coordinate of the finger to the target object's height and used the corresponding equation for a 2D approximation of the object's outline. For the round objects, we used the equation for an ellipse, and for the rectangular condition the equation for a rectangle. These equations effectively create infinite volumes, which means that the Extruded condition extended higher compared to the HoverSpace and its height of 10 cm (cf. Section 3.1.2). The objects were colored grey when the user's finger was outside the hover volume. When the finger was inside an incorrect target's hover volume, the object turned red. When the finger was in the correct target's hover volume, the object turned green as shown in Figure 3.6. After pressing the button, the next trial began. All participants were instructed to complete the task as fast and as precise as possible. The first dependent variable was the selection time, i.e., the time from the start of the trial until the participant confirmed the selection with the press of a button. The second dependent variable was the error rate, i.e., the number of times the participant pressed the button without being within the correct hover volume. The participants completed supervised training trials before the experimental phase to ensure that they understood the task correctly. The trials differed from the actual trials in that they allowed the participants to familiarize themselves with the two different hovering conditions. The training trials were excluded from the analysis.

3.1.3.4 Subjective Questionnaires

To collect subjective impressions, we utilized a comparing AttrakDiff questionnaire, which measures hedonic quality and attractiveness [HBK03]. Following an initial demographic questionnaire, after half the trials, the participants had to take a break, answer an AttrakDiff questionnaire and then continue with the other condition. After the second condition, they filled in the second part of the AttrakDiff questionnaire and a further questionnaire directly asking them to judge which technique they preferred and why they chose that technique.

3.1.3.5 Hypotheses

Based on the results of the perceptually-inspired experiment discussed in Section 3.1.1, we evaluated the following hypotheses:

- $\mathbf{H}_1$ For the HoverSpace condition, the mean selection time is lower than for the Extruded hover volume.
- $\mathbf{H}_2$ For the HoverSpace condition, the mean error rate is lower than for the Extruded hover volume.
- $\mathbf{H}_3$ The participants prefer the HoverSpace over the Extruded hover volume.

3.1.4 Results

In the following section, we summarize the results of the confirmatory experiment. We analyzed the results with a repeated measure ANOVA at the 5% significance level. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated. Since we found no difference between the results for the two round shapes, nor between the two rectangular shapes, we pooled the data.

3.1.4.1 Selection Time

The results for the selection time are shown in Figure 3.7(a). The results show that the selection time differs significantly between the Extruded ($M = 3.09$, $SD = 1.31$) and the HoverSpace ($M = 2.90$, $SD = 1.35$) conditions ($F(1,15) = 7.955$, $p<.05$, $\eta_p^2 = .347$). As expected, we found a significant influence of the target scale on the selection time ($F(1,15) = 16.294$, $p < .001$, $\eta_p^2 = .521$). We did not find a significant influence of the target shape on the selection performance ($F(3,45) = 2.46$, $p = .075$, $\eta_p^2 = .14$). We found a significant interaction effect between the hover condition and the round condition ($F(1,15) = 9.256$, $p < .05$, $\eta_p^2 = .382$). Post-hoc tests with Bonferroni correction for the interaction effect between the hover and the round condition showed significant differences

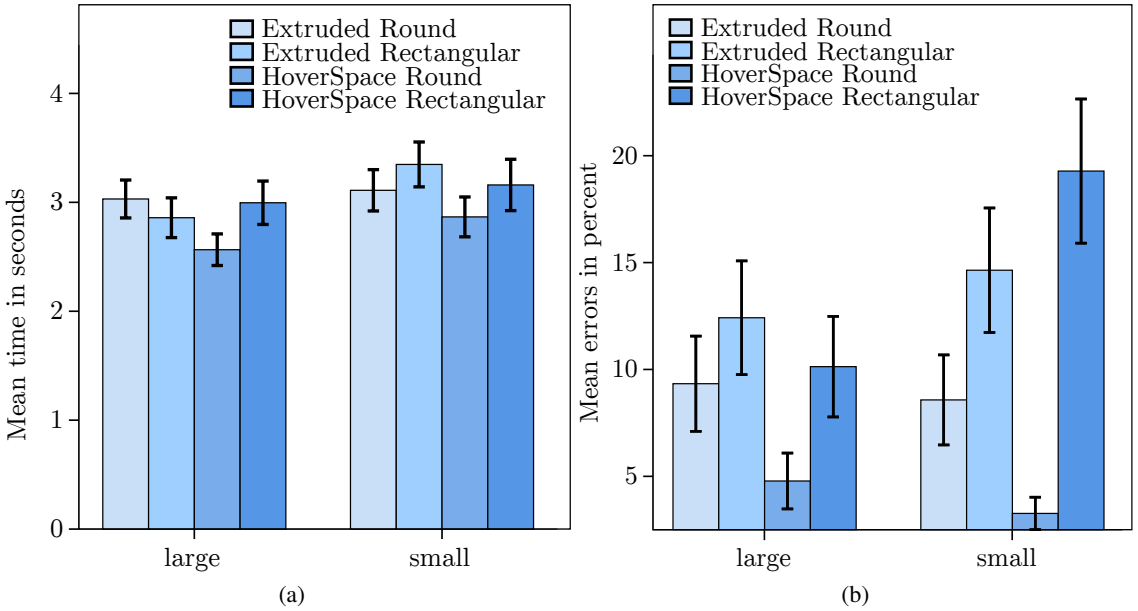


Figure 3.7: (a) Plots of the pooled results of the confirmatory experiment. The x-axes show the target scales and the y-axes show the (a) mean time in seconds and (b) the mean errors in percent. The bar plots are grouped by the hover condition and the round condition. The vertical bars show the standard error.

only between the Extruded-rectangular (M = 3.11, SD = .61) and HoverSpace-round (M = 2.72, SD = .53) conditions (T(15) = 3.164, p < .05), between the Extruded-round (M = 3.07, SD = .62) and HoverSpace-round conditions (T(15) = 4.229, p < .05), and between the HoverSpace-rectangular (M = 3.08, SD = .31) and HoverSpace-round conditions (T(15) = 3.240, p < .05).

3.1.4.2 Errors

The results for the error rate are shown in Figure 3.7(b). The results showed no significant difference in error rate between the Extruded (M = .06, SD = .24) and the HoverSpace (M = .05, SD = .21) conditions. We found a significant influence of the round condition on the errors (F(1,15) = 10.392, p<.05, $\eta_p^2 = .409$).

We found no significant influence of the hover condition (F(1,15) = .775, p = .392, $\eta_p^2 = .049$) or the scale (F(1,15) = 3.629, p = .076, $\eta_p^2 = .195$). We found a significant interaction effect between the scale and the round condition (F(1,15) = 7.304, p < .05, $\eta_p^2 = .327$). Post-hoc tests with Bonferroni correction for the interaction effect between the scale and the round condition showed significant differences only between the big-round (M = .03, SD = .06) and small-rectangular (M = .10, SD = .12) conditions (T(15) = 3.264, p < .05), as well as between the small-rectangular (M = .10, SD = .12) and small-round (M = .02, SD = .04) conditions (T(15) = -3.178, p < .05).

3.1.4.3 Selection Distribution

We evaluated how often participants selected an object in the HoverSpace condition while being with their index finger in the Orthogonal or Line-of-Sight volumes (cf. Section 3.1.2). The distribution was approximately 23% only in the Orthogonal volume, 7% only in the Line-of-Sight volume, and 65% in the overlap region of both volumes. Approximately 5% of all selections were errors. In the Extruded condition, approximately 94% were within the vertically extruded region, and approximately 6% of all selections were errors.

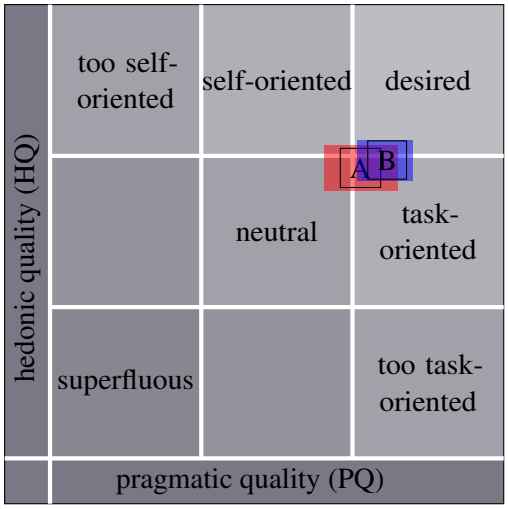


Figure 3.8: Average values and confidence rectangles for the AttrakDiff questionnaire of the two conditions: A for the Extruded approach and B for the HoverSpace.

3.1.4.4 Subjective Questionnaires

The results of the AttrakDiff questionnaire (cf. Figure 3.8) show that pragmatic quality, i.e., an indication of whether the product assists the user, reaches an average value overall. In comparison, pragmatic and hedonic qualities of the HoverSpace are higher than of the Extruded hover volume. The HoverSpace also has a smaller confidence interval for PQ and HQ, indicating a greater level of certainty on the users. Regarding the overall means, the HoverSpace approach is located in the above-average region with an overall impression of the approach as attractive. We asked the participants which of the techniques they preferred, either the first or the second one they tried. As the experiment was counterbalanced, the results were mapped towards the HoverSpace or Extruded condition. The results show a preference of the HoverSpace (rating scale 1 = Extruded, 5 = HoverSpace, Median = 4, Counts = (0,3,2,1,5)).

3.1.5 Discussion

The results of the confirmatory experiment showed that the HoverSpace outperformed the Extruded approach considering the selection time. This implies a higher overall performance, considering that we did not find a significant difference between the similar error rates. This confirms our hypothesis H_1, but not H_2. Considering the limited size of the HoverSpace, albeit wider than the infinite Extruded approach, the lower selection time implies that the perceptually-inspired hover volume is a valuable improvement over vertically extruded hover regions. The results from the subjective questionnaires support our hypothesis H_2, as they show that the participants subjectively preferred the perceptually-inspired HoverSpace over the Extruded approach. We received multiple comments such as “The second technique [HoverSpace] was much more intuitive and more precise compared to the first one.” These comments further support this hypothesis. However, some participants thought the Extruded approach was more precise and preferred it over the HoverSpace. This might be caused by the fact that the HoverSpace volumes for different target objects overlapped when they were located close together, causing two targets to change color at the same time. To disambiguate such multiple selections in future implementations, we suggest prioritizing selections in orthogonal hover volumes compared to selections in Line-of-Sight hover volumes considering the more substantial number of participants in the experiment described in Section 3.1.1 whose mental model matched these hover volumes.

3.1.5.1 Guidelines

Hovering in interactive tabletop environments allows effective decluttering of interfaces. Our analysis of the perceived spatial affordances of such hover interaction has shown that perceptually-inspired hover volumes can increase the performance, as well as the subjective attractiveness of interfaces. In the following, we summarize the lessons learned: We observed two mental models for hovering in our stereoscopic tabletop environment. The first mental model was characterized by users expecting hovering effects to occur when their hand is right above the object on a line along the display normal, and the second is characterized by users occluding the object based on their line-of-sight. The orthogonal hovering is relatively close to the naive, straight-up extruded outline solution usually implemented in related work. We suggest a combination of both of these approaches to provide a technique valid for most users. However, our results in the confirmatory experiment suggest that only seven percent of the selections were in the line-of-sight volume. This leads to the conclusion that the line-of-sight volume should be provided when available, but could be left out when head-tracking is not available, e.g., in tabletop setups with a monoscopic display.

We suggest the following guidelines for hovering in tabletop environments:

- $\mathbf{G}_1$ A combination of an orthogonal region and a line-of-sight region provides the best performance for hovering tasks.
- $\mathbf{G}_2$ Without head-tracking, using an orthogonal region with increasing width depending on the height from the object provides acceptable performance.
- $\mathbf{G}_3$ Approximate the object shape using a matching HoverSpace.

3.1.6 Conclusion

Due to recent technological advances, the combination of touch interaction on interactive surfaces with spatial interaction above the surface of a stereoscopic display has become feasible. In this section, we identified a way to improve the interaction in this holistic design space by conducting a perceptual study and evaluating the results to define a perceptually-inspired hover volume called the HoverSpace. We confirmed the advantages of this HoverSpace in an experiment and found a significant improvement in performance compared to the traditional approach. Finally, we discussed guidelines for the development of future touch-sensitive interfaces. In the perceptual experiment, we identified two mental models that users exhibit for hovering in stereoscopic 3D environments. These are grounded in the different interpretations of 2D hovering as bringing a mouse cursor over an object or occluding an object with a cursor, respectively. Our results show that both interpretations have direct implications for the design of hover interaction in the 3D space above interactive tabletops. Future research may investigate whether there are similar differences in other 3D interaction techniques derived from 2D Desktop interfaces. Additionally, future work may focus on the impact of the different parameters of the HoverSpace and determine the best possible values with the smallest necessary volume to reduce overlapping HoverSpace volumes. In particular, the difference between the round and rectangular target shapes could be investigated further to determine whether a mean between these two shapes could improve the HoverSpace. Furthermore, an investigation of training effects may show how long users need to adapt to different types of hover volumes.

3.2 Direct Selection

After investigating selections in mid-air above tabletops, we took a look at selections in fully-immersive HMD environments. This section is based on the works published in [LBS14a]. One of the unsolved challenges of 3DUIs is the user's wrong estimation of distances (cf. Section 2.4.2.1). Due to missing haptic feedback, users can touch through virtual objects. Thereby, efficient direct interaction with virtual objects is hindered. Since direct interaction is a crucial part of NUIs, in this section, we investigate the source of those wrong distance estimations in fully immersive 3DUIs.

Utilizing a Fitts' Law experiment, we analyzed the 3D *direct selection* of objects in the virtual 3D space as they might occur for 3D menus or floating objects in space. We examined the direct interaction space in front of the user and divided it into a set of interaction regions for which we compared different levels of selection difficulty. Our results indicate that selection errors are highest along the viewing axis, less along the motion axis and marginal along the orthogonal plane. Based on these results we suggest some guidelines for the design of direct selection techniques in IVEs.

IVEs have the potential to offer natural and direct interaction with objects displayed in the virtual world (cf. Section 1.1). In particular, the workspace within the user's arm's reach provides a region, in which the user can grab virtual objects similar to grabbing in the real world. However, some of the often observed drawbacks of VR technologies, for instance, distance underestimation, complicate the design of 3DUIs in IVEs [Ste+09]. Although, we have recently observed significant improvements in 3D input devices and motion tracking systems, using tracked human gestures and postures in *mid-air* introduces challenges to the design of high-performance interaction techniques in the 3D space [Cha+10; Val+11; Val+10b] (cf. Section 2.5).

Interacting with natural gestures in 3D space opens up new possibilities for exploiting the richness and expressiveness of the interaction (cf. Section 2.6). Users can control multiple DOF simultaneously, and exploit well-known real-world actions. However, interaction in the 3D mid-air is physically demanding and, therefore, often hinders user satisfaction and performance [Cha+10]. The increase in the DOFs that have to be controlled simultaneously as well as the absence of passive haptic feedback and resulting interpenetration and occlusion issues when *touching the void* [BSS13a; BSS13c; Cha+10] are often responsible for reduced performance. 3D object selection is one of the fundamental tasks in 3DUIs. It is usually the initial task followed by more complex user interactions such as 3D manipulations [Bow99; SRH04].

In this context, *virtual hand techniques* are often considered to be the most natural way of directly selecting virtual objects as they map identically virtual tasks with real tasks, which stands in contrast to indirect selection [Bow99; LaV+17]. However, direct selection of a virtual object in a fully-immersive setup significantly differs from selecting a physical object in the real world [AA13]. For instance, users perceive the VE stereoscopically with vergence-accommodation conflicts, and usually also cannot see their real body, but at most a virtual representation in the form of a virtual hand, marker or 3D point in space.

Furthermore, even small imprecisions, inaccuracies, and latency of the used tracking system may cause slight mismatches between visual appearance of the virtual hand and the user's proprioceptive and kinesthetic feedback [BSS13a; BSS13c; Cha+10]. Such a decoupling of motor and visual space during natural hand interaction may degrade performance due to the kinematics of point and grasp gestures in 3D space and the underlying cognitive functions [Mac+87; WM99]. An increased vergence-accommodation conflict has been found to effect size and distance judgments, as well as judgments of interrelations between displayed virtual objects [BSS13c; LK03]. So far, while the perceptual problems in fully immersive VR are known, it is assumed that the main error for selection tasks in HMD environments is in the movement direction [BSS13c]. We hypothesize that the main reason for errors lies in the distance estimation and is thus along the viewing axis.

Previous research focused on the selection of objects at arbitrary positions in the virtual space around the user, and different guidelines have been proposed [DKM99; Int00]. However, due to

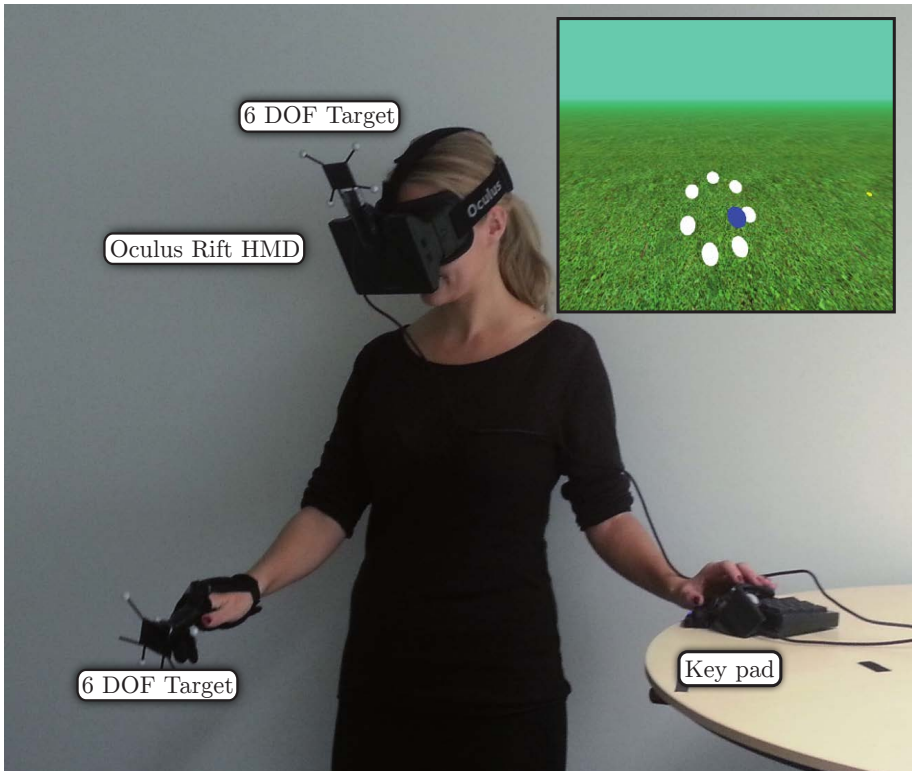


Figure 3.9: User during the experiment. The inset shows the user's view to the VE with the white and blue target spheres as well as the yellow sphere that visually represents the virtual finger tip.

the perceptual difficulties in HMD environments, it is necessary to revisit if the known guidelines from real workspaces and traditional interfaces apply for 3DUIs. One of the first steps towards that goal is to investigate whether there are any differences between the regions in the user's interaction space or not. The related work concerning 3D selection is discussed in Section 2.6.1.

In this section, we compare direct 3D selection task performance in an HMD environment using a Fitts' Law experiment. We tested different regions as well as different levels of difficulties, using the metrics movement time, error rates, error distances, and resulting effective throughput as an overall performance indicator. The results give implications for the design of 3DUIs for direct selection of objects displayed in mid-air space. In summary, our contributions include the

- Comparison of direct selection performance in different regions in the space around the user, and
- Guidelines for designing 3DUIs for direct selection in the 3D space.

3.2.1 Experiment

In this section, we describe the Fitts' Law experiment in which we analyzed direct selection in the user's arm's reach in an immersive HMD environment. Based on previous findings discussed in Section 2.4.2.1 and Section 2.6.1, we evaluate the following two hypotheses:

H$_1$ Larger selection errors occur along the view direction than along the movement direction.

H$_2$ Fewer selection errors and a higher effective throughput are present for selection targets located in lower interaction regions.

3.2.1.1 Participants

We recruited 27 subjects for our experiment. Nine of them were male and 18 female (ages 19–25, M = 21.78, heights 1.58 m to 1.86 m, M = 1.72 m). All of the subjects were students of human-

computer interaction or media communication from the University of Würzburg, who received class credit points for the participation in the experiment. One of the subjects was left-handed, the remaining 26 subjects were right-handed. All but one of the subjects had normal or corrected vision.

Using the technique proposed by Willemsen et al. [Wil+08] (cf. Section 2.4.2.1) we measured the IPD of each subject before the experiment started (M = 6.26 cm, SD = 0.31 cm). Additionally, we used the Porta and Dolman tests to determine the dominant sighting eye of subjects [MOB03] (cf. SectionSec:fund:perception:vision). The test revealed that seven subjects were left-eye dominant (3 male, 4 female), whereas 20 subjects were right-eye dominant (5 male, 14 female). We measured the subject's arm length, and celiac plexus height and calibrated the VE for each subject accordingly.

Only three subjects reported no experience with stereoscopic 3D, such as cinemas or TV. All other subjects reported at least some experience (rating scale 0 = yes, 4 = no, Median = 3, Counts = (2,2,5,9,7)). Five subjects reported that they had no experience playing video games. All other subject reported at least some experience (rating scale 0 = no experience, 4 = a lot of experience, Median = 3, Counts = (5,6,2,7,5)). The remaining seven subjects had no experience with HMDs. All subjects were naive to the experimental conditions. Nine subjects reported that they have participated in HMD studies before, and eleven subjects reported experience with HMDs (rating scale 0 = no experience, 4 = a lot of experience, Median = 0, Counts = (14,3,2,5,1)). The remaining seven subjects had no experience with HMDs. All subjects were naive to the experimental conditions. The mean total time per subject, including questionnaires and instructions, was about 55 minutes. The mean time for performing the actual experiment with the HMD worn was about 36 minutes. Subjects were allowed to take breaks at any time.

3.2.1.2 Materials

As illustrated in Figure 3.9, subjects were instructed to stand in an upright position facing away from a wall. A Razer Nostromo keypad was adjusted to a comfortable height for the non-dominant hand of the user. Subjects were instructed to keep their hand in that position during the experiment to confirm their selections.

The experiment was conducted with the user wearing an Oculus Rift (Developer Edition) HMD with an attached 6 DOF passive infrared (IR) target. An Optitrack V120:Trio tracked the target, a passive, factory-calibrated IR tracking system, allowing to track the position and orientation of the HMD and thus the subject's head. The Oculus Rift offers a horizontal FOV of approximately 90° and a vertical FOV of 110° at a resolution of 1280 × 800 pixels (640 × 800 for each eye). The hardware is described in more detail in Section 2.5.

Additionally, we attached another 6 DOF target to the index finger of the subject's dominant hand (see Figure 3.9). The tracking system uses three cameras with a resolution of 640×480 pixels, a FOV of 47° and a latency of 8.33 ms. The virtual stimulus (see Figure 3.9(inset)) used in the experiment displayed a 3D scene, which was rendered with OpenGL on an Intel computer with a Core i7 3.8 GHz CPU, 8GB of main memory and Nvidia GeForce GTX580 graphics card. For the Oculus Rift, we rendered the virtual scene side-by-side and applied a barrel distortion afterward.

Spheres represented the targets in the experiment. All but the current target sphere for one trial were always visible and colored white. The current target was colored blue when the marker was outside, and green when the marker was inside to give the subjects a visual clue. The spheres were lighted in the order specified by the ISO 9241-9 standard [Int00]. As illustrated in Figure 3.10 each trial consisted of an arrangement of 8 spheres, 7 target spheres which were forming a circle with each sphere at the same distance to the 8th start sphere which was located at a comfortable position in front of the subject's celiac plexus [Vet+11].

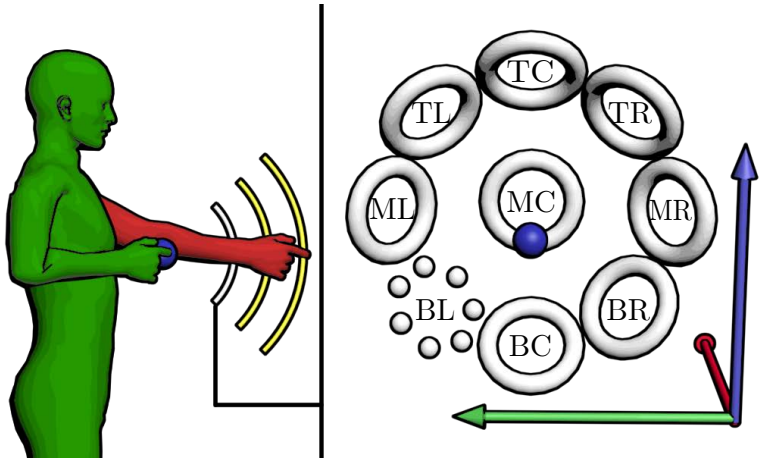


Figure 3.10: Schematic illustration depicting the layout of the spheres. The blue sphere depicts the starting position for all the trials. The circle segments show the three evaluated distances and the red arm depicts the subject's arm reaching for the furthest distance while the green person is in the starting position. The right side shows all the positions of the sphere rings on one depth level. The bottom-left (BL) ring shows an example configuration of seven target spheres during one trial. Each of the nine rings represents a setup of seven spheres used in another trial. Each trial consisted of the seven target spheres of one ring and the blue starting position.

3.2.1.3 Methods

At first, subjects completed between 2 and 6 supervised training trials for the experimental phase to ensure that they understood the task correctly. Those training trials were excluded from the analysis.

We used a $2 \times 9 \times 3 \times 2$ design with the method of constant stimuli for the experiment trials. The two sphere radii, three distances, and nine ring positions, as well as the two repetitions, were uniformly and randomly distributed between all 108 trials for each participant.

Each trial consisted of sequential selections of all 7 targets in a full ring with recurring direct selections of the 8th target in the center position, resulting in a total of 16 selections per trial and 8 visible target spheres. Using both oral and written instructions, the subjects were guided to select the targets as quickly and accurately as possible, as it is common in most Fitts' Law experiments [DKK07; Mac+87; MI08].

While selecting a target correctly, subjects received visual feedback by changing the color of that target sphere from blue to green. Like previous studies on mid-air-selection [BSS13a; BSS13c], the subjects had to confirm each selection by pressing a button on the keypad with their non-dominant hand to avoid any jitter caused by pressing buttons with their dominant hand. Alternatively, a pen might be used to point-and-click with the dominant hand, but Li et al. have shown that using the non-dominant hand might even improve performance for pen-based interfaces [Li+05].

We computed the distance of the position of the index finger to the sphere center, which indicated a selection if the distance was less than the radius of the target sphere. If subjects performed a selection while the target sphere was not highlighted green, we recorded this as a selection error and advanced the trial state. The dependent variables were movement time, error distance (deviation from target center), error rate (percentage of targets missed), and effective throughput (see Section 2.6.1.1).

With the calibrated celiac plexus height and the target's position, we defined a starting position of 35 cm in front of the user's celiac plexus. Furthermore, we defined 9 *interaction regions* (rings) around an axis straight away from the user, as illustrated in Figure 3.10. Those regions were chosen

since we wanted to evaluate whether moving in a certain direction (despite the same ID) had any influence on the performance. The regions cover the user's interaction space. To ensure equal distances between the starting position, represented in the schematic by a blue sphere, we rotated the ring of spheres by 45 degrees from the center axis.

Each of the rings in Figure 3.10 represents a ring of spheres like the one denoted as BL (bottom left). In this Figure, only the bottom left ring is shown as a ring of spheres to ensure better visibility. The visible sphere configuration in Figure 3.10 corresponds to the configuration visible in the inset of Figure 3.9. Each trial consisted of 8 spheres, the start sphere in the center (blue) and the 7 target spheres. In the following, we denote these interaction regions as top-left (TL), top-center (TC), top-right (TR), middle-left (ML), middle-center (MC), middle-right (MR), bottom-left (BL), bottom-center (BC), and bottom-right (BR). Each sphere on different distance levels has the same distance to the starting position. Using the measured length L from forearm to index finger, we calculated the distances of the spheres in the ring to the start position as follows: $D_1 = L \cdot 0.5$, $D_2 = L \cdot 0.65$, and $D_3 = L \cdot 0.8$.

To be able to compare the different target distances, we scaled the spheres to the two IDs 3.0 and 4.0. To achieve this, we computed two sphere sizes W for each distance D by rearranging Fitts' formula for the ID: $W = D/(2^{ID} - 1)$.

According to Fitts' Law, adapting the target size with respect to the distance between selections results in larger targets for longer selection distances, whereas the targets are smaller for shorter distances, thus resulting in the same task difficulty between the different interaction regions (cf. Section2.6.1.1).

Questionnaires

Additionally to the main experiment trials, we asked subjects to complete subjective questionnaires. Before and after the experiment subjects were asked to complete an SSQ [Ken+93]. After the experimental phase, subjects were asked to complete a SUS-PQ [Uso+99b] (cf. Section 2.4.2.4).

3.2.2 Results

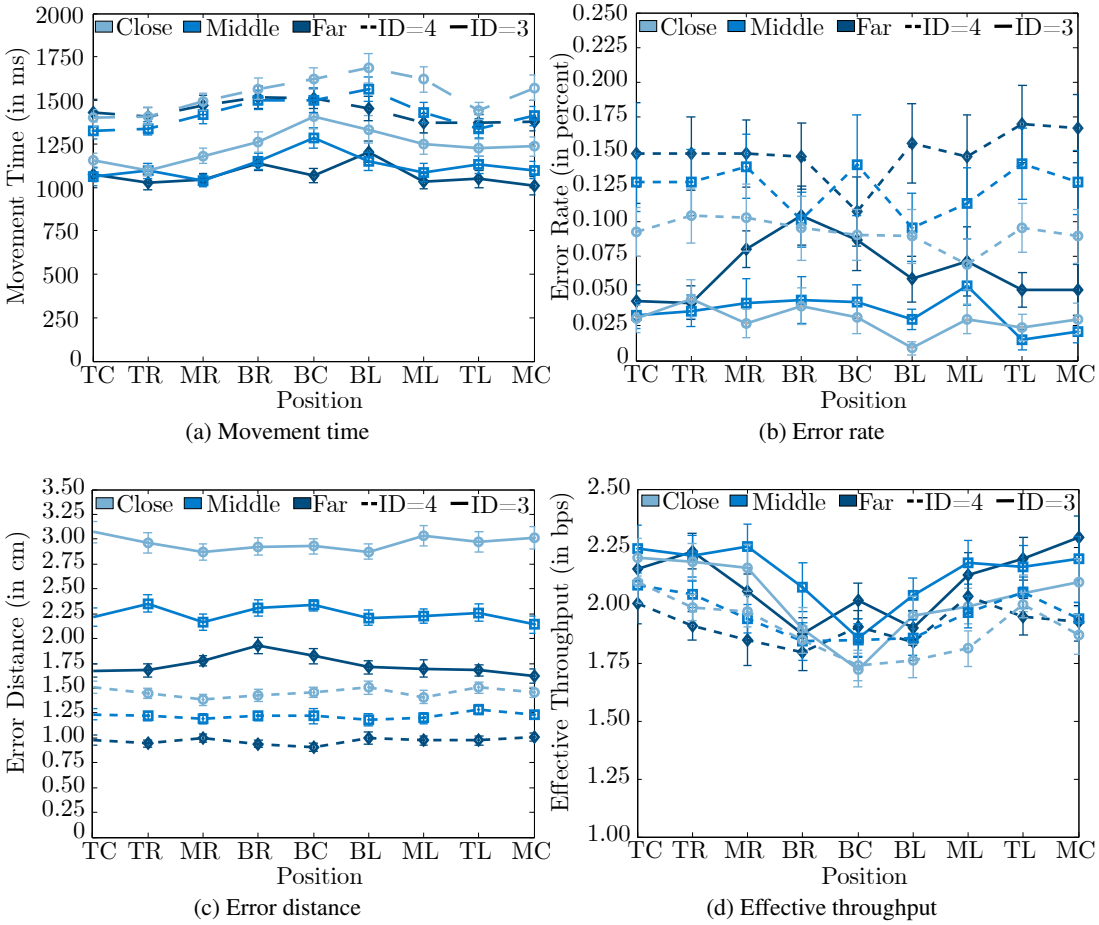
In the following section, we summarize the results of the experiment. Due to severe cybersickness, two subjects had to be excluded. A third subject did not have corrected vision and was excluded, as well. Results from the remaining subjects were normally distributed according to a Shapiro-Wilk test at the 5% level. We analyzed the results with a repeated measure ANOVA and Tukey multiple comparisons at the 5% significance level. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated.

3.2.2.1 Selection Performance

In this section, we describe the analysis of the Fitts' Law selection performance of our experiment. In particular, we analyze the movement time between subsequent selections, the rate of missed targets, error distances to target centers, and effective throughput. The effective throughput metric incorporates both the results for speed and accuracy to form an estimate of the bottom-line performance.

Movement Time

The results for the movement time are shown in Figure 3.11(a). We found a significant main effect of position ($F(4.37, 100.61) = 9.48, p < .001, \eta_p^2 = .29$) on movement time. Post-hoc tests revealed that the movement time was significantly increased when objects were displayed at positions bottom positions BL, BC, and BR (see Figure 3.10) in comparison to the other positions TL, TC, TR, ML, MC, and MR ($p < .001$). Furthermore, the movement time for the three distances differs significantly ($F(2,46) = 46.01, p < .001, \eta_p^2 = .67$). As expected, we found a significant effect of the



(a) Movement time

(b) Error rate

(c) Error distance

(d) Effective throughput

Figure 3.11: Plots of the experiment's results. The x-axis show the positions and the y-axis the a) movement time in ms, b) error rate in percent, c) euclidean error distance in cm and d) the effective throughput in bits per second (bps). The black lines show the closest distance between starting position and the target spheres. The dark gray lines stand for the middle distance and the light gray lines for the longest distance. Solid lines represent the low ID and dotted lines the high ID.

ID on movement time ($F(1,23) = 294.06, p < .001, \eta_p^2 = .93$). We found a significant interaction effect between position and distance ($F(7.04,161.91) = 3.63, p = .001, \eta_p^2 = .14$). The average movement time during the experiment was M = 1313 ms (SD = 371 ms).

Error Rate

The results for error rate are shown in Figure 3.11(b). The results show that the error rate for the three distances differs significantly ($F(1.59,36.63) = 20.56, p < .001, \eta_p^2 = .47$). We found a significant effect of the ID on error rate ($F(1,23) = 43.14, p < .001, \eta_p^2 = .65$). However, we found no significant main effect of the position ($F(8,184) = .86, p = .552, \eta_p^2 = .04$) on the error rate. The average error rate during the experiment was M = 8.8% (SD = 11.3%).

Error Distance

The results for error distances between the center of each sphere and the finger position during selections are shown in Figure 3.11(c). The results show that the error distance for the three distances differs significantly ($F(1.34,30.82) = 139.55, p < .001, \eta_p^2 = .86$). We found no significant main effect of position ($F(1.56,35.88) = .75, p = .447, \eta_p^2 = .03$) on error distance. As expected, we

found a significant effect of the ID on error distances ($F(1,23) = 204.97, p < .001, \eta_p^2 = .90$). The average error distance during the experiment was M = 1.9 cm (SD = 1.3 cm).

Effective Throughput

The results for the effective throughput are shown in Figure 3.11(d). We found a significant main effect of position ($F(8,184) = 12.07, p < .001, \eta_p^2 = .34$) on throughput. Post-hoc tests revealed a significant decrease of the effective throughput when objects were displayed at bottom positions BL, BC, and BR (see Figure 3.10) in comparison to the other positions TL, TC, TR, ML, MC, and MR ($p < .001$). The results show that the throughput for the three distances differs significantly ($F(1.46,33.56) = 9.61, p < .001, \eta_p^2 = .30$). Furthermore, we found a significant interaction effect between position and distance ($F(9.48,218.11) = 2.26, p < .02, \eta_p^2 = .09$). Matching the expectations from Fitts' Law, we found a significant effect of the ID on throughput ($F(1,23) = 35.97, p < .001, \eta_p^2 = .61$). The average throughput during the experiment was M = 1.98bps (SD = .44bps).

3.2.2.2 Error Distribution

In this section, we analyze the distribution of selection errors along

- the finger's movement vector between selections,
- the viewing axis from the dominant eye to the target, and
- the viewing axis from the non-dominant eye to the target.

We could not find any significant main effect for errors in the three directions ($F(1.03,23.76) = .61, p = .449, \eta_p^2 = .03$). However, we found a significant interaction effect between the three distances and the three considered directions ($F(2.04,46.96) = .86.90, p < .001, \eta_p^2 = .79$). Post-hoc tests revealed that for the shortest distance errors significantly differed between movement direction and dominant eye ($t(23) = -5.26, p < .001$), as well as between movement direction and non-dominant eye ($t(23) = -5.52, p < .001$), but not between dominant and non-dominant eye ($t(23) = .80, p = .432$). However, we could not find any significant differences for the larger distances. The error distributions are shown in Figure 3.12.

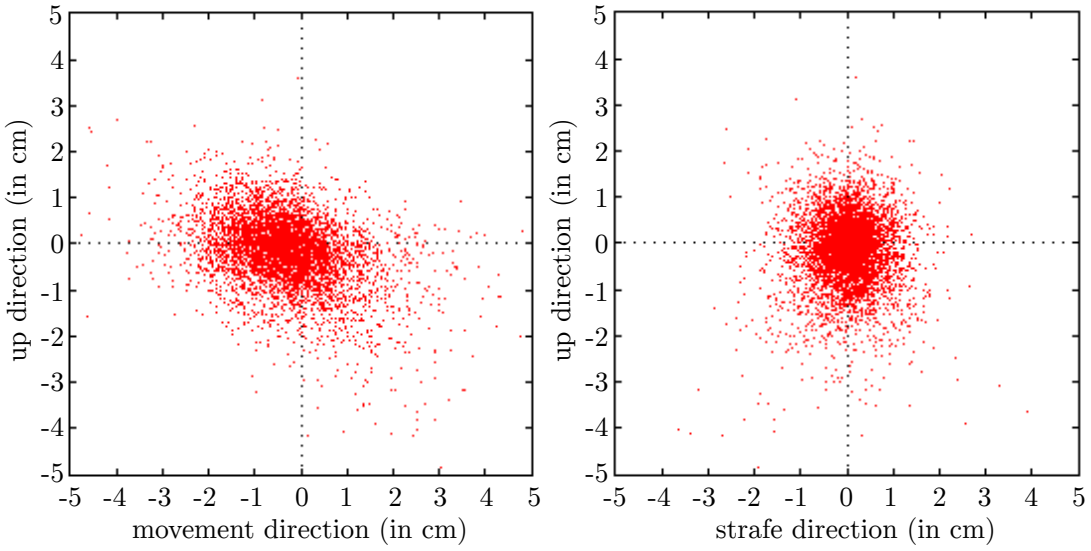
The results for the short distances show that errors were distributed mainly along the view direction and less along the movement direction. The lack of significant difference for larger distances can be explained by the fact that for these targets the view directions and movement directions converged.

Movement Direction

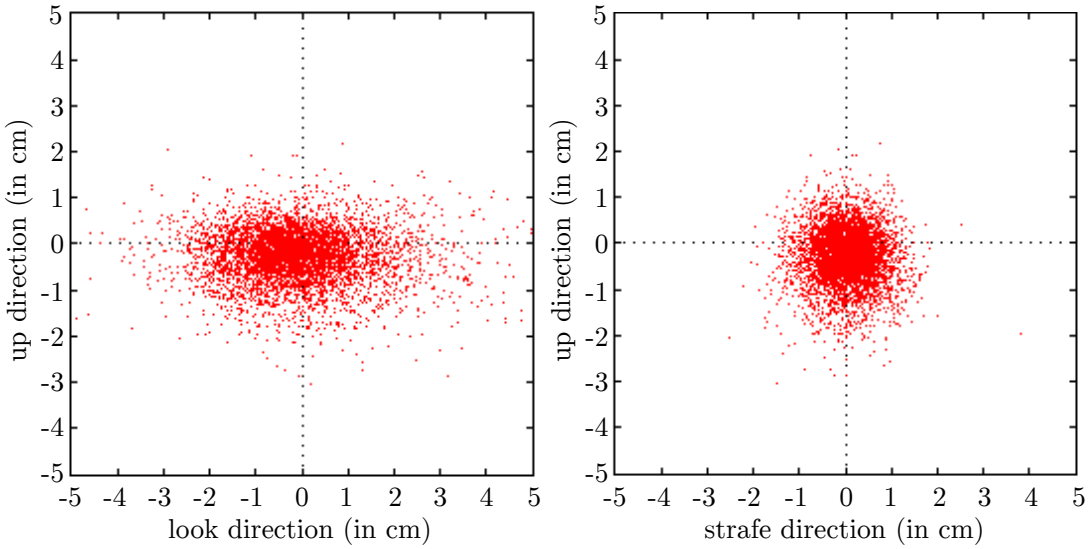
We found a significant difference between errors in movement direction and along the orthogonal axes ($t(23) = 2.81, p < .02$). The error distribution along the finger's movement vector for each selection is shown in Figure 3.12(a). The average error in movement direction was M = 1.21 cm (SD = 0.69 cm) and the average error along the orthogonal axes to the movement direction was M = 1.09 cm (SD = 0.40 cm). For the shortest distance, the average error in movement direction was M = 0.83 cm (SD = 0.13 cm) and along the orthogonal axes M = 0.93 cm (SD = 0.14 cm).

Dominant Eye

The error distribution along the vector from the subject's dominant eye to the target position for each selection is shown in Figure 3.12(b). We found a significant difference between errors in the dominant eye's view direction and along the orthogonal axes ($t(23) = 2.59, p < .02$). The average error in the dominant eye's view direction was M = 1.20 cm (SD = 0.60 cm). The average error along the orthogonal axes to the dominant eye's view direction was M = 1.10 cm (SD = 0.50 cm). For the shortest distance, the average error in the dominant eye's view direction was M = 0.93 cm (SD = 0.15 cm) and along the orthogonal axes M = 0.83 cm (SD = 0.14 cm).



(a) Movement direction



(b) Dominant eye direction

Figure 3.12: Scatter plots showing the error distribution of selection points for the closest tested distance in the experiment over all three IDs: Errors in (a) movement direction from the starting position to the target position as well as the axes orthogonal to the movement direction, and (b) the viewing direction from the dominant eye towards the target as well as the axes orthogonal to the viewing direction during the experiment.

Non-dominant Eye

The error distribution along the vector from the subject's non-dominant eye to the target position for each selection is shown in Figure 3.12(c). We found a significant difference between errors in the non-dominant eye's view direction and along the orthogonal axes ($t(23) = 2.53, p < .02$). The average error in the non-dominant eye's view direction was M = 1.19 cm (SD = 0.59 cm). The average error along the orthogonal axes to the non-dominant eye's view direction was M = 1.11 cm (SD = 0.51 cm). For the shortest distance, the average error in the non-dominant eye's view direction was M = 0.92 cm (SD = 0.14 cm) and along the orthogonal axes M = 0.83 cm (SD = 0.13 cm).

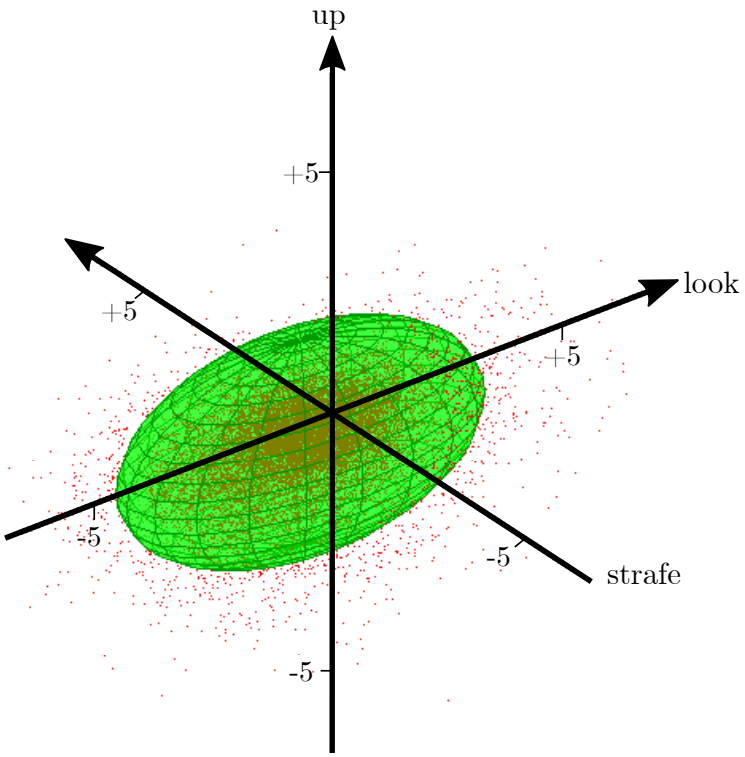


Figure 3.13: 95% confidence ellipsoid encompassing the selection points around the target center during the experiment.

3.2.2.3 Questionnaires

Before and after the experiment, we asked subjects to judge their level of cybersickness as well as their subjective sense of presence. While we measured an average pre-SSQ score of M = 10.29 (SD = 9.80), the post-SSQ score was M = 34.15 (SD = 21.99). The increase in cybersickness over the time of the experiment was significant ($t(23) = -5.83, p < .001$). The mean SUS-PQ score for the reported sense of feeling present in the virtual scene was M = 3.33 (SD = 1.52), which indicates a reasonably high level of presence.

3.2.3 Discussion

Fitts' Law was developed with a focus on two-dimensional interfaces, offering valuable and precise estimations on the potential time necessary to fulfill a task and hence the performance [Fit54]. Afterward, MacKenzie et al. have shown that Fitts' Law also applies to 3D spaces [Mac+87], and Teather and Stuerzlinger expanded the application of the law to VEs [TS11]. All of those works focused on the errors along the movement direction since the very definition of Fitts' Law includes the width along the axis of motion. However, our results imply that in HMD environments the errors along the viewing axis are dominant. Our results suggest that the errors are larger in perception space than in motor space.

These results support our hypothesis H_1, i.e., the movement direction is less a cause of errors than the view direction, which is in line with results found for touch interaction in stereoscopic tabletop environments [BSS13a; BSS13c]. An increase of the interaction distance causes a convergence of the motion axis and the viewing axis in touch selection tasks. This could explain why at increased distances we could not find a significant difference. However, the errors do not disappear but instead accumulate at a distance, which might be the reason for the misconception that the errors are mainly due to the movement direction.

While analyzing the distribution of errors along the viewing direction, we calculated a 95% confidence ellipsoid [SF87], which is visualized in Figure 3.13. On the axes strafe, up, and view the center is located at position $(0.16, -0.47, -0.59)$ with radii $(1.83, 2.11, 3.91)$. Considering the errors were made mostly along the viewing axis, further investigations might show that increased tolerance along the viewing axis would allow faster and more precise selections in HMD environments.

Moreover, we found differences regarding the performance depending on the location of the interaction spaces as illustrated in Figure 3.11(d). Contrary to our hypothesis H_2, our results show that, what we thought were the more comfortable lower regions offer significantly reduced performance compared to the top positions, despite the same distances to the starting position and the same index of difficulty. These results match guidelines from occupational ergonomics, which suggest placing the table height, and thus the interaction height, closer to the user's head when doing precision work [GK97].

Since all constellations had the same two indexes of difficulty in motor space, this behavior could be caused by the increased distance along the viewing axis and thus the increased difficulty in the perceptual space as discussed above. Despite the increased radii of the spheres at a distance, which was done to ensure that the ID remained the same at different distances, the low resolution of the Oculus Rift HMD may also be considered a contributing factor to the effect for more distant targets due to the reduced visual acuity. Future work may focus on evaluating higher resolution displays, which will help to improve the understanding of the reasons for the observed effects. Furthermore, future work should consider if microlens array HMDs, which address the accommodation-convergence conflict [LL13], could provide better distance estimations.

3.2.4 Conclusion

We investigated interaction spaces for 3DUIs in HMD environments. Our Fitts' Law experiment shows that the main factor causing errors in 3D selection is the visual perception and not, as presumed, the motor movement direction during selections. The error in the movement direction, which was presumed to be paramount, induces only part of the error.

We found significantly more errors along the viewing axis than along the orthogonal axes. With these results, we suggest increased tolerance along the viewing axis and proposed an ellipsoid cursor. Further research could investigate the effects of an increased error tolerance along the viewing axis, which could potentially improve selection task accuracy in IVEs. Moreover, our results indicate that the bottom regions, which are farther away along the viewing axis, are worse for precise tasks, despite the fact that they have similar IDs compared to visually close regions.

Considering that with increasing distance to the user, the viewing and motoric axes converge, the implications are mainly for close tasks. However, our results show that for precise tasks a close position to the eyes is preferable. Our results suggest that the ellipsoid cursor shown in Figure 3.13 can significantly improve selection performance over the tested interaction regions.

We propose the following guidelines for immersive HMD environments:

G_1 3D selection tasks, which require high precision, should be restricted to a level close to the eyes despite more comfortable interaction in the lower regions.

G_2 For accurate selections of spherical targets, such as points, we suggest inflating the selection space using an ellipsoid cursor as illustrated in Figure 3.13, which covers the 95% confidence region for touch selections.

3.3 Summary

In this chapter, we investigated selections in mid-air. First, in Section 3.1, we evaluated how users would expect to do hover selections for stereoscopic multi-touch tabletop setups. In a perceptual

experiment, we had participants mark areas above virtual objects, where they would expect a hover interaction to happen. We identified two mental models from our participants, which either expect hovering to happen when the target is visually occluded by their hand, or when their hand is above the object. The derived volumes, coined HoverSpace, were evaluated in a confirmatory study, wherein users performed better using the HoverSpace. However, should head-tracking be unavailable, the part of the HoverSpace just above the target object is adequate.

Whereas hover-interactions are indirect since the target object is not touched, we decided to investigate performance for direct, mid-air 3D interaction. In the second part of this chapter, Section 3.2, we investigated interaction spaces for 3DUIs in HMD environments. We could show with an experiment using Fitts' Law that the main factor for errors during 3D selection is caused by visual misperception and not imprecisions in the movement. Thus, we suggest to accommodate for these misperceptions, for example, by using ellipsoid selection volumes.

The results of our experiments allow improving the performance during 3D hover selections, as well as direct, mid-air 3D interaction. Nonetheless, especially during the experiment in Section 3.2, out participants complained about prolonged mid-air selections being too physically demanding. These insights led us to investigate how to make 3DUIs more comfortable since we expected prolonged physical exhaustion to have a negative impact on performance.

4. Improving Comfort for 3D Interaction

Abstract

During the work on performance, we noticed users subjectively complaining about aching limbs during longer experiments, which correlates with potential discomfort during prolonged use in production setups.

Comfort and fatigue are key factor to take into account during the design of 3DUIs [LaV+17]. However, the connection between comfort and high performance for 3D mid-air interaction has to be proven first. This chapter consists of three sections. In Section 4.1, we investigate the influence of the region in a user's interaction volume on performance. Afterward, we show that providing users with a more comfortable work-setup, by utilizing physical arm support, results in higher performance in selection tasks. Based on these results we developed, joint-centered user interfaces (JCUIs), which adaptively place interactive elements in a way that is most comfortable to reach when a user rests their elbow or wrist on a physical support. We present the JCUIs in Section 4.2. Finally, we in Section 4.3.1, we present a more innovative form of SNUIs by giving users more than one hand pair. These quad- or multi-manual user interfaces reduce the necessary movement distance to interactive objects.

4.1 Comfort and 3D Task Performance

The first experiment described in this chapter was conducted by [Hof+08], while the second experiment and the revision were conducted by [LBS15].

IVEs can offer more natural interaction in the three-dimensional (3D) space around a user. However, interaction performance in 3D mid-air is often reduced compared to 2D interaction (cf. Section 2.6). The performance depends on a variety of ergonomics factors, for example, the user's endurance, muscular strength, as well as fitness. In contrast to traditional desktop-based setups, users often cannot rest their arms in a comfortable pose during the interaction.

In this section, we analyze how comfort impacts 3D selection tasks in an *immersive desktop setup*. First, in a pre-study, we identified comfortable and uncomfortable interaction poses for users who are standing upright. Then, we investigated differences in 3D selection task performance when

users interact using their hands in a comfortable or uncomfortable body pose, while sitting on a chair at a table while the VE was displayed with an HMD. We conducted a Fitts' Law experiment to evaluate selection performance in different poses. The results suggest that users achieve a significantly higher performance in a comfortable pose when they rest their elbow on a table.

Stereoscopic visualization is getting becoming more critical for a range of application fields (cf. Section 2.3.2). Moreover, advances in the field of unobtrusive body-tracking, such as the Microsoft Kinect or the Leap Motion controller, afford natural interaction with 3D data sets (cf. Section 2.6). Stereoscopic display supports near-natural spatial impressions of virtual objects and scenes. Using the hands and body to touch and manipulate virtual objects, provides an intuitive, direct interface for interaction with stereoscopically displayed 3D content [Ste+12]. IVEs, such as tracked HMDs or CAVEs, thus have the potential to provide natural and intuitive interaction with virtual objects located in the proximity of the user's viewpoint. If an interactive virtual object is located within arm's reach, users can perform natural reach and touch gestures similar to the real world, whereas different IVEs provide users with different affordances for traveling to objects that are located at a larger distance (cf. Section 2.6).

Since interactive virtual objects could be located anywhere within arm's reach, the question arises how its position in different regions around the user's body affects comfort as well as interaction performance. Interaction in regions that users judge as *comfortable*, i.e., feeling physically relaxed without any pain or displeasing posture, could be beneficial for long interaction sessions. Performance may also differ depending on the muscle strength or flexibility of the user, and may be affected by the use of different muscle groups and levels of energy expenditure of the body. VR researchers and practitioners had observed through recorded participant comments and occasional complaints, that mid-air interactions can lead to arm fatigue and discomfort [LaV+17]. This was known before Hollywood popularized 3D direct interaction in mid-air, e.g., in different movies and TV series, such as *Minority Report* or in the *Iron Man* trilogy, where protagonists can interact with virtual content by touching virtual objects with outstretched arms. Mid-air interaction requires significantly more muscular energy than desktop interaction with a mouse and keyboard, and may thus not provide high performance during prolonged use, although the performance during the first minutes may be encouraging (cf. Section 2.4.4).

In this section, we compare direct 3D selection performance between users interacting in comfortable or uncomfortable poses in an immersive desktop setup. In such *immersive desktop setups*, users sit on a chair in front of a table while the VE is displayed on a HMD (cf. Figure 4.4). First, we performed a pre-study in which we determined regions that subjects rated as comfortable or uncomfortable. Then, we conducted a Fitts' Law experiment in which we compared 3D selection performance between users interacting in a comfortable and uncomfortable pose in the immersive desktop setup [LBS15]. In the comfortable pose, users rested the elbow of their dominant arm on the table, whereas in the uncomfortable pose, the users had to interact in mid-air above the table surface without resting the elbow.

In summary, our contributions include:

- An analysis of comfortable and uncomfortable regions within arm's reach,
- A comparison of selection performance in comfortable and uncomfortable poses in an immersive desktop setup, and
- Guidelines for designing 3DUIs for immersive desktop setups.

4.1.1 Pre-Study

In this section, we summarize the pre-study that we conducted to determine comfortable regions in 3D mid-air interaction within arm's reach, as well as the first study we conducted to find an influence of comfort on selection performance. Further details can be found in [Hof+13].

Figure 4.1: Position grid as used in the pre-study.

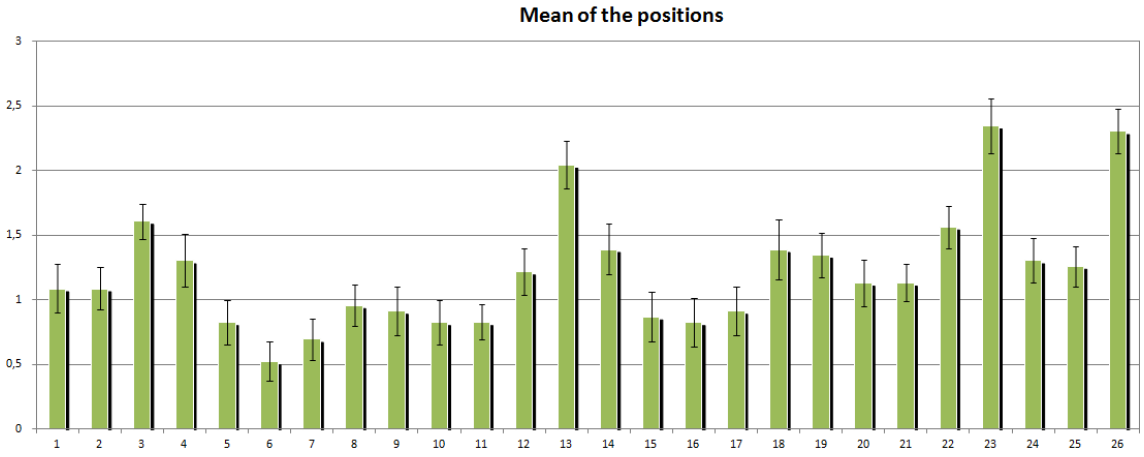


Figure 4.2: Means of the comfort values of every position.

4.1.1.1 Setup

For the study, a Sony HMZ-T1 HMD was tracked in 6 DOF by an 8 camera IOTracker while the subjects' index finger was tracked with a 3 DOF target.

4.1.1.2 Study Design

In a pre-study, we recruited 27 participants (16 female, 11 male, age 14–57, M = 31.5) and asked them to take 26 predefined positions with their dominant hand and arm (cf. Figure 4.1). All subjects were right-handed. As illustrated in Figure 4.1, we defined the positions based on a grid with two different depths and six different heights. The different depths were realized by instructing the

Figure 4.3: Heatmap illustration of comfortable and uncomfortable poses; red regions were evaluated as uncomfortable, green regions as comfortable poses.

subjects to flex or to stretch out their arms.

Subjects received minimal instructions about the goals of the study to prevent bias on the results. We randomized the positions, and every subject had to hold every position for precisely 10 seconds indicated by an acoustic signal. After every position, the subjects had to judge the comfort of that position using a 5-point ranking scale (rating scale 0 = very comfortable, 4 = very uncomfortable). The total time per subject was less than 20 minutes.

4.1.1.3 Results

We compared the means of the comfort values of every position (cf. Figure 4.2). The results imply that the most comfortable way to interact is close to the body with elbow flexing. The most comfortable positions had a mean below 1: positions 5, 6, 7, 10, and 11. In contrast, the least comfortable positions are the positions 3, 13, 24, 25, and 26. As illustrated in Figure 4.3, we found that positions with flexed arms are mostly considered more comfortable than positions with stretched-out arms.

In the first study, described in [Hof+13], direct mid-air selection performance was compared between users interacting in comfortable or uncomfortable regions within arm's reach in an HMD setup while standing upright. The results of the pre-study were generalized into flexed arms at a distance of less than about 65% of the maximum arm's reach for comfortable positions and stretched-out arms at a distance of more than 65% of the maximum arm's reach for uncomfortable positions. While we found a significant influence of the comfortable condition on the error distance, with larger errors in the uncomfortable condition, we did not find main effects of the comfortable condition on the error rate, movement time or throughput. However, we found a trend on the error rate with the mean error rate being lower for the comfortable condition. We also analyzed SSQ and SUS-PQ questionnaires but did not observe significant differences between the two conditions (cf. Section 2.4.2.4). However, all positions required some effort of the subjects. Also, the subjective opinion of our subjects clearly showed differences in the different conditions.

This motivated us to consider situations in which a more significant distinction can be made between a comfortable and uncomfortable pose (cf. Section 4.1.2).

4.1.2 Experiment

In this section, we describe the Fitts' Law experiment in which we analyzed direct 3D selection in the user's arm's reach in an immersive desktop-based HMD environment.

Motivation

While we could not find a significant main effect of comfort on selection performance measured by throughput in the previous studies, the subjective opinions, as well as the trends we found motivated us to investigate the influence of comfort in a further study.

Instead of defining stretched out arms as uncomfortable and arms with a degree of elbow flexing as comfortable, requiring participants to keep their arms aloft nonetheless (cf. [Hof+13]), we decided to evaluate a different definition of comfort, namely arm support. Requiring participants to keep their arm aloft is considered less comfortable than providing an armrest for the participant to place their elbow on, removing the tension from the shoulder and arm muscles. Although the concept of comfort is broad and should not be overly simplified, we believe that a reduction of muscle tension over time is an adequate facet of the broad concept of comfort. For more references on fatigue please refer to Section 2.4.4.

While the first evaluation was in an upright position, we decided to conduct the study in a seated position in both conditions, comfortable due to arm support and uncomfortable without. Supporting the arm has been mentioned in research before, but fatigue and performance in 3DUIs have not been evaluated in detail before [Bér+09] (cf. Section 2.4.4).

4.1.2.1 Participants

We recruited 12 participants for our experiment. Nine of them were male, and three were female (ages 19–36, M = 25.33). All of the subjects were either students or professionals in HCI or media communication from the University of Würzburg. The students received class credit points for the participation in the experiment. Two subjects were left-handed, the remaining ten subjects were right-handed. All subjects had normal or corrected vision.

Using the technique proposed by Willemsen et al. [Wil+08] we measured the IPD of each subject before the experiment started (M = 6.35 cm, SD = 0.44 cm). Additionally, we used the Porta and Dolman tests to determine the dominant sighting eye of subjects [MOB03] (cf. Section 2.4.2.1). The test revealed that two were left-eye dominant (2 male), whereas eight subjects were right-eye dominant (6 male, 2 female). For two subjects, both tests yielded different results (1 male, 1 female). We measured the lower arm length (M = 37.00 cm, SD = 3.47 cm) and calibrated the VE for each subject accordingly (cf. Section 4.1.2.2).

All subjects reported at least some experience with 3D stereoscopy (rating scale 0 = no, 4 = yes, M = 3.42, SD = .79). Five subjects reported that they have participated in HMD studies before, and eleven subjects reported experience with HMDs (rating scale 0 = no experience, 4 = much experience, M = 2.00, SD = 1.86). The remaining seven subjects had no experience with HMDs. All subjects were naive to the experimental conditions.

The mean of the total time per subject, including questionnaires and instructions, was about 40 minutes. The mean time for performing the actual experiment while wearing the HMD was about 26 minutes (M = 25.52, SD = 06.24). Subjects were allowed to take breaks in between trials.

4.1.2.2 Materials

As illustrated in Figure 4.4, subjects were instructed to sit in an upright position facing towards the cameras of the optical tracking system. A Razer Nostromo keypad was adjusted to a comfortable position for the non-dominant hand of the user. Subjects were instructed to keep their hand in that position during the experiment to confirm their selections.

The experiment was conducted with the user wearing an Oculus Rift (Developer Edition) HMD with an attached active infrared (IR) target. The target was tracked by an optical WorldViz Precision

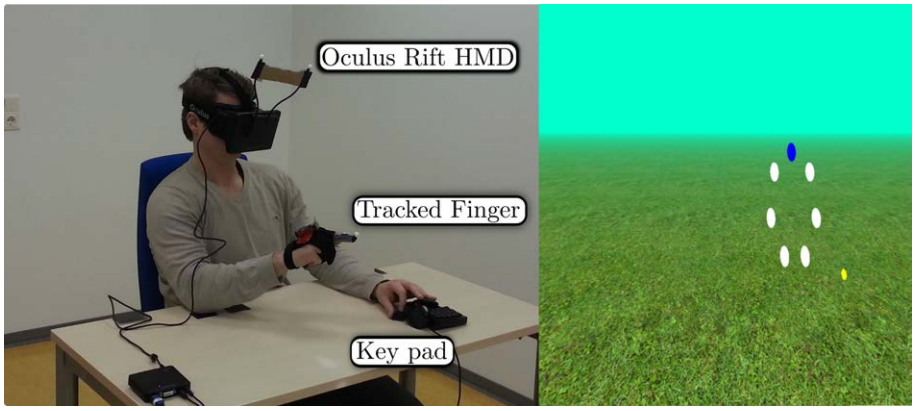


Figure 4.4: (left) User during the experiment. (right) The user's view to the VE with the white and blue target spheres as well as the yellow sphere that visually represents the tracked finger tip.

Position Tracking (PPT X4) system with sub-millimeter precision for view-dependent rendering. We combined the PPT's optical heading with the inertial orientation of the Oculus Rift to provide robust head tracking without drifts. The Oculus Rift offers a horizontal FOV of approximately 90° and a vertical FOV of 110° at a resolution of 1280 × 800 pixels (640 × 800 for each eye)(cf. Section 2.5).

Additionally, we attached a single IR target to the index finger of the subject's dominant hand (cf. Figure 4.4(left)). The virtual stimulus (cf. Figure 4.4(right)) used in the experiment displayed a 3D scene that was rendered with OpenGL on an Intel computer with a Core i7 3.8 GHz CPU, 8GB of main memory and Nvidia GeForce GTX580 graphics card. For the Oculus Rift, we rendered the virtual scene side-by-side and applied a barrel distortion.

Spheres represented the experiment targets. All target sphere for one trial were always visible and colored white, except for the current target. The current target sphere was colored blue when the marker was outside, and green when the marker was inside to give the subjects a visual cue. As typical in Fitts' law tasks, the odd number of spheres were lighted in the order specified by the ISO 9241-9 standard [Int00] (cf. Section 2.6.1.1). As illustrated in Figure 4.5, each trial consisted of an arrangement of 7 target spheres, forming a circle with each sphere at the same distance to the subject's elbow.

4.1.2.3 Methods

We used a 2 × 2 × 2 × 3 within-subject design with the method of constant stimuli for the experiment trials. In condition *comfortable*, subjects rested their elbow on a table, whereas in condition *uncomfortable*, subjects were not allowed to rest their arm during the trial.

Besides, we used two distances (8 cm and 12.5 cm) between spheres as well as two sphere radii (2 cm and 3 cm). This resulted in four IDs (2.32 bps, 1.87 bps, 2.86 bps, 2.37 bps) representing a valuable range of task difficulties for such 3DUI setups. IDs outside this range seemed impractical for 3D user interfaces, as the spheres would be either too large, taking up too much space or too small, to be precisely selected in mid-air.

The two sphere radii, two distances, and three interaction positions were uniformly and randomly distributed. Each resulting condition was tested five times, resulting in 60 trials for the comfortable and uncomfortable condition respectively. We counterbalanced the order between subjects; half of the subjects started with the comfortable condition followed by the uncomfortable one and vice versa.

Following both oral and written instructions, for both conditions, subjects completed between 2 and 10 supervised training trials for the experimental phase to ensure that they understood the task

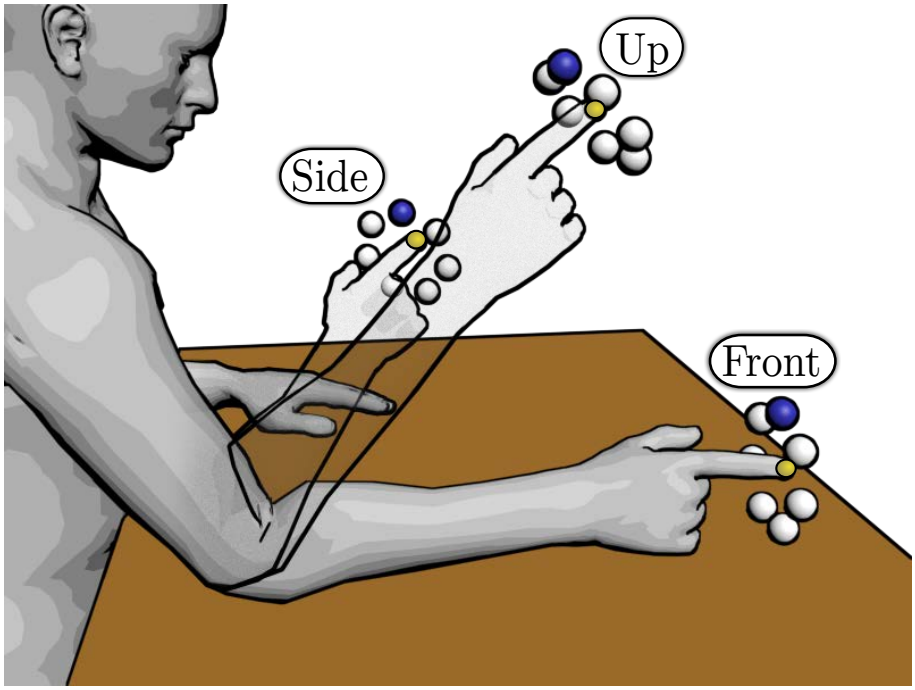


Figure 4.5: Schematic illustration depicting the three interaction positions. Only one ring of spheres was active during a single trial. The blue sphere depicts the target sphere. The yellow sphere depicts the sphere the user saw at the location of the fingertip.

correctly. Those training trials were excluded from the analysis.

Each trial consisted of sequential selections of all 7 targets in a full ring, resulting in a total of 7 selections per trial. The starting position for the trial was the first lit sphere, meaning that the measured distances were always between two target spheres. Since the distances between the spheres were the same, this ensured the selections were valid for the computation of the ID. The subjects were instructed to select the targets as quickly and accurately as possible, as it is common in most Fitts' Law experiments [DKK07; Mac+87; MI08](cf. Section 2.6.1.1).

While selecting a target correctly, subjects received visual feedback by changing the color of that target sphere from blue to green. Like previous studies on mid-air-selection [BSS13a; BSS13c], the subjects had to confirm each selection by pressing a button on the keypad with their non-dominant hand to avoid the Heisenberg effect caused by pressing buttons with their dominant hand [Bow+01].

We computed the distance of the position of the index finger to the sphere center, which indicated a selection if the distance was less than the radius of the target sphere. If subjects performed a selection while the target sphere was not highlighted green, we recorded this as a selection error and advanced the trial state. The dependent variables were movement time, error distance (deviation from target center), error rate (percentage of targets missed), and effective throughput (cf. Section 2.6.1.1).

For both conditions, we calibrated the elbow position. With the calibrated position of the user's elbow, we defined 3 *interaction positions* (rings) in front of the user, as illustrated in Figure 4.5. The regions are reachable without moving the elbow rested on a table. Previous studies indicate that regions close to the user's eyes offer increased performance over far regions [LBS14a]. Also, according to our previous studies, the positions with elbow flexion are comfortable and divide the interaction space for the hand [Hof+13]. We ruled out üositions further away from the nondominant hand, i.e., taking into account Figure 4.5, the other side. These positions would force the participant

to either move their elbow, which would complicate further selections, or twist their arm. The positions are the same for both conditions. To ensure equal distances between the elbow and the target spheres, we rotated the ring of spheres accordingly. Only one of the sphere rings in Figure 4.5 is visible at a time.

Additionally to the main experiment metrics, we asked subjects to complete subjective questionnaires (cf. Section 2.4.2.4 and Section 2.4.2.4). Before and after the experiment subjects were asked to complete an SSQ [Ken+93]. After the experimental phase, subjects were asked to complete an SUS-PQ [Uso+99b]. Moreover, the subjects had to judge their level of comfort after each condition (comfortable or uncomfortable).

4.1.2.4 Hypotheses

Based on previous findings (cf. Section 2.6.1 and Section 2.4), we evaluate the following three hypotheses:

- $\mathbf{H}_1$ The performance measured by throughput is higher in a comfortable pose with an armrest.
- $\mathbf{H}_2$ The number of selection errors is lower in a comfortable pose with an armrest.
- $\mathbf{H}_3$ The subjective comfort level is higher when the subjects can rest their arm during selection tasks.

4.1.3 Results and Discussion

In the following section, we summarize the results of the experiment. Results from the subjects were normally distributed according to a Shapiro-Wilk test at the 5% level. We analyzed the results with a repeated measure ANOVA and Tukey multiple comparisons at the 5% significance level. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated.

4.1.3.1 Selection Performance

In this section, we describe the analysis of the Fitts' Law selection performance of our experiment. In particular, we analyze the movement time between subsequent selections, the rate of missed targets, error distances to target centers, and effective throughput. The effective throughput metric incorporates both the results for speed and accuracy to form an estimate of the bottom-line performance (cf. Section 2.6.1.1).

Error Distance

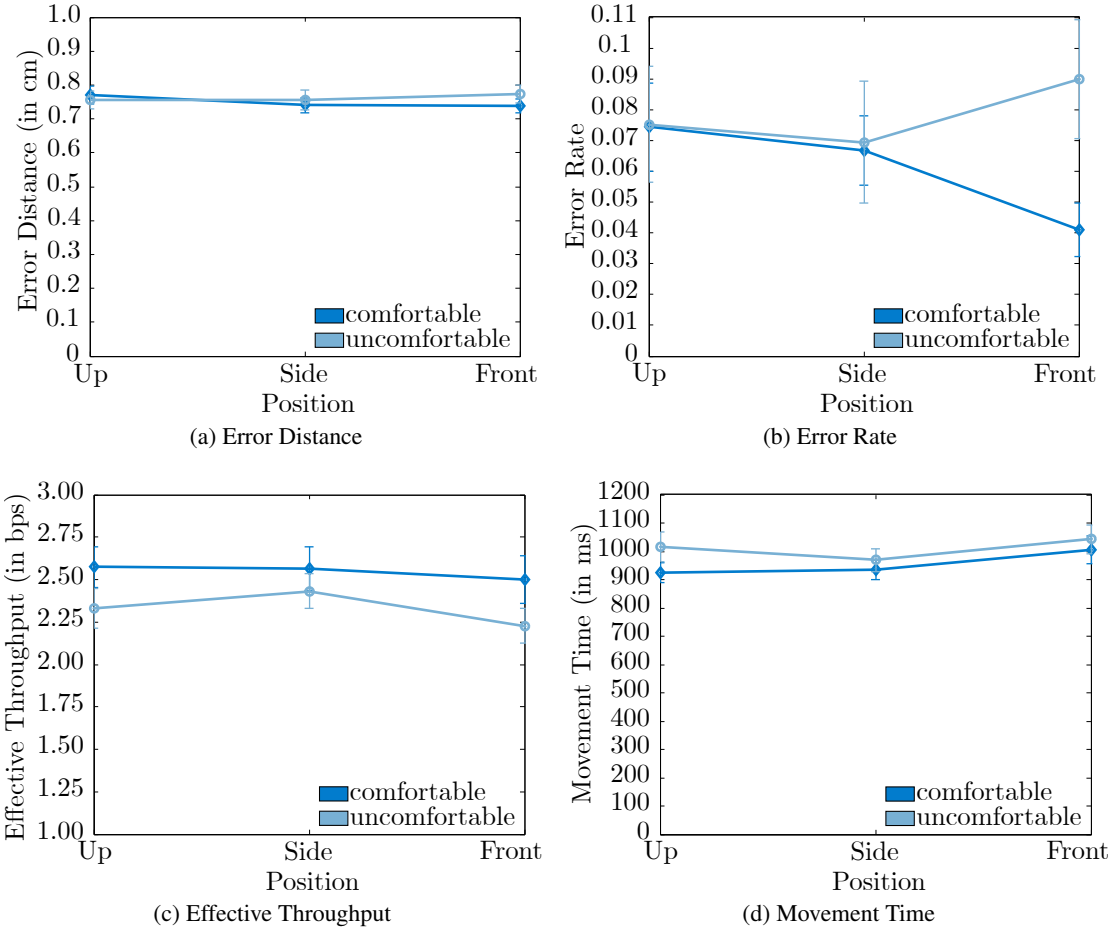
The results for error distances between the center of each sphere and the finger position during selections are shown in Figure 4.6(a). As expected by Fitts' Law, the results show that the error distance for both sphere distances significantly differs ($F(1,11) = 6.116$, $p < .05$, $\eta_p^2 = .357$). As expected, the sphere radius had a significant effect on the error distance ($F(1,11) = 212.734$, $p < .001$, $\eta_p^2 = .951$).

The average error distance during the experiment was M = 7.556 mm (SD = 1.603 mm). For the comfortable condition the average was M = 7.497 mm (SD = 1.596 mm), which was lower than for the uncomfortable condition (M = 7.613 mm, SD = 1.614 mm).

For the small sphere distance, the average error distance M = 7.485 mm (SD = 1.164 mm) was significantly lower than the error distance for the large sphere distance M = 7.626 mm (SD = 1.565 mm).

For the small sphere radius, the average error distance M = 6.351 mm (SD = 0.997 mm) was significantly lower than the error distance for the large sphere radius M = 8.760 mm (SD = 1.115 mm).

The results show an interaction effect of position and distance ($F(1.324, 14.564) = 5.223$, $p < .05$, $\eta_p^2 = .322$) and of position, distance and radius ($F(2,22) = 4.22$, $p < .05$, $\eta_p^2 = .277$). We found no significant main effect of position ($F(2,22) = .931$, $p = .449$, $\eta_p^2 = .07$) on error distance. Furthermore, we could not find any significant effect of comfort ($F(1,11) = .251$, $p = .626$, $\eta_p^2 = .022$) on error distance.



(a) Error Distance

(b) Error Rate

(c) Effective Throughput

(d) Movement Time

Figure 4.6: Plots of the experiment's results. The x-axis show the interaction positions and the y-axis the a) error distances in cm, b) error rate, c) effective throughput in bits per second (bps) and d) the movement time in ms. The green lines show the comfortable pose. The blue lines show the uncomfortable pose

Finally, an Independent Samples T-Test resulted in no significant difference in error distance between left- and right-handed subjects ($p = .934$).

Error Rate

The results for error rate are shown in Figure 4.6(b). The results show that the error rate for the two sphere distances significantly differs ($F(1,11) = 9.009$, $p < .05$, $\eta_p^2 = .45$). As expected, also the radius had a significant effect on the error rate ($F(1,11) = 26.224$, $p < .001$, $\eta_p^2 = .705$).

The average error rate during the experiment was M = 6.94% (SD = 7.34%). For the comfortable condition the average was M = 6.07% (SD = 6.13%) lower than the average for the uncomfortable condition, which was M = 7.81% (SD = 8.31%).

We measured a significantly lower average error rate M = 6.31% (SD = 7.63%) for the small sphere distance compared to the large sphere distance (M = 7.58%, SD = 7.01%).

As expected, for the small sphere radius the average error rate M = 9.22% (SD = 8.56%) was significantly higher than for the large sphere radius (M = 4.66%, SD = 4.93%).

We found an interaction effect of comfort and position ($F(2,22) = 6.794$, $p < .05$, $\eta_p^2 = .382$) and the results show an interaction effect of position, distance and radius ($F(2,22) = 5.683$, $p < .05$, $\eta_p^2 = .341$). However, we found no significant main effect of the position ($F(2,22) = 1.242$,

$p = .308$, $\eta_p^2 = .101$) on the error rate, or of the comfort on the error rate ($F(1,11) = 1.646$, $p = .226$, $\eta_p^2 = .130$).

Finally, an Independent Samples T-Test revealed a significant difference in error rate between left- and right-handed subjects ($p < .05$). The mean error rate for left-handed subjects (M = 8.89%, SD = 7.54%) was significantly higher than the error rate of right-handed subjects (M = 6.55%, SD = 7.25%).

Effective Throughput

The results for the effective throughput are shown in Figure 4.6(c). We found a significant main effect of the comfortable condition ($F(1,11) = 6.157$, $p < .05$, $\eta_p^2 = .359$) on throughput. Additionally, we found a significant main effect of sphere radius ($F(1,11) = 8.065$, $p < .05$, $\eta_p^2 = .423$) on throughput. The average throughput during the experiment was M = 2.44 bps (SD = .47 bps). Also, the average throughput for the comfortable condition M = 2.55 bps (SD = .49 bps) was significantly higher than for the uncomfortable condition M = 2.33 bps (SD = .43 bps). For the small sphere radius the average with M = 2.378 bps (SD = .103 bps) was lower than for big sphere radius M = 2.498 bps (SD = .105 bps). The results show no significant effect of the three positions ($F(1.38, 15.176) = 2.157$, $p = .159$, $\eta_p^2 = .164$) on throughput. Also we found no significant effect of the sphere distance ($F(1,11) = 4.078$, $p = .068$, $\eta_p^2 = .270$) on the effective throughput. Furthermore, we found a significant interaction effect between position and distance ($F(2,22) = 4.795$, $p < .05$, $\eta_p^2 = .304$). An Independent samples T-Test revealed no significant difference in effective throughput between left- and right-handed subjects ($p = .887$).

Movement Time

The results for the movement time are shown in Figure 4.6(d). We found a significant main effect of the target position ($F(2,22) = 4.117$, $p < .05$, $\eta_p^2 = .272$) on movement time. As expected, the movement time for the two distances differs significantly ($F(1,11) = 152.64$, $p < .001$, $\eta_p^2 = .933$) and also the radius had a significant effect on the movement time ($F(1,11) = 60.72$, $p < .001$, $\eta_p^2 = .847$).

The average movement time during the experiment was M = 982 ms (SD = 205 ms). For the comfortable condition, the average movement time M = 955 ms (SD = 194 ms) was shorter than for the uncomfortable condition, for which it was M = 1010 ms (SD = 212 ms).

As expected, for the small distance the movement time M = 903 ms (SD = 164 ms) was significantly faster than for the large distance M = 1061 ms (SD = 212 ms).

Also, the mean movement time for the small sphere radius M = 1061 ms (SD = 210 ms) was slower than for the large sphere radius M = 904 ms (SD = 166 ms).

Post-hoc tests revealed that the movement time was significantly lowers when the target was displayed at the Side position, compared to the Front position ($p < .05$). However, we found no significance between the Side and Up ($p = 0.481$) or the Front and Up ($p = .096$) conditions. We found a significant interaction effect between position and radius ($F(2,22) = 4.904$, $p < .05$, $\eta_p^2 = .308$) and distance and radius ($F(2,22) = 14.24$, $p < .05$, $\eta_p^2 = .564$). An Independent Samples T-Test revealed no significant difference in movement time between left- and right-handed subjects ($p = .172$).

4.1.3.2 Modeling

Fitts' Law can be used as a predictive model by regressing movement time on the index of difficulty. The regression lines for movement time are presented in Figure 4.7. The predictive quality of the model, as expressed by χ^2 values $(.12, .27)$, is very high for both the comfortable and uncomfortable condition.

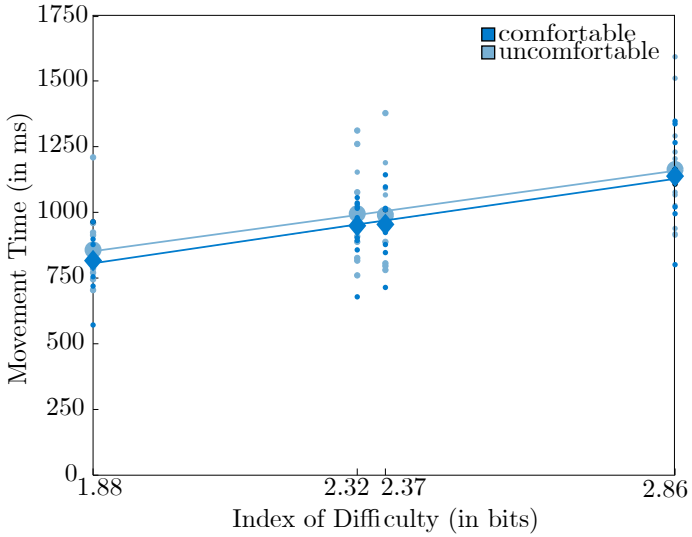


Figure 4.7: Models for the selection: The lines show the regressions of the measured movement time for the comfortable (green) and uncomfortable (blue) pose.

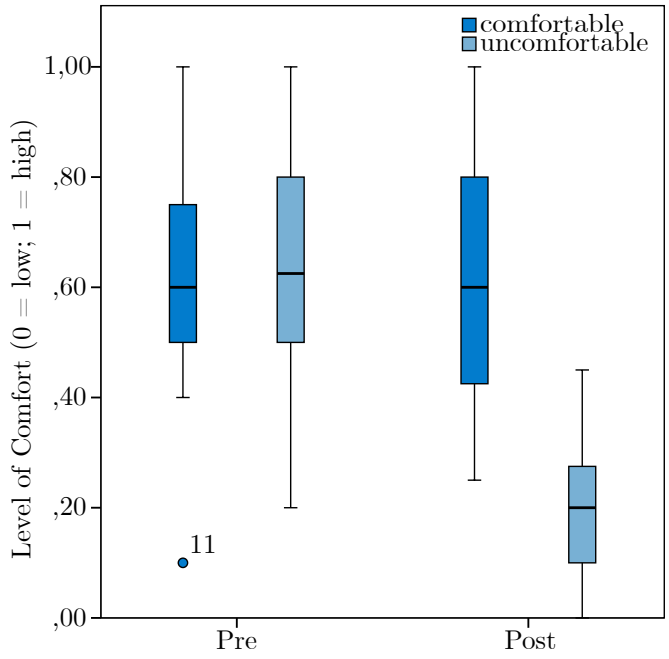


Figure 4.8: Boxplots showing the subjective comfort level before (pre) and after (post) the condition: the green bars show the comfortable condition, the blue bars show the uncomfortable condition.

4.1.3.3 Questionnaires

Before and after the experiment, we asked subjects to judge their level of cybersickness as well as their subjective sense of presence. While we measured an average pre-SSQ score of M = 8.10 (SD = 9.14), the post-SSQ score was M = 13.09 (SD = 13.29). We found no significant effect of the increase in cybersickness over the time of the experiment ($t(11) = -1.71$, $p = .116$). The mean SUS-PQ score for the reported sense of feeling present in the virtual scene was M = 3.65 (SD = .82), which indicates a reasonably high level of presence.

Besides the SSQ, we asked subjects to judge their subjective level of postural comfort during the experiment by giving subjective values on a scale of 0 (uncomfortable) to 1 (comfortable). We

analyzed these subjective comfort levels, as shown in Figure 4.8. The plot shows the subjective comfort levels before (pre) and after (post) the condition was tested. It shows a lower average in the condition without an armrest (uncomfortable). Taking into account the participants comment, that the table was hard and uncomfortable, the outlier can be explained. Paired T-Tests show no significant change of subjective comfort in the condition with an armrest (comfortable) ($t(11) = .238$, $p = .817$). However, a Paired T-Test for the uncomfortable condition shows a significant decrease of subjective comfort ($t(11) = 4.343$, $p < .05$). We compared the measured subjective comfort of both conditions before the respective trials (pre) using an Independent Samples T-Test and found no significant difference ($t(22) = .299$, $p = .768$). However, the Independent Samples T-Test for the comfort levels after the condition was tested (post) shows a significant difference between the subjective comfort ($t(22) = 5.367$, $p < .001$).

4.1.3.4 Discussion

The results of the Fitts' Law experiment revealed the importance of comfort for 3DUIs. Offering subjects a possibility to rest their arm, or specifically their elbow, comfortably allows them to utilize 3DUIs in an IVE efficiently. The significantly increased throughput with an armrest indicates a significant improvement over the free mid-air interaction. This result confirms our hypothesis H_1.

Additionally, because comfort is very subjective, the evaluation of the subjective estimation of the participants' comfort level showed that the condition with the armrest was preferable to the condition without. While at the beginning of the trials we could not find a significant difference between the conditions, by the end of the trial the difference was significant. This result supports our hypothesis H_3. Another interesting note is that, as Figure 4.8 indicates, some subjects judged an increase of comfort over the course of the experiment in the comfortable condition. Some subjects informally noted that they were not used to the pose at first, but got used to it quickly. Also, several subjects left comments such as, "Resting my arm helped a lot!"

However, the results show that, contrary to H_2, the comfortable pose with an armrest did not have a significant influence on the error rate. This could be due to the limited precision a finger offers during manual selection due to hand jitter. Another limiting aspect might be the low resolution of the utilized Oculus Rift HMD. Future work could evaluate higher resolution HMDs to determine whether the low resolution was a factor limiting the influence of comfort. However, while we conducted our studies in a fully IVE, as previous studies concerning fatigue suggest [Sha98], our results are most likely valid for semi-immersive VR setups. The main difference between those setups is the visualization. However, in such setups, the user's hands always occlude virtual objects and the vergence-accommodation conflict arises, which might cause further problems.

While we found different interaction effects for the different performance metrics related to Fitts' Law, we also found an interaction effect between comfort and position for the error rate. We measured fewer errors in the comfortable condition of the front position than in the uncomfortable condition and positions. Additionally, we found an interaction effect of position and distance on error distance and throughput. The effect may be explained by the effects of depth perception in VR during selection tasks [LBS14a].

4.1.4 Conclusion and Future Work

In this section, we investigated the importance of comfort for the performance in 3DUIs. We first summarized the pre-study in which we determined which regions are considered comfortable when interacting in mid-air while standing. Based on the results, we decided to analyze a more comfortable body poses for the interaction. In particular, we analyzed the influence of a comfortable armrest on performance in immersive desktop setups. Many Science Fiction movies and novels envision a plethora of vivid virtual systems which demand that the user stands within them to navigate and control the system. Meanwhile the user has to move, utilizing their whole arms. Prolonged use of such systems turns out to be immensely tasking for the user. To permit users to

experience the immersive experience such systems potentially offer without the negative aspects, we evaluated possibilities to increase the comfort. One possible choice would be to allow users to rest their arm and to seat them in something similar to standard desktop workspaces. For many application scenarios, for instance, when navigation is not required, this pose has the potential to be beneficial for long-term use. The conducted Fitts' Law experiment yielded interesting results for such immersive desktop setups. To our knowledge, our experiment is one of the first studies that showed that comfort has a significant effect on effective throughput according to Fitts' Law in IVEs. Despite the high immersion of modern HMDs, it is necessary to design the real environment to the user's comfort carefully.

These results offer interesting guidelines for future 3DUIs in immersive desktop setups:

G_1 If it is possible to use an elbow rest, make use of *elbow-centered 3DUI menus*, characterized by elements that are optimally reachable along a partial sphere centered around the elbow rest without the necessity to lift the arm.

G_2 Along the elbow-centered partial sphere, menu elements should be displayed in decreasing order of importance from the head position, i.e., the most important elements should be displayed close to the eyes whereas the least important objects should be displayed towards the far end of the partial sphere (cf. [LBS14a]).

Following these guidelines, future work could evaluate the effectiveness of established navigation and selection techniques adapted to the immersive desktop setup. Even novel input or haptic feedback devices included in an armrest could be developed, which could further increase the level of immersion in VEs. Further studies have confirmed our results, showing that supported gesture interactions cause less exertion [Han+17].

Knowing that performance can decrease after prolonged use, we aim to use this knowledge to develop UIs that are physically less demanding during prolonged use.

4.2 Joint-Centered User Interfaces

This section is based our work previously published as [Lub+16b]. Recent developments in the field of VR and AR HMDs such as the Oculus Rift, HTC VIVE, or Hololens have received much public interest and praise for their design, tracking, and visual quality. However, they also received much critique for their spatial interaction and, in particular, for how they handled interaction with 3D GUIs. By combining HMDs with depth sensors, such as the Leap Motion, it has become possible to leverage direct hand input for spatial interaction, as also seen in the Meta and Hololens. User interfaces based on direct hand interaction, however, received less user acceptance than expected, such that Oculus and HTC decided to ship their HMDs with hand-held controllers. While portions of the challenge lie on the hardware side with still limited tracking ranges and accuracy, it has become necessary to revisit spatial menu designs and to develop methods to improve the user acceptance of free-hand spatial user interfaces by considering the human factors of 3D mid-air interaction [Man01].

In this section, we revisit mid-air interaction based on the observation that for the human jointed arm structure the reachable space from each joint is naturally limited. We refer to such an interactive volume of space around the center of a user's kinematic joint as a *joint-centered kinesphere* (cf. Section 2.4.4). While traditional head-centered menus are by design placed within arm's reach in front of the user, it is seldom the case that they can be reached without requiring movements of the shoulder joint. Such lifting of the arm and shoulder induces high levels of fatigue over time [CMR91]. Moreover, most traditional menus are not matched to each user's arm length but rather use population means. This means that for some users the menus are placed closer to their maximal reachable arm distance than for others. Requirements for the visual placement and size of spatial GUI elements for good visibility and high interaction performance have been extensively

researched [Bow99; LBS14a], but their optimal location in the kinematic chain has received less attention.

To evaluate the optimal placement of menus in the scope of joint-centered kinespheres, we conducted an experiment in which we collected quantitative and qualitative data on usability and selection performance. We compared interaction targets located at different distances on such kinespheres with a Fitts' Law task. The results support the notion that the efficiency of selections in spatial menus within arm's reach can be improved by placing interactive elements on joint-centered kinespheres at certain distances, in particular at the maximal reachable distance. We confirm the practical usability of joint-centered menus in an experiment by comparing them against traditional head-centered GUIs. We discuss the advantages and limitations of this method focusing on differences in the length of the kinematic chain for kinespheres centered around the shoulder, elbow or wrist joint.

In this work we provide the following contributions:

- We evaluate the placement of interactive elements in joint-centered kinespheres with a view on usability and touch performance, and
- We present a JCUI and compare user acceptance against head-centered menus.

4.2.1 Joint-Centered Interaction

In this section, we describe the motivation and rationale of joint-centered interaction, as well as the single components.

4.2.1.1 Kinespheres

Figure 4.9 illustrates the kinespheres at the maximal reachable distance for three joints typically involved in 3D mid-air interaction. According to the biomechanics and human factors literature, when assuming that a joint (e.g., wrist, elbow or shoulder) is rested and the remaining joints are stretched, the user's fingertip can approximately trace a partial sphere centered on the joint (cf. Section 2.4.4 and Figure 4.9). Finger positions close to a kinesphere are usually regarded as easily reachable and comfortable, whereas it is harder to reach the mid-region of a kinesphere closer to the center joint. This fact has been utilized in previous research using joint angles and hyperplanes [Gui+15; JPG14].

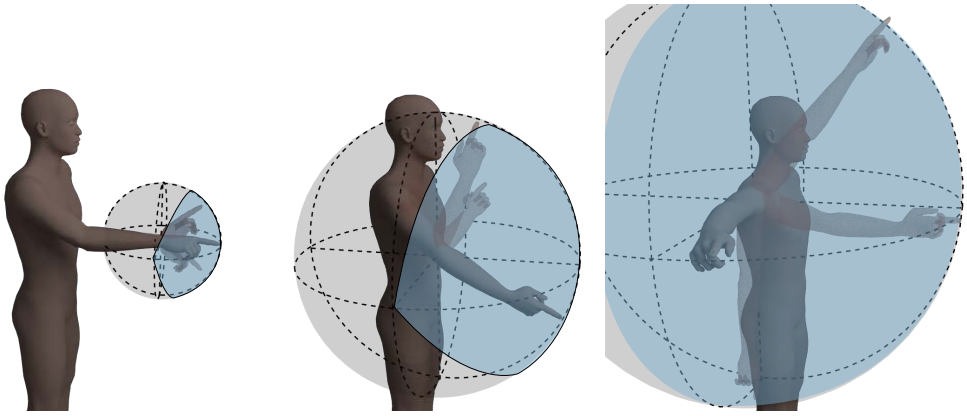
4.2.1.2 Kinematic Chain

We refer to the term kinematic chain as a series of links connected by joints, or more specifically, a human's arm from the shoulder joint to a fingertip (cf. Section 2.4.4). Previous studies have shown that interaction higher up the kinematic chain is less accurate and causes higher energy expenditure if the corresponding muscle groups are used, and more accurate and less fatiguing further down [BM97; SM87]. Optimally, physical support at a joint as low as possible in the kinematic chain allows users to relax muscle groups higher in the kinematic chain, reducing exertion and increasing comfort. In contrast, traditional head-centered spatial menus are placed within arm's reach in front of the user's eyes, requiring movements of the entire kinematic chain.

4.2.1.3 Boundary

The boundary of a joint-centered kinesphere defines a 2D manifold in 3D space that, conceptually, reduces the interaction complexity of touch gestures concerning the DOF of arm movements from three to two since the finger's distance from the joint is clamped to its maximum and does not have to be controlled consciously. Since large overshooting and undershooting are not possible as long as the corresponding kinematic chain is stretched, selection performance can be described by Fitts' Law in 2D during the correction phase, using visual perception only to correct for lateral offsets from the target.

2D GUI elements may be placed anywhere on the accessible part of a kinesphere, i.e., on the



(a) (b) (c)

Figure 4.9: Illustrations showing the joint-centered kinespheres around (a) the wrist joint, (b) the elbow joint, and (c) the shoulder joint for a standing user for 3D mid-air interaction. If the user is seated or uses an arm rest the comfort is increased while the corresponding kinespheres are reduced. The reachable area on the kinesphere is shown in green.

opposite side from the user's head, while ensuring visibility and low muscle exertion. Arranging 2D GUI elements or widgets at the maximal reachable distance, e.g., from the shoulder, elbow or wrist, might thus provide advantages in performance based on Fitts' Law over placing them anywhere else within arm's reach. Moreover, in contrast to pointing techniques, this approach has the advantage of maintaining the superimposed perceptual and motor spaces. It also maintains direct interaction with the graphical elements since they are located within arm's reach. However, we have to consider that fully-stretched kinematic chains are regarded as less comfortable than slightly relaxed joints. Thus, although placing menus slightly inwards from the boundary may lose the advantages described above, it may lead to higher user acceptance.

4.2.1.4 Implementation

For joint-centered menus, we align the touch-enabled 2D GUI elements in a local coordinate system around the center joint. The distance is limited by the maximal reachable distance and biomechanics (cf. Section 2.4.4). Concerning the kinematic chain we have implemented joint-centered 2D GUIs around the following joints:

The shoulder joint is a ball-and-socket joint providing the largest kinesphere (cf. Figure 4.9(c)). With a stretched elbow and wrist such that upper arm, forearm and index finger are colinear, the user can select targets on the kinesphere.

The elbow joint is the synovial hinge joint proving the second largest kinesphere (cf. Figure 4.9(b)). Assuming a stretched wrist such that forearm and index finger are colinear, the user can select targets on the kinesphere.

The wrist joint that bridges the hand to the forearm proving the smallest kinesphere in our implementation (cf. Figure 4.9(a)). With the stretched index finger the user can select targets displayed on the surface of the kinesphere.

In our implementation, the application can choose whether the shoulder, elbow or wrist kinespheres are used to present a joint-centered menu. Informal testing also included various other joints of the human body at finger level, legs, trunk, and back. We always chose the joint furthest down in the kinematic chain as the center of our kinesphere, if they are rested using physical support in the workspace. For instance, when a user is seated in an office chair, they can place their wrist or elbow on the armrest, and the interface elements are centered around the respective

joint (cf. Figure 4.10). Alternatively, without physical support, a kinesphere is chosen to minimize occlusion of the environment and proximity to affected targets while maximizing user preference and selection accuracy (cf. Section 4.2.2).

Interaction with the joint-centered GUI elements is implemented following the literature based on the three steps [AA13; Bow99]:

1. **Indication of the target object** by touching,
2. **Confirmation of selection**, e.g., with a touch or flip gesture with the dominant or non-dominant hand, voice command or timed response, and
3. **Feedback** generated by the system.

The joint-centered kinespheres described in this section contribute mainly to the first component. Interaction designers may combine the joint-centered approach to indicate target objects with any of the confirmation and feedback methods presented in the literature as desired (cf. Section 2.6.1).

Overall, this describes the conceptual part of joint-centered approaches. We implemented the ideas using the Unity engine and a PPT active IR optical tracking system. More information about the hardware and the implementation is described in Sections 4.2.2, 4.2.3 and Section 2.5.

4.2.2 Experiment

We conducted a Fitts' Law experiment to evaluate selection performance of GUI elements within arm's reach focusing on the relationships between joint, distance, performance, comfort and user acceptance.

4.2.2.1 Participants

20 participants (5 female and 15 male, ages 20–36, $M = 26.1$) took part in the experiment. The participants were members or students of the local department of computer science, who obtained class credit for their participation. All of our participants had normal or corrected-to-normal vision. Seven participants wore glasses, and one participant wore contact lenses during the experiment. None of our participants reported a disorder of equilibrium or a motor disorder such as impaired hand-eye coordination. One of our participants reported color blindness; no other vision disorders have been reported. 10 participants had participated in an experiment involving HMDs before. 18 participants were right-handed, and 2 were left-handed. We measured the IPD of our participants before the experiment using the measurement tool of the Oculus Rift configuration utility. The IPDs of our participants ranged between 6.2 cm to 7.2 cm ($M = 6.5$ cm, $SD = 0.2$ cm). We used the IPD of each participant to provide a correct perspective and stereoscopic rendering on the HMD. We measured each participant's maximal reaching distance from joint to fingertip for the shoulder joint ($M = 0.726$ m, $SD = 0.055$ m), for the elbow joint ($M = 0.423$ m, $SD = 0.036$ m), and for the wrist joint ($M = 0.195$ m, $SD = 0.027$ m). We confirmed each participant's ability to perceive binocular depth before the experiment via stereograms. 17 participants had prior experience with 3D stereoscopic display (cinema, games, etc.). The total time per participant, including pre-questionnaires, instructions, experiment, breaks, post-questionnaires, and debriefing, was 1 hour. Participants wore the HMD for approximately 45–50 minutes. They were encouraged to take regular breaks between trials to rest their arm.

4.2.2.2 Material

As illustrated in Figure 4.10, participants were instructed to sit in an upright position facing towards the cameras of the optical tracking system. The experiment was conducted with the user wearing an Oculus Rift (Developer Kit 2) HMD with an attached active infrared (IR) target. Additionally, we attached an IR target to the index finger of the participant's dominant hand (cf. Figure 4.10). The targets were tracked by an optical WorldViz Precision Position Tracking (PPT X4) system with sub-millimeter precision, which was chosen to offer maximal precision for the Fitts' Law task. The visual stimulus displayed during the experiment showed a 3D scene, which was rendered with

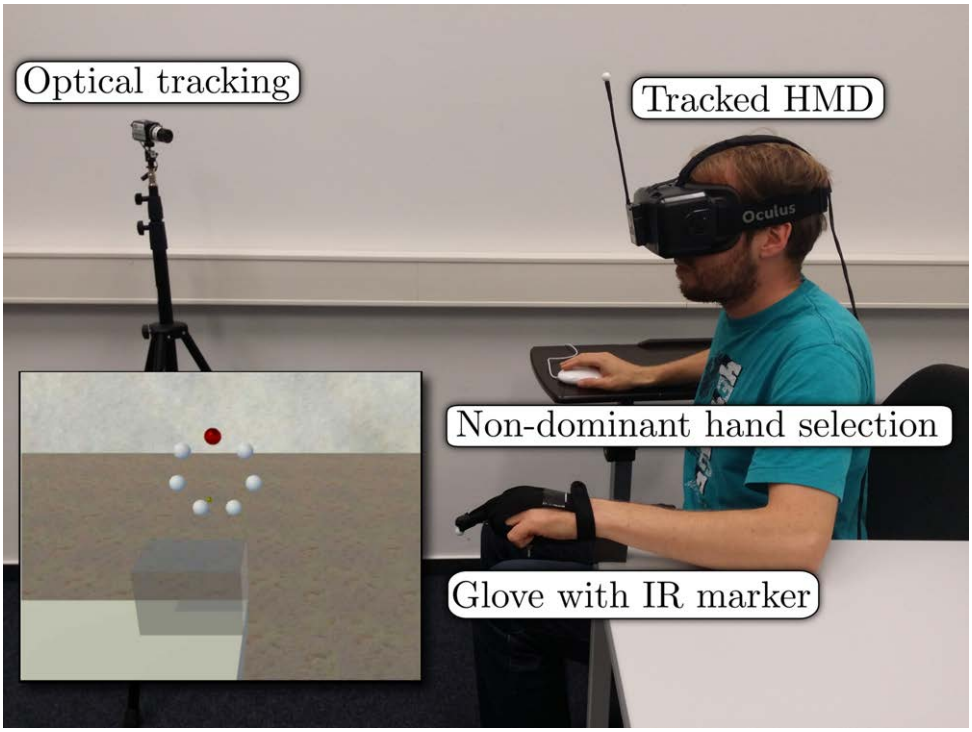


Figure 4.10: Experiment setup: Participant with a tracked Oculus Rift HMD and glove; the inset shows the participant's view of the VE.

the Unity3D engine with an Intel computer with a Core i7 3.4 GHz CPU and an Nvidia Geforce GTX780TI (for more information on the hardware cf. Section 2.5)

4.2.2.3 Methods

Spheres represented the targets in the experiment. All target spheres for one trial were always visible and colored white, except for the current target. The current target sphere was colored red when the marker was outside, and green when the marker was inside to give the participants a visual cue about selection performance. The spheres were highlighted in the order specified by the ISO 9241-9 standard for Fitts' Law evaluations [Int00]. As illustrated in the inset of Figure 4.10, each trial consisted of an arrangement of 7 target spheres, forming a circle with each sphere at the same distance to the participant's joint.

In the experiment, we used a within-subjects, repeated measures 3 (joints) $\times$ 4 (distances) $\times$ 4 (index of difficulties) $\times$ 3 (repetitions) full-factorial design. As described before, we examined the shoulder, elbow and wrist joints. As Figure 4.11 illustrates, we considered three distances (with j being the distance between the joint and the finger in a comfortably stretched pose: $0.618 \times j \mathrel{\widehat{=}}$ short, $0.854 \times j \mathrel{\widehat{=}}$ medium, $1.0 \times j \mathrel{\widehat{=}}$ long) in the experiment. These distances were chosen using the knowledge that the distances between a human's joints can be approximated with the golden ratio [MP07]. The fourth distance condition denotes the boundary technique. When touching an object at the maximal reachable distance it might be advantageous in practical scenarios to include a tolerance range to account for slight overshooting or undershooting in depth (cf. Section 2.4.2.1), hence in this technique, we considered a range stretching from $[0.854, 1] \times j$. We evaluated the ecologically viable IDs 2, 2.5, 3 and 3.5. All conditions were tested three times, and their order was fully randomized. We included 38 training trial (one for each condition), which were excluded from the analysis.

Before the experiment, all participants filled out an informed consent form and received

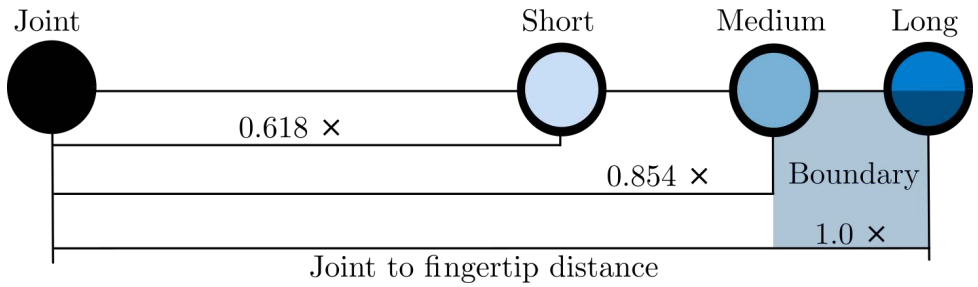


Figure 4.11: Evaluated distances: Taking into account the joint position and the position of the user's fingertip in a comfortably stretched pose, this illustration shows the distances tested in the experiment as detailed in Section 4.2.2.3.

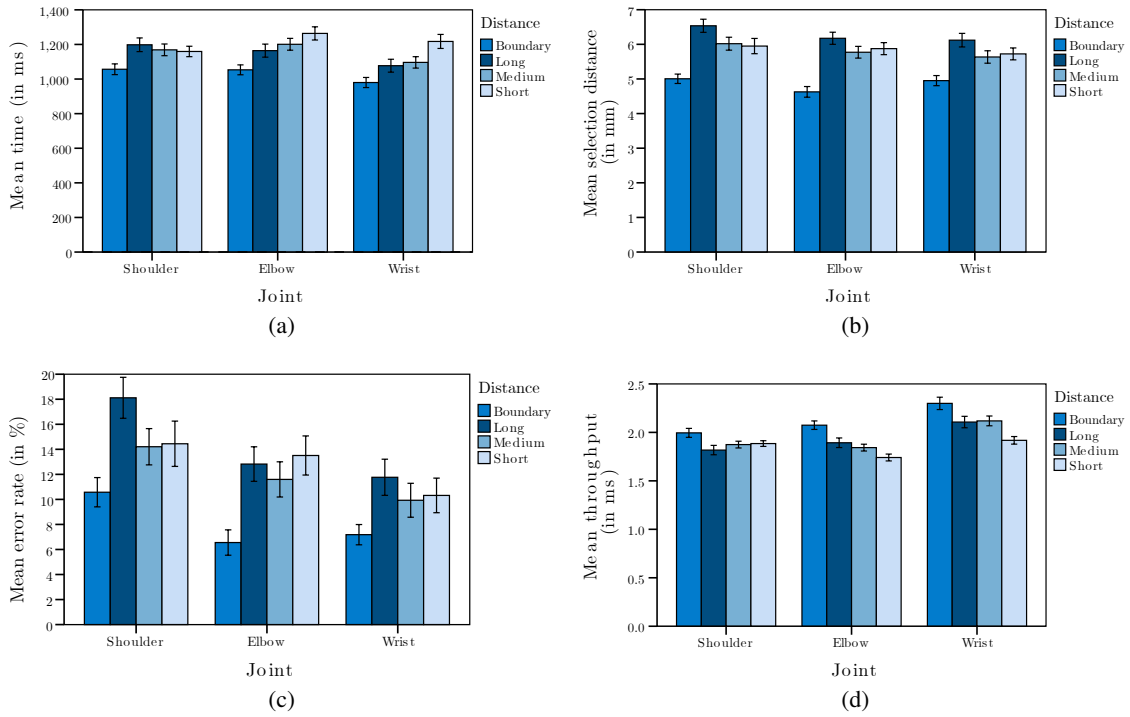


Figure 4.12: Results for Fitts' Law trials with the joint on the horizontal axis and pooled results on the vertical axis for (a) movement times, (b) error distances, (c) error rate and (d) effective throughput. The error bars show the standard error. Red bars show the results for the short distance, yellow for medium distance, green for long distance, and blue for the boundary technique.

instructions on how to perform the task. Furthermore, they filled out the SSQ [Ken+93] immediately before and after the experiment, the SUS-PQ presence questionnaire [Uso+99b], and a demographic questionnaire (cf. Section 2.4.2.4). We further observed the behavior of the participants during the experiment, and debriefed them after the experiment.

4.2.2.4 Results

We analyzed the results with repeated-measure ANOVAs and multiple pairwise comparisons with Bonferroni's correction at the 5% significance level. We confirmed the assumptions of the ANOVA for the experiment data. A Shapiro-Wilk test did not indicate that the assumption of normality had been violated. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated.

joint	M	SD	F	p	η_p^2
wrist	1093ms	202ms	$(2,38)$		
elbow	1171ms	203ms	$=$	$< .001$	.26
shoulder	1146ms	195ms	6.56		

Post-hoc	p
wrist–elbow	$< .001$
wrist–shoulder	.18
elbow–shoulder	.79

distance	M	SD	F	p	η_p^2
short	1214ms	193ms	$(3,57)$		
medium	1155ms	172ms	$=$	$< .001$	.67
long	1146ms	236ms	38.30		
boundary	1030ms	186ms			

Post-hoc	p
medium–long	.99
others	$< .05$

ID	M	SD	F	p	η_p^2
2	861ms	144ms	$(1.41,26.7)$		
2.5	1028ms	184ms	$=$	$< .001$	.95
3	1230ms	205ms	375.89		
3.5	1426ms	247ms			

Post-hoc	p
all	$< .001$

Table 4.1: Statistical analysis for the movement time

Selection Performance

Figure 4.12(a) shows the mean time elapsed until a participant selected a target object for each condition in the experiment. Table 4.1 shows the results of the statistical analysis. Figure 4.12(b) shows the mean error distances, i.e., the Euclidean distance between the participant's fingertip and the center of the target sphere when a participant selected the target object for each condition in the experiment. Table 4.2 shows the results of the statistical analysis. Figure 4.12(c) shows the mean error rate between the times when the participant's fingertip was within or outside of the target sphere when a participant selected the target object, for each condition in the experiment. Table 4.3 shows the results of the statistical analysis. Figure 4.12(d) shows the mean effective throughput for each condition in the experiment. Table 4.4 shows the results of the statistical analysis.

Questionnaires

Asked to rate their level of comfort during the joint conditions with 5-point Likert scales (0 = very low, 4 = very high) (cf. Figure 4.13).

We further asked the participants for each joint condition which distance they preferred with the alternatives boundary, long, medium and short (cf. Figure 4.14). The responses show that most participants preferred the boundary technique and medium distance, followed by the long distance, but not the short distance.

We measured a mean SSQ score of 15.15 ($SD = 17.77$) before and 26.18 ($SD = 29.67$) after the experiment. The results indicate an average increase in cybersickness symptoms with an HMD over the time of the experiment. The mean SUS-PQ score for the sense of feeling present in the VE was 4.57 ($SD = .97$), which indicates a high sense of presence [Uso+99b].

joint	M	SD	F	p	η_p^2
wrist	5.6mm	1.0mm	$(2,38) = 2.78$	.075	.13
elbow	5.6mm	1.0ms			
shoulder	5.9mm	1.2mm			

Post-hoc	p
all	$> .05$

distance	M	SD	F	p	η_p^2
short	5.9mm	1.2mm	$(1.93,36.57) = 55.8$	$< .001$	.75
medium	5.8mm	1.1mm			
long	6.3mm	1.1mm			
boundary	4.9mm	.9mm			

Post-hoc	p
short–medium	.99
short–long	.05
others	$< .001$

ID	M	SD	F	p	η_p^2
2	6.7mm	1.0mm	$(3,57) = 165.14$	$< .001$	.9
2.5	4.8mm	1.0mm			
3	6.4mm	1.2mm			
3.5	4.9mm	1.0mm			

Post-hoc	p
2.5–3.5	.99
others	$< .001$

Table 4.2: Statistical analysis for the error distance

We further collected informal comments during the debriefing after the experiment. Notably, one participant explained the preference of the wrist condition by, *"My arm was too heavy for keeping it up in the air with the shoulder-based interface."* Another participant stated, *"I liked the wrist condition; similar to a mouse."*

4.2.2.5 Discussion

Overall, the results show that the interaction performance highly depended on the placement of targets within arm's reach. In the following, we discuss the findings, focused on the kinematic chain and kinespheres and summarize the main ones as practical guidelines.

Kinematic Chain

We found the significantly highest effective throughput for the wrist joint condition, whereas we found lower effective throughputs for the elbow and shoulder conditions, between which the results showed no significant difference. There are multiple possible explanations for this effect. First, from a biomechanical point of view, a longer kinematic chain leads to more weight and inertia while requiring more muscle exertion and might thus have caused longer movement times [MP07; SM87]. Moreover, the presence of physical support through an elbow or a wrist cushion might have reduced fatigue and hand tremors by shortening the kinematic chain [Gui+15; LBS15]. Additionally, we have to consider that the wrist condition was similar to using a mouse and might have benefited from this similarity.

joint	M	SD	F	p	η_p^2
wrist	9.8%	9.1%	$(2,38)$		
elbow	11.9%	9.1%	$=$	$< .001$	.36
shoulder	14.3%	11.3%	10.85		

Post-hoc	p
wrist–elbow	.33
wrist–shoulder	$< .001$
elbow–shoulder	$< .05$

distance	M	SD	F	p	η_p^2
short	12.8%	11.5%			
medium	11.9%	10.2%	$(2.22,42.11)$		
long	14.2%	10.6%	$=$	$< .001$	.42
boundary	8.1%	7.0%	13.47		

Post-hoc	p
short–medium	.99
short–long	.99
others	$< .05$

ID	M	SD	F	p	η_p^2
2	7.4%	7.4%			
2.5	11.7%	10.3%	$(1.74,32.99)$		
3	11.1%	10.3%	$=$	$< .001$	.6
3.5	16.8%	11.8%	27.95		

Post-hoc	p
2.5–3	.99
others	$< .001$

Table 4.3: Statistical analysis for the error rate

Placement in Kinespheres

As expected, the practically-inspired *boundary technique* with the short tolerance range near the maximal reachable distance showed the overall significantly lowest errors, shortest movement times, and consequently highest effective throughput. This result might be explained by the reduction from 3 DOF to 2 DOF due to the stretched-out kinematic chains and short tolerance range, which additionally eliminated errors along the depth axis over the similar long-distance condition.

While we expected an overall high performance for the *long distance* condition second only to the boundary technique due to the matching stretched-out kinematic chains, our results instead show performance that is very similar to the medium distance condition. Hence, we further analyzed the distribution of errors in the joint-centered kinespheres. With stretched-out limbs, we observed mean selection errors of -1.76 mm ($SD = 1.43$ mm) for the shoulder condition, -2.97 mm ($SD = 1.48$ mm) for the elbow condition, and -2.83 mm ($SD = 1.40$ mm) for the wrist condition as measured from the center of the kinesphere to the target. The negative values might be explained by the participants calibrating their hand position with a genuinely outstretched arm, while they later drew their arm back a few millimeters into a more comfortable pose. Without this systematic undershoot in depth, it appears that the long distance condition would have resulted in higher performance. However, most participants indicated the boundary technique or the medium distance condition as their preference while estimating the long distance (without boundary) and the short distance as the least comfortable.

joint	M	SD	F	p	η_p^2
wrist	2.11 bps	.38 bps	$(1.4, 26.51) = 25.31$	$< .001$	.57
elbow	1.89 bps	.29 bps			
shoulder	1.89 bps	.26 bps			

Post-hoc	p
wrist–elbow	$< .001$
wrist–shoulder	$< .001$
elbow–shoulder	.99

distance	M	SD	F	p	η_p^2
short	1.85 bps	.21 bps	$(1.5, 28.57) = 12.15$	$< .001$	.39
medium	1.95 bps	.25 bps			
long	1.94 bps	.4 bps			
boundary	2.12 bps	.39 bps			

Post-hoc	p
short–long	.93
medium–long	.99
others	$< .05$

ID	M	SD	F	p	η_p^2
2	2.05 bps	.32 bps	$(1.98, 37.61) = 12.43$	$< .001$	.4
2.5	1.95 bps	.34 bps			
3	1.98 bps	.29 bps			
3.5	1.87 bps	.28 bps			

Post-hoc	p
2–2.5	$< .01$
2–3.5	$< .01$
3–3.5	$< .01$
others	$> .05$

Table 4.4: Statistical analysis for the throughput

Guidelines

We found the highest performance for the wrist joint using the boundary technique, which we recommend for practitioners in the field of spatial user interfaces. However, overall performance and user acceptance also showed that placing menus at the medium tested distance is an ecologically viable alternative. We do not recommend placing menus near the long or short tested distances. We found similar performance for the elbow and shoulder conditions, but user preference based on estimated comfort clearly shows that the shoulder joint should be ruled out in practical applications.

4.2.3 Confirmatory Study

In this section, we present a confirmatory study in which we compared a traditional head-centered menu with a joint-centered menu that we designed based on the results of the first experiment. The menus were compared in a practical interior design application. While direct interaction is a excellent choice for many applications in spatial user interfaces, interior design is merely one of them. The concept of JCUIs is applicable for any applications with context changes and tasks that require menus and which can benefit from direct interaction. These could be applications involving planning, construction, design, entertainment or smart environments. In this case, the example of interior design was picked because it offers many opportunities to provide the user a selection of multiple items. Our menu implementation is similar to marking menus, as merely moving into one

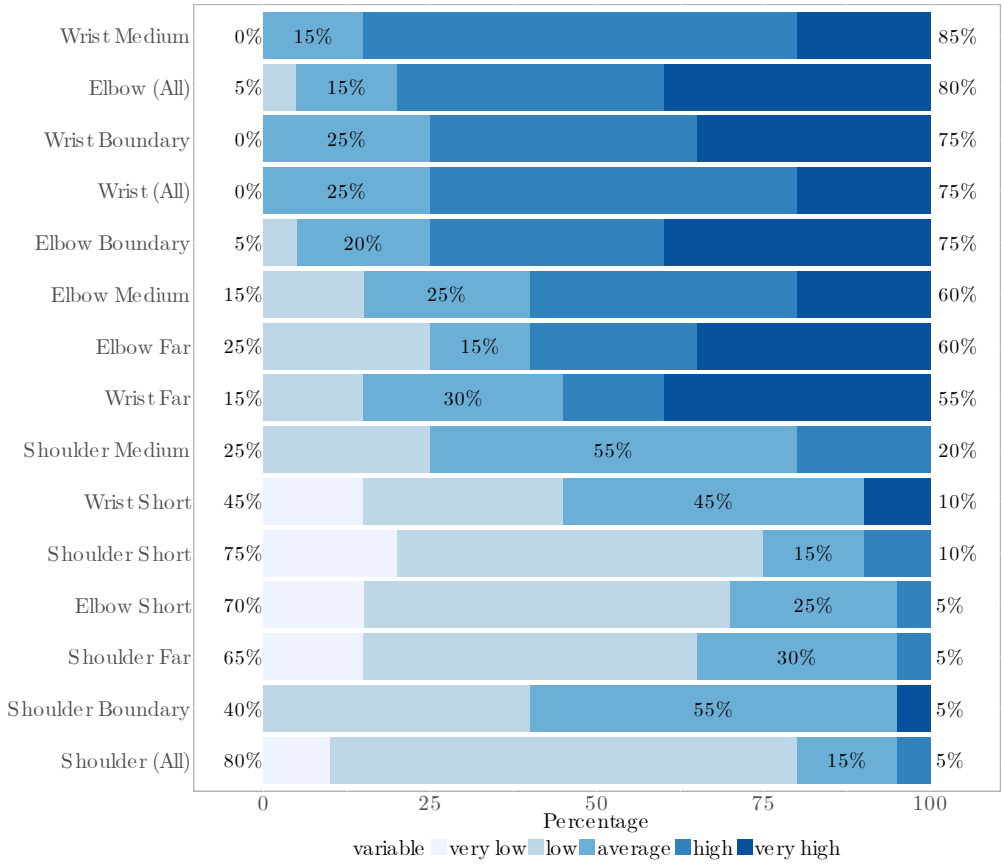


Figure 4.13: The results showing the subjective comfort ratings (x), and the joint and distance condition (y).

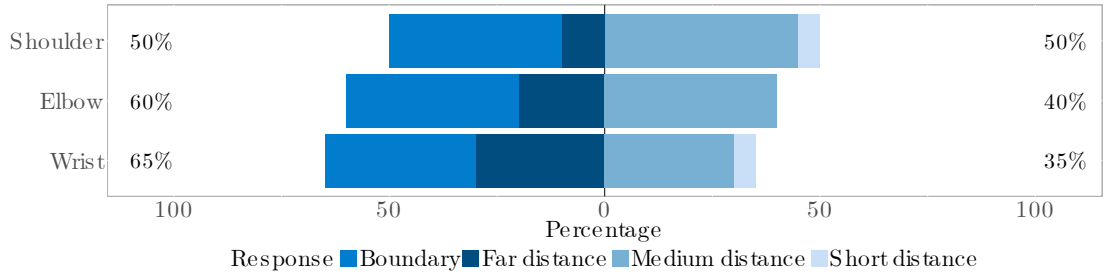


Figure 4.14: The user preference regarding the evaluated distances (x), and the joint (y).

of the elements is enough to trigger a selection and the items are close together [Kur93].

4.2.3.1 Participants

20 participants (2 female, 18 male, ages 19–44, $M = 27.6$) took part in the experiment. The participants were students or experts from computer science or human-computer interaction, and one journalist. Most participants had normal or corrected-to-normal vision (10 no correction, 7 glasses, 3 lenses, 1 with strong eye dominance, 1 with dyschromatopsia (red-green color weakness) and strong eye dominance and astigmatism). 14 of the participants took part in a study with an HMD before. All participants were right-handed. 18 participants reported at least some experience with computer games. The total time per participant was approximately 30 minutes.

Figure 4.15: Illustration of the confirmatory study setup.

Figure 4.16: The virtual room for the confirmatory study.

4.2.3.2 Material

The experiment setup is illustrated in Figure 4.15. Participants were seated in an MWE Lab Emperor Chair 1510 and wore an Oculus Rift CV HMD on their head, IR markers on their wrist and elbow and a haptic ring input device on their index finger [ALS15]. The IR markers and the ring were tracked in 3 DOF using an optical WorldViz PPT X4 tracking system with sub-millimeter precision. Additionally, the ring includes an IMU, allowing 6 DOF tracking of the user's index finger. The living room shown in the virtual scene was rendered in Unity3D on an Intel computer with a Core i7 4 GHz CPU, 16 GB RAM and an Nvidia GeForce GTX 980 (cf. Section 2.5 for more information about the hardware). The in-house implemented interior design application is illustrated in Figure 4.16.

4.2.3.3 Methods

Half of the participants started with the JCUI and afterward completed the second part of the experiment with the head-centered user interface (HCUI) and vice versa. The tasks for the participants were:

1. Familiarize themselves with the spatial user interface through toggling the light by pointing at it and pressing the button on the input device.
2. Moving a table by pointing at it, pressing a button and then pointing to the target location and pressing the button to confirm the position.
3. Changing the wall, ceiling, and floor materials by pointing at them, respectively, and then touching one of the appearing menu buttons of their choice.

In the head-centered condition, the appearing menus were either positioned centered around the head. In the joint-centered condition, the menus were positioned centered around one of the three joints. The menu center depended on whether the participants were resting their wrist, elbow, or

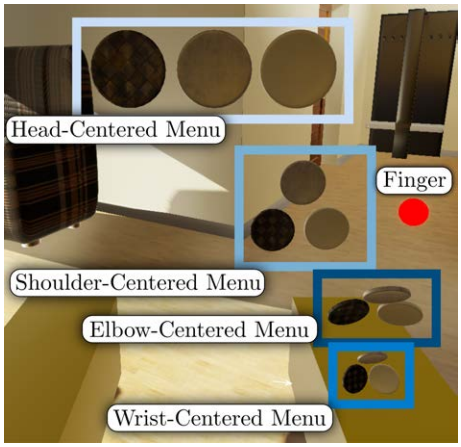
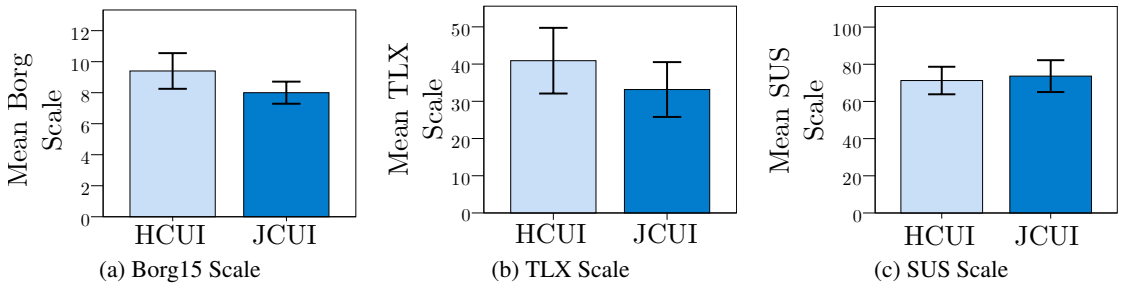


Figure 4.17: The menus that participants would see with annotations. Maximally one menu was visible at a time.

Questionnaire	M	SD	$T(19)$	p	d_{Cohen}
SUS JCUI	73.625	18.308	.373	.713	.373
SUS HCUI	71.25	15.822			
TLX JCUI	33.16	15.740	-1.538	.14	-.344
TLX HCUI	40.91	18.854			
Borg15 JCUI	8	1.522	-2.463	<.05	-.551
Borg15 HCUI	9.4	2.458			

Table 4.5: Statistical analysis for the confirmatory study



(a) Borg15 Scale (b) TLX Scale (c) SUS Scale

Figure 4.18: Results of the confirmatory study.

nothing (shoulder) as shown in Figure 4.17. Next, the participants had to move the couch and then change its color. For this they had to point at the couch, press a button, touch a context button (move or color), and then either move it or color it as described above. We positioned the menus and determined input based on the *boundary technique* described in Section 4.2.2.3.

After the experiment, the participants had to complete different questionnaires comparing the conditions. As a measure of usability, we used the simple usability scale (SUS) [Bro+96]. Additionally, the participants had to complete a NASA TLX [Har06] (cf. Section 2.4.3) and a Borg15 Scale [Bor82] to measure the task difficulty and physical exertion (cf. Section 2.4). Afterward, the participants were asked to provide subjective feedback and answer questions regarding the demographics in a questionnaire.

4.2.3.4 Results

The results were analyzed with a Paired-Samples T-Test for each of the three main questionnaires and are shown in Figure 4.18 and Table 4.5.

Additionally, we analyzed the subjective feedback to the questions whether participants liked the joint-centered approach (15/20 positive), whether they liked head-centered interfaces (12/20 positive), whether joint-centered interfaces could be applied in other domains also (16/20 positive), which technique was less tiring (15/20 joint-centered), and which technique was more fun (11/20 joint-centered, 5/20 indifferent).

4.2.3.5 Discussion

Overall, the feedback was very positive concerning the joint-centered menu, and the results from the Borg15 Scale show that it is estimated as significantly less tiring than head-centered menus. Our results did not show significant differences for the SUS and NASA TLX usability questionnaires (cf. Section 2.4.3).

Improvements

We received qualitative feedback and suggestions for improvement that we plan to incorporate in the next iteration of the design. For instance, the tracking of the different joints with IR markers occasionally caused an unintended switch of the interface to another joint, which we aim to resolve using improved skeletal sensors. Moreover, our participants remarked that the level of fatigue during the few minutes with the interface in the experiment was still tolerable using either head-centered or joint-centered menus, but they estimated that their preference would be shifted towards joint-centered interfaces during more prolonged use. Also, as two participants commented, the joint-centered menu is positioned around the user's arm, which is not necessarily always within the user's FOV. However, due to the matching perceptual and motor space it might be possible to interact even without vision, potentially supported by (vibro-)tactile feedback (cf. [ALS15]).

Implications

Our results imply multiple advantages of presenting 2D GUIs on joint-centered kinespheres instead of traditional head-centered GUIs:

- Physical support, if available, can be used to rest joints without the need to lift the arm up for GUI interaction.
- Users can move the GUI closer to their eyes if need be to increase accuracy and precision during interaction [LBS14a; LBS15].
- Matching joint-centered perceptual and motor spaces support motor training and may even be used for GUI selections without visibility, for experienced users.
- Joint-centered menus located close to the maximal reachable distance ensure that false positives of unintended selections, when moving the hands casually within arm's reach, are highly unlikely.
- Using the boundary technique that combines the positioning of menus on kinespheres with a small tolerance area can significantly improve interaction performance.

4.2.4 Conclusion

In this section, we presented and evaluated joint-centered kinespheres for efficient and comfortable spatial interaction in virtual and augmented reality. We performed a Fitts' Law experiment that showed that the interaction performance largely depends on the placement of menus along the kinematic chain and within the kinespheres centered around the shoulder, elbow or wrist joint. In particular, the effective throughput was highest for the wrist joint and menus that were placed on the boundary of the kinesphere with a small tolerance region for touch input. Based on the results, we implemented a joint-centered user interface and compared it to a traditional head-centered user

interface in an in-house developed interior planning application. We discussed our observations and presented practical guidelines for how to leverage joint-centered kinespheres for the design of menus for spatial user interfaces. Potentially, having a longer practice period may yield even larger performance differences and should be considered for future studies [Sch+12].

There are several possible paths to extend and to adapt the method of joint-centered spatial user interfaces. For instance, kinespheres could be applied at finger level, but also at the legs, trunk or back to incorporate further modalities. Moreover, placing interactive objects on joint-centered kinespheres may require a new class of optimized widgets or layered menu structures. Furthermore since we observed both high efficiency and high comfort in the wrist boundary condition in our experiment, following up on this path may lead towards spatial user interfaces for productive long-time use.

JCUIs can be positioned outside the user's FOV, which may not be desirable for all use-cases and including the necessary tracking to allow these adaptive interfaces may not always be feasible. We developed a supernatural alternative to these JCUIs that may be easier to incorporate into existing applications. In this alternative, we gave a user more than one hand pair, reducing the potential distance between selections by placing a helping hand closer to the target.

4.3 More Than Two Hands

Having shown the benefits of comfortable interfaces, we wanted to provide an alternative to JCUIs (cf. Section 4.2), which is focused on the placement of UI elements within the user's comfortable reach. Taking into account the advantages and shortcomings of virtual hand techniques discussed in Section 2.6.1.4, we propose a novel way of bimanual interaction. This novel interaction evolves around simulating additional virtual hands for a user resulting in *quadmanual user interfaces (QUIs)* or even *Multi-manual User Interfaces (MMUIs)*. Using this approach, homing movements of the hands can be effectively reduced by dividing the interaction space into interaction volumes. With this approach we transfer a solution previously devised for large display environments to IVEs, aiming for similar benefits (cf. [KI08]).

4.3.1 Quadmanual User Interfaces

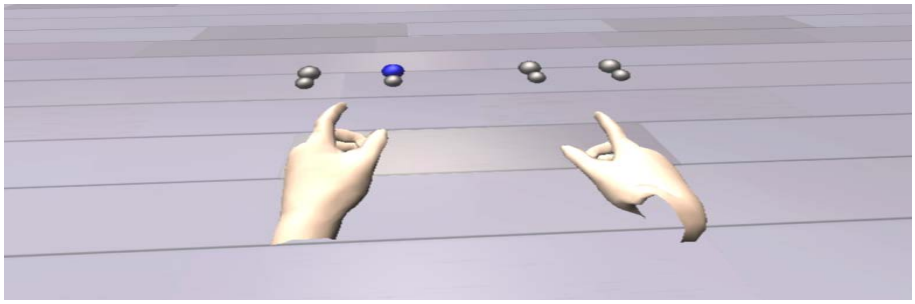
Figure 4.19(a) shows a conventional *bimanual* user interface in which movements of a user's tracked real-world hands are mapped one-to-one to virtual hands. The QUI is illustrated in Figure 4.19(b). With QUIs the user can control two pairs of hands. One hand pair is active, whereas the other pair is inactive and displayed semi-transparently. Using these two virtual hands, it is possible to reduce homing movements of the hands as well as the distance between them and target locations by dividing the interaction space into smaller interaction volumes. This section is based our work previously published as [LBS14b].

4.3.1.1 Controlling Four Hands

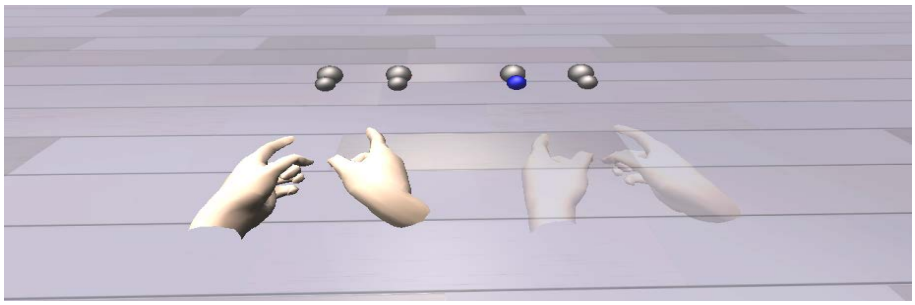
Since the user has only two real hands available, a mapping strategy is required to map movements of the user's two real hands to four virtual hands. Two straightforward approaches are possible:

Simultaneous control The user controls both virtual left hands with their real left hand and both virtual right hands with their real right hand. Although this approach is easy to implement, it has the drawback that all virtual hands are active, even if the user is not focusing on them. Hence, it becomes possible that hands outside the view interact with the VE, even if not intended to.

Selective control The user controls the virtual left hand pair with their real left and right hand, or they control the virtual right pair accordingly. This approach appears to be more feasible, requiring focus only on the active hands while the inactive hands do not affect the virtual space.



(a) Regular Bimanual User Interface



(b) Quadmanual User Interface

Figure 4.19: Illustrations of (a) bimanual user interaction and (b) a quadmanual user interface (The active hands are shown opaquely while the inactive hands are shown semi-transparently).

4.3.1.2 Activating Hands

To determine which two of the virtual four hands should be active, we decided to exploit the gaze direction of the user, approximated by the head position and orientation. Thus, if the user looks to one hand pair, this pair is activated, which is visually indicated by an opaque textured rendering, whereas the inactive hand pair is shown semi-transparently. The active hands stay active until the user focuses on the other hands. In this case, the former active hand pair is visualized semi-transparently, and all virtual hand movements freeze for this pair.

In theory, users should be more efficient using four than two hands, e.g., if we only consider movement distances to spatial targets for selections as predicted by Fitts' Law [Fit54]. However, it is a challenging task to control four virtual hands with only two tracked hands. Hence, the question arises how much additional perceptual, cognitive and motor effort is required for such an unnatural, or in other words, *supernatural* way of spatial bimanual user interaction.

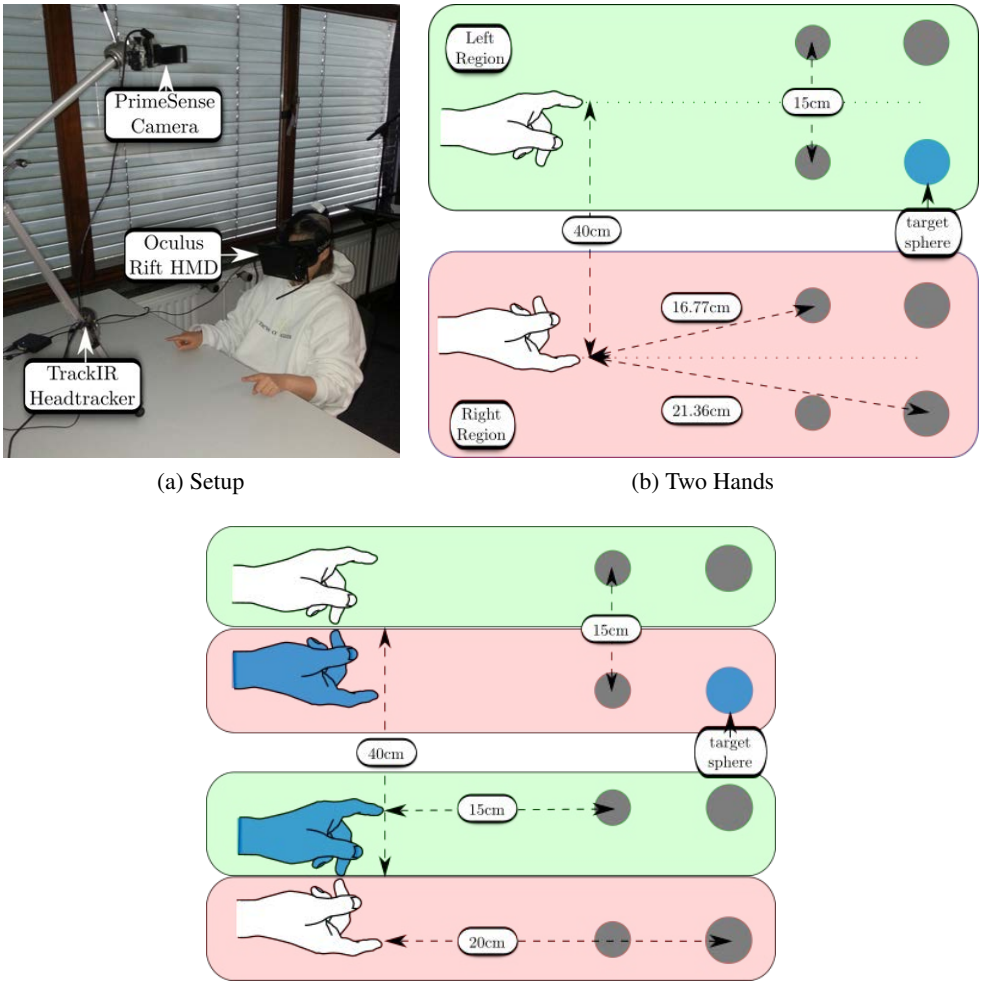
4.3.2 Experiment

In this section, we describe a Fitts' Law inspired experiment, in which we explore how much learning is required until QUIs have the potential to outperform bimanual user interfaces. We analyzed direct 3D selection in the user's arm reach in an HMD environment using two-handed interaction to control bimanual and QUIs. We evaluate the following two hypotheses:

- $\mathbf{H}_1$ 3D selection performance is initially higher with bimanual compared to quadmanual user interfaces.
- $\mathbf{H}_2$ With training, the performance difference between bimanual and quadmanual user interfaces decreases.

4.3.2.1 Participants

We recruited 11 participants for our experiment. Nine of them were male, and two were female (19–45, M = 27.27). One participant was left-handed, the others were right-handed. All of



(a) Setup

(b) Two Hands

(c) Four Hands

Figure 4.20: Illustration of (a) a participant during the experiment, as well as experimental conditions with target spheres for (b) two hands and (c) four hands. The virtual, duplicated hands are dark blue.

them had normal or corrected vision. We measured the IPD before the experiment (M = 6.54 cm, SD = 0.32 cm), and determined their sighting dominant eye (2 left). All participants were naive to the experimental conditions. The duration of the experiment was one hour.

4.3.2.2 Materials

As illustrated in Figure 4.20(a), users wore an Oculus Rift DK1 HMD tracked in 6 DOF a Naturalpoint TrackIR 5 camera. We used a PrimeSense Carmine sensor and the 3Gear NimbleSDK for skeletal tracking and hand reconstruction. The visual stimulus consisted of a 3D scene (cf. Figure 4.19), which was rendered with Unity3D on an Intel computer with a Core i7 3.4 GHz CPU and Nvidia GeForce GTX780TI. For more information about the hardware, please refer to Section 2.5.

Gray colored spheres represented selection targets in the experiment. The current target was colored blue when the user's finger was outside the target and green when the user's finger was inside. Each trial consisted of 8 spheres on a plane in front of the starting positions, four of which were at a close distance and the others at a far distance (cf. Figure 4.20). For the two-hand condition, the targets were divided into a left region for the left hand and the right region for the right hand.

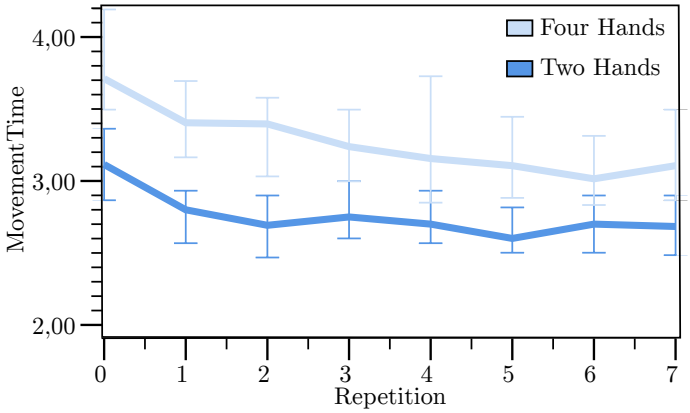


Figure 4.21: Results of the experiment: The y-axis shows the measured movement time and the x-axis shows the repetitions.

The close targets were at a distance of 16.77 cm, the others at 21.36 cm. In the four-hand condition, two targets in a region, at distances 15 cm and 20 cm and IDs of 2 and 3, were assigned to each of the four hands. Taking the different distances into account, we scaled the targets in the two hands condition up to IDs of 2 and 3 to be able to compare the movement time between the two and the four hands conditions. According to Fitts' Law [Fit54], adapting the target size with respect to the distance between selections results in larger targets for longer selection distances, thus resulting in the same task difficulty between the different interaction volumes (cf. Section 2.6.1.1).

4.3.2.3 Methods

To counterbalance, we used a within-subject $2 \times 2 \times 8 \times 8$ design. The two conditions (bimanual vs. quadmanual), two indices of difficulty (IDs) and eight target positions, were uniformly and randomly distributed among each repetition. We repeated all conditions eight times to measure learning effects. Each trial consisted of a single selection of a single target, where participants had to keep one finger within the target sphere for one second. To account for perceptual limitations, we implemented an ellipsoid selection volume as suggested by Lubos et al. [LBS14a]. To make sure all virtual hands were used, we only accepted selections by the correct finger in the correct sphere. Before each trial, participants held both index fingertips within spheres at the starting position, i.e., the initial positions were the same for each trial.

4.3.2.4 Results

We analyzed the results with a repeated measure ANOVA and Tukey multiple comparisons at the 5% significance level. DOF were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated.

The results for movement time, including the selection time of one second, are shown in Figure 4.21. We found a significant difference between the conditions on movement time ($F(1,10) = 26.88$, $p < .001$, $\eta_p^2 = .73$). The mean movement time for the four hands was M = 3.338, SD = .464 and for the two hands M = 2.870, SD = .400. We also found a significant main effect of repetition on movement time ($F(1.79,17.85) = 13.65$, $p < .001$, $\eta_p^2 = .58$). We found a significant effect of hand dominance on movement time ($t(10) = -3.96$, $p < .01$). The mean movement time for the dominant hand was M = 2.97, SD = .330 and M = 3.13, SD = .411 for the non-dominant hand.

We analyzed differences in movement time between the co-located hands and the shifted, virtual hands. The mean movement time for the co-located hands was M = 3.046, SD = .375 and M = 3.509, SD = .633 for the non-co-located hands. We found a significant difference ($t(10) = -3.57$, $p < .01$). We didn't find significant interaction effects.

Additionally, participants answered subjective questions on a scale of 5-point Likert scales

(1 = yes, 5 = no). We asked whether having four hands made the task easier (M = 3.55, SD = 1.21), whether seeing the transparent hands helped (M = 2.55, SD = 1.64), whether it was hard to control four hands (M = 3.82, SD = 1.25), and whether they wanted four hands (M = 2.18, SD = 1.47).

4.3.2.5 Discussion

The results show that for four hands the movement time is initially higher than for two hands. Since we adjusted the targets to the same IDs, the main difference between these selections was the time necessary to process the input, which can be explained by the human action cycle [Nor98]. Compared to simple bimanual selections, an additional step is needed to turn the head towards the target if the corresponding virtual hands are not activated. Nevertheless, this result confirms our hypothesis H_1.

Furthermore, our results show a training effect, as the mean selection times decrease per repetition, which confirms our hypothesis H_2. As the experiment was designed to ensure there was no difference in task difficulty between two and four hands, meaning that the target size was artificially increased for the two hands conditions, the difference would be even smaller with targets of the same size, i.e., selections with four hands appear to be feasible. For time-critical tasks, bimanual or single hand interaction might be more viable, but as a training effect is evident in our results, and since it offers reduced homing times, future studies might show that a quadmanual approach is faster for specific time-critical tasks. By reducing the need for long homing motions to switch from one interactive element to another in a virtual scene, QUIs reduce the tracking space necessary to track the user's hands accurately. Current hand tracking sensors, such as the Leap Motion, have a limited tracking volume. This drawback may be overcome with QUIs.

4.3.3 Multi-manual User Interfaces

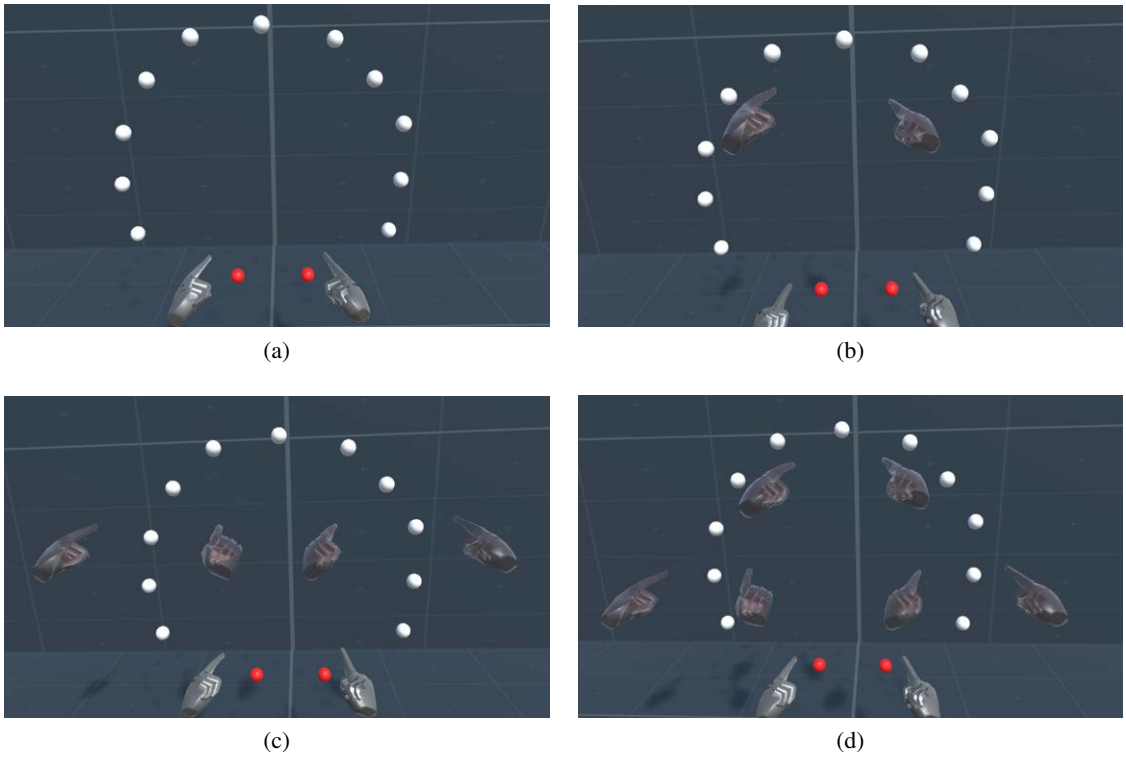
To improve upon QUIs, we developed various switching methods and compared bimanual user interfaces with more than one hand pair. Figure 4.22(a) shows the control setting, a regular bimanual user interface, where the users have to move their real hands for the full distance. Thus, this setting has the longest ballistic phase. Multimanual user interfaces are shown in Figure 4.22(b), Figure 4.22(c) and Figure 4.22(d). In those, the user can operate multiple pairs of hands. However, only one hand pair is active at any given time. The inactive hand pairs are displayed semi-transparently and can be activated by one of the four given switch methods. The active hand pair is then displayed fully opaque. This results in the mapping of the real-life hands to the selected hand pair. Depending on the setting, the number of hands available varies. The general interaction space is divided into smaller interaction sections, each equipped with an inactive hand pair. This method reduces, in general, the distance between targets and another hand pair, shortening the ballistic phase [LBS14b]. This section is based on work currently under review.

4.3.3.1 Switch Methods

We designed four different methods for switching between the present pairs of hands. Neither restricted the hand usage. The user's hands remain freely movable while switching.

Line of gaze We implemented the selection method from [LBS14b], which allows users to select hand pairs by looking at them. However, we extended the setup by adding neutral areas in-between the hands to keep the current selection active. This allows users to freely look around without losing control of the active hands. This removes one of the shortcomings of the original implementation, in which hand pair switches happened too quickly.

Button This switching method is using a button placed on the ground, close to the user's feet. By pushing the button, the selection cycles clockwise one after another through all available hand pairs. This method is heavily dependent on the condition, as one cycle is always as long as the number of hand pairs available.



(a) (b) (c) (d)

Figure 4.22: (a) Control Setting. (b) Two hand pairs. (c) Three hand pairs. (d) Four hand pairs.

Line of gaze & button method This method is a combination of aforementioned methods. The hand pair, which would be activated by the line of gaze, first gets highlighted before getting selected, giving the user visual feedback. To activate the highlighted pair, the user has to press the button.

Speech recognition Using the Cortana API, we assigned keywords to the different pairs of hands. The pair could then be activated by saying the corresponding keyword. Keywords were chosen by the general direction in which the hand pair were located. Note that the entire experiment was conducted in German and therefore the German words for "up, down, left, right" were used. There was a slight input delay ($< 1\,\text{s}$) to recognize the spoken words.

4.3.4 Experiment

Our experiment consisted of 13 blocks. Each blocks was defined by a different combination of hand pairs and switch methods. We had 4 switch methods and 3 settings (2, 3 and 4 pairs of hands), resulting in 12 blocks, and an additional block with just one hand pair without any switch method. After the completion of each block, participants answered four questionnaires, including the Simulator Sickness Questionnaire [Ken+93], the NASA-TLX [Har06], the System Usability Scale [Bro+96] and a variation of the Borg-Questionnaire for perceived exertion [Lin+12] (cf. Section 2.4.4 and Section 2.4.2.4). At the end of our 2.5 hours experiment, the participants answered questionnaires concerning demographics, prior VR-experience and their overall favorite switch method, for short-time tasks, as well as for long-time tasks.

During one block participants had to complete 44 trials in VR. The VR environment consisted of a circle of targets in front of the participant, a dark ground plane, and a dark checkered background. The targets were small, white spheres. While these elements are in the background and do not improve the performance, research indicated that they are helpful for the user, e.g., orientation and

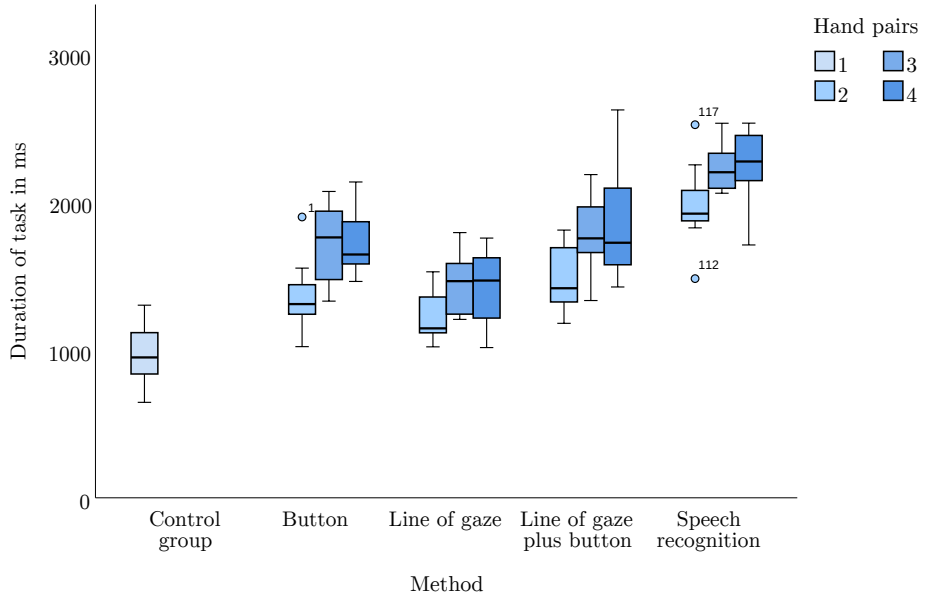


Figure 4.23: The median of the duration of each method.

depicting scale [TS14]. In each block, the participants had three test trials to get familiar with the switch method. Each trial started with the two bottom spheres (the reset spheres) being highlighted in red. The participant then had to touch both spheres simultaneously. After two seconds of continuous contact, another sphere in the circle of targets turned red. The participant then had to reach the highlighted target. Upon completion, the bottom two spheres turned red again, and a new trial started. The circle consists of 11 targets that are arranged symmetrical (excluding the two reset spheres). Throughout one block, each sphere, except the reset spheres, was highlighted randomly four times, resulting in 44 trials. Therefore we had a total of $44 * 13 = 572$ trials per participant. At the start of the experiment, participants were instructed to be as fast and precise as possible.

We evaluated the following hypotheses:

H_1 The participants prefer using multiple hands over only one hand pair.

H_2 Having only one hand pair for completing the task is physically more demanding compared to having multiple pairs of hands available.

H_3 Participants are slower using multiple pairs of hands compared to bimanual interaction.

4.3.4.1 Participants

11 participants took part in our experiment. Ten of them were male, and one was female. The age varied from 19 to 30 ($M = 22.1$). All the participants were right-handed. None of the participants had any medical conditions regarding their shoulders or arms. Nine of the participants had already made experiences with VR, and two participants had not made any previous experiences with VR. Eight participants had never used a Leap Motion before, and three had made experiences with the Leap Motion.

4.3.4.2 Materials

Users wore an Oculus Rift CV1 HMD, tracked by the Oculus Sensor. The distance measured between the Oculus HMD and Sensor was 1 meter. The Leap Motion Controller was attached to the table in the optimal calibrated position for each participant (cf. Figure 4.26). The 3D Environment (cf. Figure 4.22) was rendered with Unity3D on a computer with an Intel Core I7 3.4 GHz CPU and Nvidia GeForce GTX 780Ti, at more than 90 frames per second the entire time. For the button condition, a Griffin PowerMate USB was used. Both chair and table, where the participants were seated at, were adjusted in height to account for different statures. The participants were seated

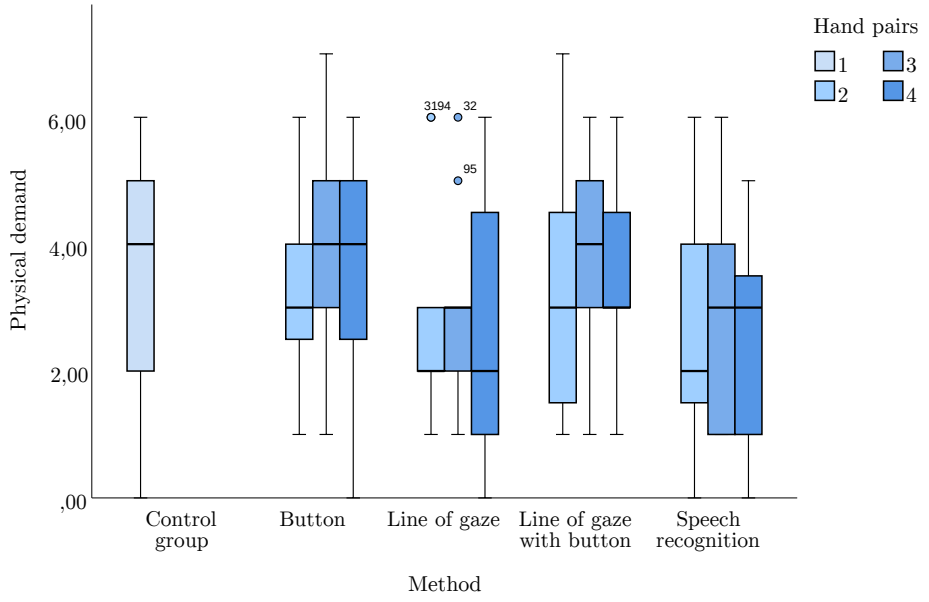


Figure 4.24: The median of the TLX physical demand of each method.

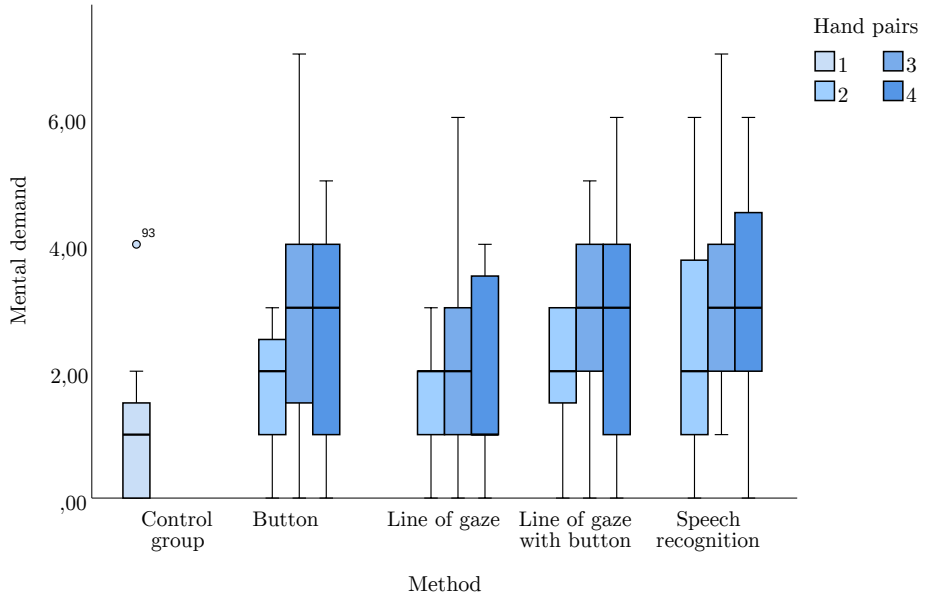


Figure 4.25: The median of the mental demand of each method.

according to the Guidebook of Human Factors [SM87]. With a particular focus on the 90°-bend in their knees, an upright back, elbows on the table, and the Leap Motion positioned at the middle of the participant's palm, while their arms are flat on the table. This brought the participant's hands in the optimal tracking position above the Leap Motion while maintaining a 90°-bend in their elbow during the trials.

4.3.4.3 Methods

For our experiment, we chose a $3 \times 4 \times 44 + 44$ within-subject design. The factors were the number of hand pairs available (setting: 1, 2, 3, 4) and the switch method used for selecting the hand pairs (button, line of gaze, line of gaze & button, speech recognition). The setting and control were ordered according to a Latin square. During each setting, the participants used all switch

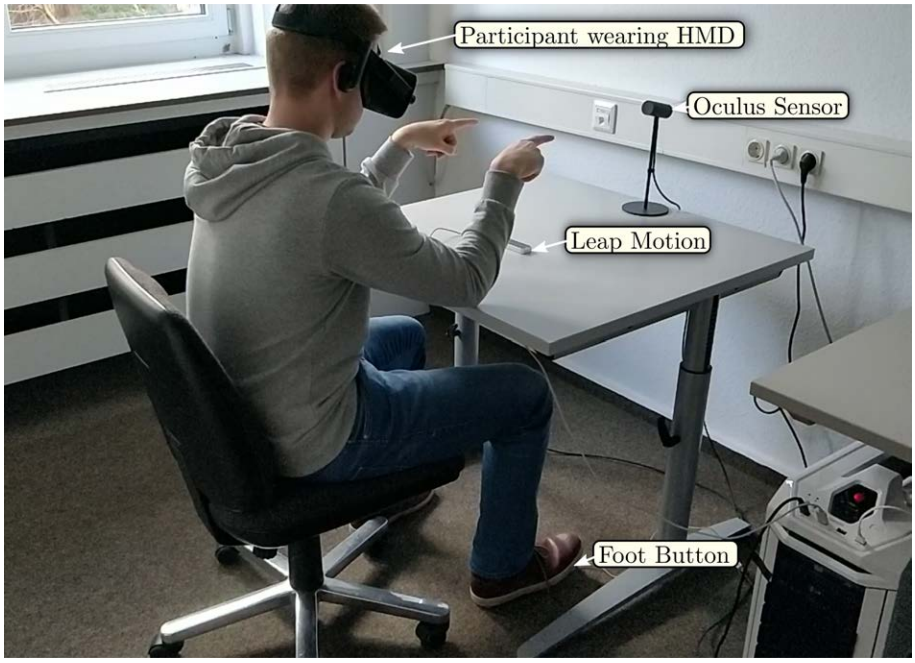


Figure 4.26: Illustration of the experiment setup.

methods. We tested every switch method in every setting. Note that, only after the completion of one setting with all switch methods, the next setting would be tested. For example, participants first used all switch methods in a 4-pair setting, before using them again in a 3- and 2-pair setting. The participants could freely choose which hand pair to activate, with the only restriction being that it was only possible to touch the reset targets with the bottom hand pair.

4.3.4.4 Results

The results were evaluated at the 5% significance level. Since the results were not normally distributed, they were evaluated using non-parametric Two-Way Friedman ANOVAs. The times needed to complete the task were significantly different, $\chi^2(12) = 108.563, p < .001$. We did three Wilcoxon signed-rank tests that showed that the users were slower with two, three and four pairs of hands (independent of the switch method) compared to only one hand pair ($Z = -2.934, p < .05$). The result of the measurement of the time required by the participants to complete the tasks is shown in Figure 4.23. The results from the *mental demand* part of the NASA-TLX showed that there were significant differences ($\chi^2(12) = 19.234, p < .05$) regarding the mental demand for the combinations of settings and switch methods. A Wilcoxon signed-rank test showed that the switch method line of gaze & button had been rated significantly more demanding than the control setting for any number of available hand pairs. The switch method line of gaze had only been rated significantly more demanding in the setting with three available hand pairs ($Z = -2.232, p < .05$). For the *physical demand* component we found no significant differences regarding the physical demand ($\chi^2(12) = 9.199, p = .686$). However, we found significant differences when comparing the Borg questionnaire data ($\chi^2(12) = 27.666, p < .05$). Post-Hoc tests showed that the line of gaze condition was significantly less demanding than other conditions. Compared to the control condition, the a Bayesian T-Test showed no significance, but only anecdotal evidence for H_2 ($T(10) = 2.014, p = .72. BF_{01} = .891$) [LW13]. After completing every task, the participants were asked to fill out a subjective questionnaire containing questions about the switch methods. When asked which switch method they liked most overall, 7/11 of the participants said they would prefer the switch method line of gaze. 2/11 preferred having only one hand pair and no switch

method. One participant favored the speech recognition method, one last the line of gaze & button method. When the participants were explicitly asked which method they would use for short tasks, 6/11 participants said they would choose the method line of gaze, 3/11 said they would choose one hand pair without any switch methods, 1 participant chose speech recognition and another one chose the button method. When the participants were explicitly asked which method they would use for long tasks, 7/11 participants said they would choose the method line of gaze, 2/11 said they would prefer line of gaze & button and another 2/11 said they would choose speech recognition.

4.3.5 Discussion

The results from the subjective questionnaire show that the participants generally preferred using multiple pairs of hands. Specifically for long tasks, participants wanted to use multiple pairs of hands combined with the switch method line of gaze. For short tasks, some participants preferred having one hand pair. This might be because short tasks are less physically demanding and therefore the increased mental demand of controlling multiple hands is not beneficial. This partly confirms our hypothesis H_1. Results from the NASA-TLX showed that there was no significant difference in the physical demand, results from the Borg-questionnaire showed only anecdotal evidence for a difference between the best switch method (line of gaze) and the control condition. Thus we cannot confirm our hypothesis H_2. A further study with more participants would have to evaluate H_2. As seen in Figure 4.25, the mental demand for most switch methods was higher for three hand pairs than for four. We argue that this is because in the setting with three pairs of hands available, the participants had to decide which hand pair they use for targets located in the upper center. For the setting with two and four pairs of hands, this decision was omitted, because there was a section especially assigned to the top center. (cf. Figure 4.22(c), Figure 4.22(d)). The time needed to complete the tasks was higher when participants were to use multiple pairs of hands. The time considerably varied depending on the combination of the setting and which switch method was used. Overall, there was no combination of settings and switch methods where the mean duration of trials was lower than the mean duration of trials in the control setting. This confirms our hypothesis H_3 as well as corresponding with the findings of [LBS14b]. Whereas the lower speed indicates a lower performance, the user's preference for multiple hand pairs suggests that further evaluation of MUIs should still be considered.

4.3.6 Conclusion

To evaluate our switch methods, we let participants test the methods in each setting. From our analysis, we conclude that the best switch method is line of gaze, as it was the fastest overall and was picked as the favored switch method by most of the participants. We evaluated physical and mental demand for each switch method and compared it with a bimanual user interface. This comparison indicated that controlling multiple pairs of hands is mentally more demanding.

Further research in this topic could take a deeper look at how to divide the interaction space into smaller sections, for example, by dynamically creating more pairs of hands, so it is not necessary to have them in fixed positions in the scene. Our proposed switch methods could be extended by new and different methods, but also be enhanced by, i.e. using a more ergonomic foot button. For example, a gas pedal, where the users can rest their foot on could reduce the physical exertion, especially for long tasks. Even though there were no significant differences in physical demand, participants preferred using multiple pairs of hands and having a higher mental demand than having only one hand pair available and having a lower mental demand. Once more, we want to remark that when asked which switch method the participants would choose for long tasks, not a single participant wanted to use the regular hand pair. This leaves a very promising foundation for future research on the topic of MUIs.

4.4 Summary

In this chapter, we investigated comfort for selection in 3DUIs and how to design them to be more comfortable for long-term use. In the first section (4.1), we evaluated a standing user's interaction space to identify which regions therein are more or less comfortable. Whereas we could not find a difference there, we conducted a second experiment, directly comparing 3D selection with an armrest to selections without. Therein, we could show that having a more comfortable setup can increase performance for long-term use.

We used these results to develop JCUIs, which we present in Section 4.2. JCUIs are a concept of creating more comfortable 3DUIs by placing UI elements adaptively depending on which joint of a user's arm has physical support. Therefore, we first had to conduct an experiment to determine comfortable selection distances between a joint in the arm and the fingertip. In a confirmatory study, we could show that JCUIs are preferred by users and cause less fatigue.

Considering JCUIs don't necessarily keep UI elements within a user's view, we developed an alternative, by decoupling virtual hands from their physical counterparts. We placed first two and then more hand pairs in the VE, which the user can switch between. This concept of Quadmanual- and Multimanual UIs is described in Section 4.3.1.

The results described in this chapter show the importance of comfort in 3DUIs, which was previously neglected in research. Using these results, 3DUI developers can create applications allowing longer use than previously possible.

5. Enhancing Traveling and Exploration of Virtual Worlds

Abstract

As one of the fundamental tasks of 3DUIs, navigation encompasses traveling to a target, as well as the cognitively demanding stage of navigating through a VE [LaV+17]. In this chapter, we investigate supernatural means of traveling through environments. First of all, we develop new, supernatural, leaning-based steering techniques (cf. Section 2.6.2) and evaluate them in user studies. The goal of the first part of this chapter is not to find subtle, true-to-nature navigation techniques, but offer user overtly supernatural means of exploring a VE [Dib+16; Sch+16]. During the second part we investigate walking travel techniques and determine how cognitively demanding RDW is [BLS15]. We try to improve upon existing RDW techniques by employing a secure area without redirection and overt warning cues to take the user's fear of colliding with physical objects they cannot see [LBS14c]. Finally, the exploration of VEs is often limited by hardware limitations, such as a low FOV and we investigate whether adding low-resolution peripheral displays encourages users to explore a VE [Lub+16a].

5.1 Leaning Based Flying

Flying in VEs is a supernatural technique, as humans are incapable of flight. We investigate whether leaning-based steering can be used efficiently for flying in VEs. First, we will describe a seated, leaning-based steering technique to control a supernatural, fictional creature [Sch+16].

Most of the travel techniques discussed in Section 2.6.2 are general-purpose and often aim to make travel subtle and similar to real travel, such as walking. When goal is to increase immersion, the application context has to be taken into account. For instance, while virtual airplane metaphors are well suited for navigation in virtual worlds inspired by the real world, their applicability for virtual fantasy environments remains in question. Whereas in an airplane context, a joystick would be a reasonable input device, we assume that another technique would suit the supernatural context better. We decided to develop a travel technique that fits this supernatural context, based on previous research regarding leaning-based travel techniques [Kru+15a](cf. Section 2.6.2 for more related work). Leaning-based techniques, or more generally, weight-shifts, are used to control various

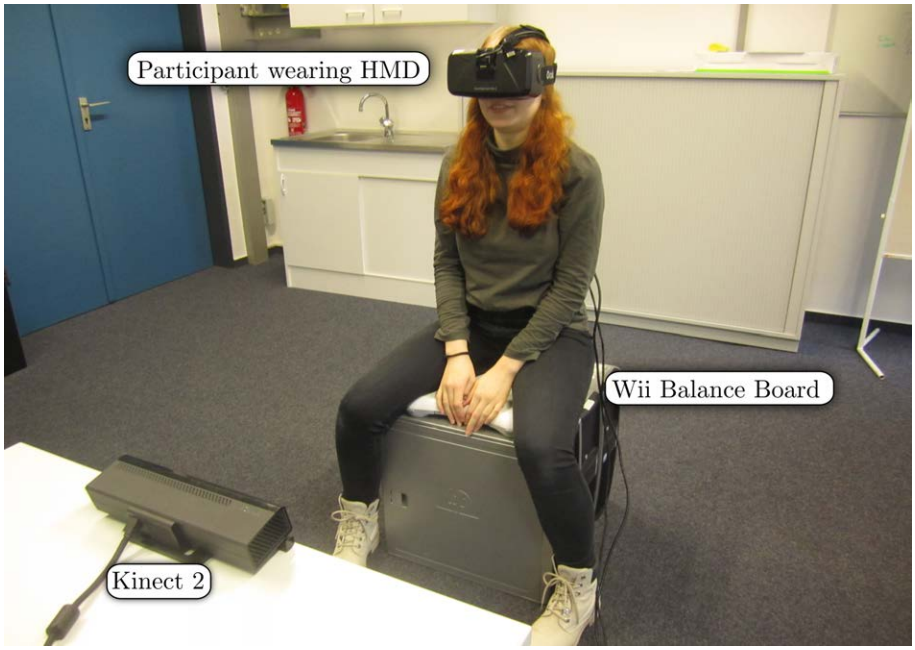


Figure 5.1: Illustration of the dragon rider setup with the two control methods based on either the Wii Balance Board or the Microsoft Kinect 2 sensor. The blanket was omitted for the illustration to show the Wii Balance Board.

Figure 5.2: The user's viewpoint during the experiment.

flying vehicles, such as hang gliders, and animals, such as horses.

In this section, we introduce a virtual flight interface that allows the user to travel through a virtual fantasy world by riding on a dragon controlled by leaning one's upper body while seated. In this context, we present and evaluate two approaches to track and map the leaning state of the user's upper body to steer the virtual dragon flight.

5.1.1 Dragon Rider Flight Technique

The hardware setup of our dragon rider interface (shown in Figure 5.1) consists of a seat with a Wii Balance Board mounted on top as well as a Microsoft Kinect 2 facing the seat from ahead. The Kinect 2 allowed us to track the upper body of a seated user at different leaning angles. In our current implementation, users are wearing an Oculus Rift DK2 HMD. The HMD is connected to a graphics workstation running the Unity3D game engine for rendering and movement control in the

VE (cf. Section 2.5). Additionally, we use wireless headphones to provide audio feedback from the VE. In the experiment described in Section 5.1.2, the Wii Balance Board and the seat below were covered by a blanket hiding the board from the user's view. Either the Balance Board or the Kinect were used to track the user's leaning. We also used the Kinect to track the user's skeleton and detect an additional gesture in our dragon rider interface: When the user performed a fast up and down movement of the hands, similar to the gesture used to slap a horse's reigns, they triggered a speed boost.

5.1.1.1 Calibration

Before starting the experiment, we calibrated the system according to the participants maximal comfortable leaning angles and mapped these on a scale from -1 to 1 for the transverse axis and the sagittal axis of the user's body. For the balance board, we used the distribution of the user's weight on the four sensors such that a shift of the entire weight to one side of the board corresponded to a value of -1 or 1 depending on the side of the board that the weight was shifted to.

5.1.1.2 Steering Methods

Our approach to steering the dragon's movement direction while flying through the VE is based on the leaning direction of the user's upper body. In a way, the dragon is a virtual vehicle controlled by the user. The user leaning to the left and right in the real world leads to the dragon flying in the same direction in the virtual world. In doing so, a leaning value of -1 or 1 corresponds to an angular velocity of 75 $^\circ$/s, the values between are linearly interpolated. We combine a yaw and roll movement to make it appear as if the dragon is leaning into the turn. We differ from the joyman metaphor and use leaning forward and backward to control the dragon's upward and downward movement via a pitch rotation in the corresponding direction (cf. Section 2.6.2.4). The dragon in our implementation has a constant forward flying speed, except for the speed burst by slapping the reigns. This gesture leads to the dragon tripling its speed for three seconds and then slowly decelerating back to normal traveling speed.

5.1.2 Experiment

The Wii Balance Board is not prone to visual interferences, such as the Kinect. However, the Balance Board cannot differentiate between a user shifting to the left or right or leaning left or right. To find out whether the input device results in different usability, we decided to compare these against one another in an experiment, as well as to test the travel technique for usability.

5.1.2.1 Participants

After removing 13 participants from further analysis who quit the experiment prematurely due to cybersickness, the sample for our study consisted of 17 participants, 11 male and 6 female (age from 18–37 years, average 27.5 years). Participants were students or professionals in the fields of HCI or computer science. Two of our participants reported color blindness, and the others had normal or corrected-to-normal vision. Six participants wore glasses, and one wore contact lenses during the experiment. None of the participants suffered from a known disorder of equilibrium or binocular vision disorders. 11 participants had already participated in other studies using an HMD. The experiment was counter-balanced, meaning that half of the users started with the Kinect condition and the other half started with the Balance Board condition, without being informed which control mechanism would be used. Participants wore the HMD for approximately 20 minutes. They were allowed to take breaks or terminate the experiment at any time.

5.1.2.2 Material and Methods

The visual stimulus of the experiment was rendered using the VR-mode of the Unity3D game engine and consisted of a vast landscape with tree-covered mountains. We placed 29 waypoints in the landscape. The waypoint was constructed of an egg that we used as a target and a tower

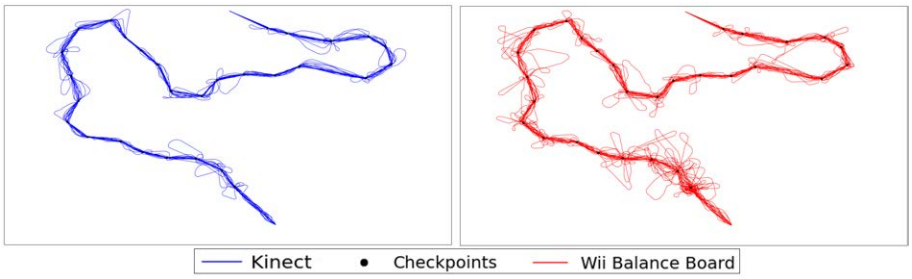


Figure 5.3: The routes flown in the different conditions (top-down perspective).

underneath for better visibility. Waypoints were set on a roughly circular path spanning across the map. Distances as well as differences in height and curvature between waypoints varied. This way, participants were required to apply different degrees of steering in every direction to reach each waypoint without backtracking. The eggs were lit one at a time to indicate which one should be aimed for next and were removed from the landscape when they were passed. The user's representation in the VE was a 3D model of a human moving according to the postural data received from the Kinect, placed on the animated model of a 3D fantasy dragon in flight. Looking forward in an upright sitting pose, participants could see the back of the dragon's head, as shown in Figure 5.2. With head movements tracked by the Oculus Rift DK2, users were able to look around in the VE without changing the direction of their flight. To enhance the immersion and tune out ambient noise, audio feedback was used. The audio feedback consisted of a steady wind sound and a sound of flapping wings, which grew faster when the dragon flew with increased speed and stopped when the dragon was gliding when flying downwards. A roaring sound was played at random times and flying through a checkpoint was confirmed by a popping sound.

5.1.2.3 Procedure

We used a within-subject design in which we compared both control methods: Kinect and Balance Board. Following pre-questionnaires, a short explanation of the controls as well as the calibration process, the participants got accustomed to the controls in a short phase of flying without any given task. After this training stage was over, they were asked to collect all eggs in the landscape by flying through them as fast as they were able to, using the speed gesture at any time they wanted. To control the dragon in the Balance Board condition, it is not necessary to lean the whole upper body. Since the Balance Board detects weight distribution, the user can achieve full results by a slight movement of the lower body. To avoid this method and to test both conditions as leaning-based control methods, we did not inform participants about the used measuring condition. Before and after the experiment we asked participants to fill out several questionnaires, which are explained below.

5.1.2.4 Results and Discussion

The recorded flight paths in Figure 5.3 show many loops around waypoints in the Balance Board condition since participants missed them and had to retry. This resulted in longer flight times.

We measured symptoms of cybersickness before the experiment and after each of the conditions using the SSQ (cf. Section 2.4.2.4). By computing the differences in the increase of cybersickness in both conditions, we found a similar increase for the Balance Board condition ($M = 22.66$, $SD = 26.69$) and the Kinect condition ($M = 26.62$, $SD = 25.94$). A Wilcoxon Signed Rank Test showed no significant differences between Balance Board and Kinect conditions ($Z = -.189$, $p = .85$). As most participants seemed to prefer the comfort levels of the Kinect condition in their informal statements and additional questions during debriefing we also looked at the data we gathered from the participants that were unable to finish the flights. When asked to comment on the

Test	Result	Balance Board condition	Kinect condition
NASA TLX	Z = -2.406, p < .05, r = 0.38	Mean = 5.5, SD = 1.16	Mean = 4.5, SD = 1.38
SUS PQ	Z = -.686, p = .49	Mean = 4.5, SD = 1.16	Mean = 4.29, SD = 1.23
Intuitiveness	Z = -2.216, p < .05, r = 0.38	Mdn = 6	Mdn = 5
Ease of use	Z = -3.371, p < .001, r = 0.58	Mdn = 3	Mdn = 6
Feeling of control	Z = -3.370, p < .001, r = 0.58	Mdn = 5	Mdn = 6
Fun	Z = -2.648, p < .05, r = 0.45	Mdn = 6	Mdn = 6

Table 5.1: Questionnaire results.

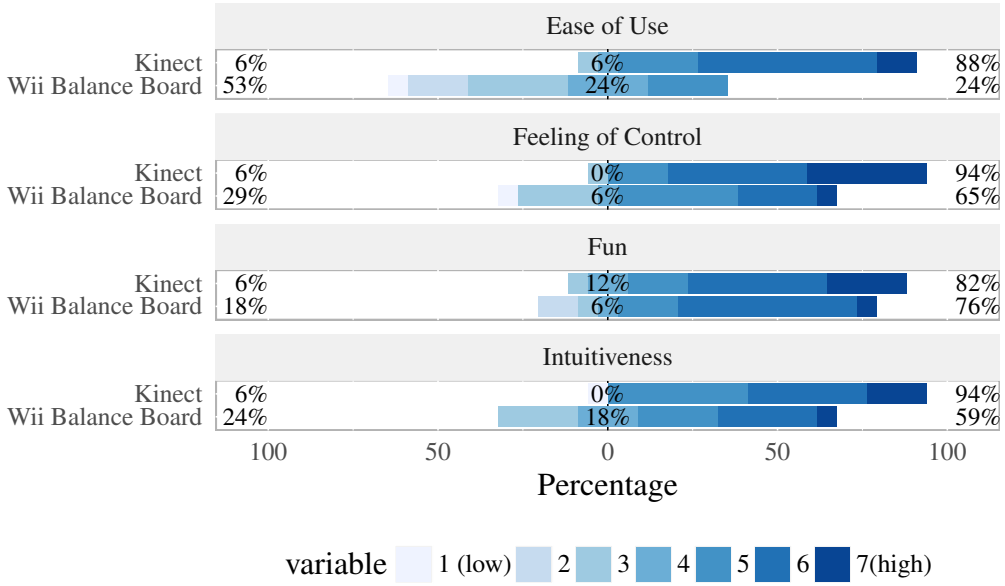


Figure 5.4: Plot showing the average scores for the usability factors that participants were asked to rate after each condition in the experiment

different conditions, many participants explained that the right and left turns in the Balance Board condition felt almost like a discrete button instead of like an analogue value. Some participants also had difficulties in achieving the change of altitude needed to complete the route, which they stated were important factors inducing cybersickness. Apart from the informal comments mentioned above, we did not find clear correlations between cybersickness and the types of control movements in the dragon rider interface, but we plan to address these points and reevaluate them for the next iteration in the human-centered design process of the interface. We assume that the high count of participants who quit the experiment early may be connected to either overall sensitivity to flying or to the rather high turning speed of 75 °/s. These assumptions would have to be investigated in further research.

We performed a Wilcoxon Signed-Rank Test to compare the NASA Task Load Index (TLX) scores between the Kinect and Balance Board conditions (Table 5.1). We asked the participants to

rate the intuitiveness, ease of use, feeling of control and their fun on 7-point Likert scales (1 = very low, 7 = very high)(cf. Figure 5.4 and Table 5.1). The varying levels fit our previous observation of the usability in the different conditions. Nonetheless, the perceived presence in the conditions showed no significant difference, which is likely caused by using the same visual and auditory stimuli for both conditions.

5.1.3 Conclusion

In this section, we presented a novel dragon rider interface and described two methods to measure the leaning state of the user's upper body, which we used to control the movement direction of a virtual fantasy dragon. We compared the two conditions which revealed that using the Kinect to measure the leaning angles of the upper body shows more promising results than using a Wii Balance Board to measure shifts in weight. This is shown not only in the speed at which participants were able to complete a 3D navigation task but also in the overall comfort levels of our participants during the different stages of the experiment. However, our results also show that such leaning-based interfaces in VR, when used for 3D flying, can induce very high levels of cybersickness. This has to be considered in future research by focusing on approaches to reduce the conflicting motion cues, e.g., by including stationary reference frames for the users to rest their eyes upon. An example could be a fixed cockpit around the user that does not move by itself as the breathing and flying dragon in our implementation did.

5.2 Suspended Flying

Similarly to the silver surfer technique [WL11], another inspiration for the development of novel, supernatural travel techniques came from the comic-book and movies about Superman [Kru+15b] and Iron Man [Dib+16]. Whereas the dragon-rider interface in Section 5.1 was a seated experience, the project described in this section is built around suspending users from the ceiling, by using a climbing harness. While various concepts of flying were presented in science-fiction literature and movies, there are few scientific evaluations of these control concepts. We first describe the hard- and software setup for our suspended flying setup. In the first experiment, we describe the two flight techniques, one inspired by gliding birds and the second by Superman's iconic way of flying with one outstretched arm.

Partly outside of the research context, several projects tried to create supernatural flying experiences. These facilities can be applied to mental and physical training purposes or leisure activities. In each case, they open a new field of research for HCI and VR.
Figure 5.5(a) shows a setup with VR glasses and an IMU for orientation tracking of the user allows for a flight experience likewise Tony Stark's (aka Iron Man) everyday life [Dre10]. Participants are hanging inside a hang glider harness in a small crane. Their bodies rest on a wing of an unmanned aerial vehicle that was tracked using a Nintendo Wii remote. To increase the degree of immersion, they added a wind machine. An HMD shows live images from the Google Earth Flight Simulator. The ICAROS VR project is part of a diploma thesis [SS17] (cf. Figure 5.5(b)). It is designed for physical training and leisure experiences with VR. Integrated within the joints of the movable parts are 16 servo motors for actuation or tracking. Data from these motors is extracted as control signals for the VE. Four additional motors are placed in the base of the system. The system is used with an HMD and headphones. Finally, Birdly is a bird simulator, using a rig comparable to an actuated comfortable table [Rhe+15](cf. Figure 5.5(c)). Users place their hands on wing-like plates with buttons. The wings are connected to joints, allowing users to utilize them as their movable wings. The visual stimulus is shown through an Oculus Rift DK2 and shows a bird's body as an avatar, which is controlled by the user's movement. The experience is enhanced by an adjustable propeller in front of the participant's head. Glomberg et al. developed a hang-glider

(a) (b) (c)

Figure 5.5: Current next generation flight simulation systems. (a) *The Iron Man Flight Simulator*, with mounting facility composed of a hang glider harness and a small crane, is built by a team of VR enthusiasts in Israel [Dre10].(b) *ICAROS VR* prototype utilizes an IMU from a smartphone to extract control signals. Image adapted from [Sto15].(c) *Birdly* simulator uses actuators as input device for controlling the virtual flight. Image adapted from [Rhe+15].

simulator, which also required the user to put on a suit and be suspended [Glo+14]. Movement within the hang-glider resulted in convincing reactions of the virtual hang-glider. More extensive instrumentation can provide a user's vestibular sense with an even better flight sensation [Mie+16]. However, considering the visual sense's dominance, it is possible to achieve a high degree of presence with appropriate displays (cf. Section 2.4.2).

Based on these setups and the related work discussed in Section 2.6.2, we decided to create a similar setup and use it with a focus on the flight controls.

5.2.1 System Setup

Our supernatural flight setup, dubbed *Zero-Gravity*, was built around an existing set-up of hanging metal trusses in our lab, which we extended with a motorized mounting system for humans limited to payloads up to 100 kg. A Kinect 2 camera sensor (cf. Section 2.5) is utilized for body tracking of the participant. The visualization is realized with an HMD with an integrated IMU. A powerful graphics workstation renders the 3D scene in real time depending on user's input by body movements with only low latency (cf. Section 5.2.3). To limit the risk of injury and to avoid the need for special security arrangements (cf. [Deu11]) we limited the maximum suspension height for participants to 2 m.

5.2.1.1 Hardware

In the following paragraphs, we describe the setup in more detail (cf. Figure 5.7). The required equipment for the whole simulator setup splits up into three different groups. The mounting part brings the participants into a comfortable and trackable position. The tracking hardware is used to steer the flight of the participants and to give them the possibility to look about. The visualization components are utilized to present the rendered environment to the user in real time.

Mounting

The user has to wear a climbing harness (Skylotec CS 4), which is professional industrial climbing equipment that is very robust, comfortable and offers various adjustments to the user's body size. Using the integrated hoops, we attached the climbing harness to three ropes using carabiners. These ropes are conducted over pulleys mounted to the metal trusses at on laboratory's ceiling. The other ends of the ropes are connected to a motor-winch attached to the metal trusses. The teleoperable motor winch is capable of lifting 400 kg in its current configuration. To further increase the comfort in our setup, we added foam plastics at the inside of the harness. However, these harnesses are

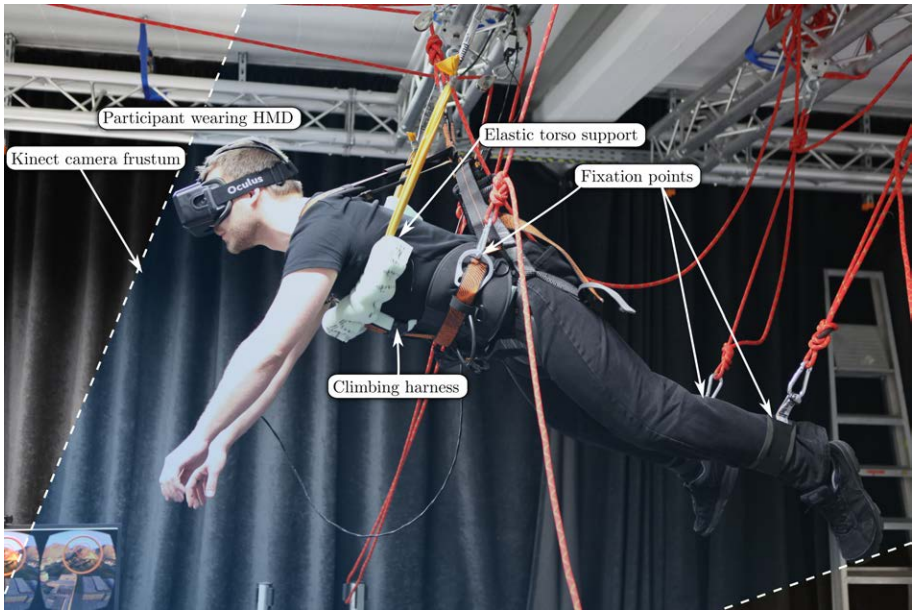


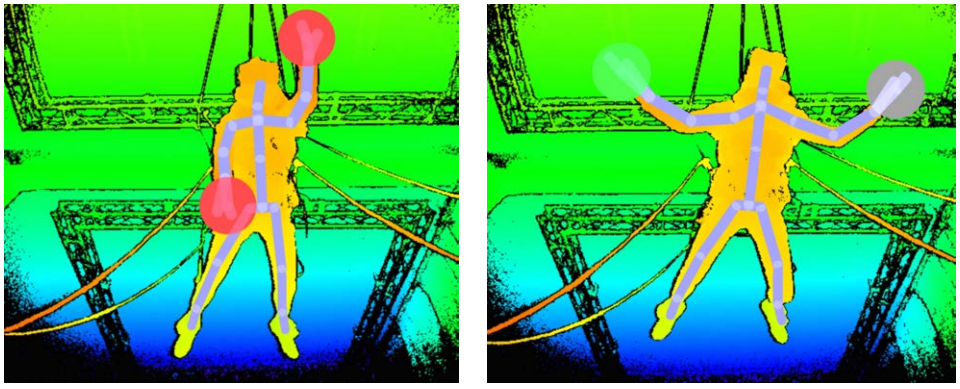
Figure 5.6: Basic setup of the Suspended Flying Simulator: Participants' weight is distributed among three main fixation points, one at the torso and two at the hips. Two secondary points at the ankles provide a more comfortable position for the user. Users are wearing a Oculus Rift DK2 and their body movements are tracked by a Microsoft Kinect for Xbox One standing on the ground in 3D-printed stand in an up-facing direction.

Figure 5.7: Components of the mounting system. From left to right, there is the motor winch mounted horizontally at on the truss, the connection between the carabiner of the winch and the climbing ropes, the climbing harness, the Thera-Band sling for load reduction at the chest and the fixation straps at the ankles.

designed for use in upright or dorsal positions. For our flying setups, we needed a harness that is applicable for face-down positions and decided to dress the harness with the front part in the back. Additionally, we added a padded sling with a robust elastic rubber band (Thera-Band gold) to optimize the weight distribution at the upper part of the body. This gives the users better movement capabilities and reduces the back muscle stress. Finally, for an equal distribution of the body weight, we used straps at the ankles for the final setup.

Tracking

The direction of flight in the simulation is controlled by a Kinect 2 sensor on the ground and the Kinect for Windows SDK 2.0. In our experiments, it required a minimal distance of 1.5 m to achieve stable recognition of bodies. Possibly due to the ropes being recognized as limbs, the tracking is more reliable from the ground than from the ceiling. The head orientation tracking is done with the integrated IMU of the HMD.



(a) (b)

Figure 5.8: Recognition of the hanging body is done by the Kinect's SDK skeleton tracking. (a) Tracked body in Superman method. (b) Tracked body in Bird method.

Visualization

The visual stimulus is shown using an HMD, which was an Oculus Rift DK2 during the development time (cf. Section 2.5). The 3D visualization of the rendered scene is shown on the display of an HMD to the user. We decided to use an Oculus Rift DK2 because of, at the time of development, its low weight, relatively good display quality and its integration into state-of-the-art game engines. In the VE, subjects make use of the integrated relative head tracking to look around. During operation, the rendered scene is additionally displayed on a flat screen connected to the graphics workstation, allowing the supervisor to monitor the scene. Whereas the hardware setup is independent of the software, we used the Unity engine in our experiment. The visual representation was realistic, to invoke a sense of presence. The user's virtual body is implemented as an object in the VE, which can collide in the VE, preventing users from flying through obstacles, such as buildings. Collisions slow down users and urge them to avoid further collisions.

To determine the flight direction, we used the user's body, tracked through the Kinect 2 with the Kinect SDK [Kni13; MZC12]. More specifically, we used the skeletal tracking, which proved robust and fast enough during testing. We employed exponential smoothing on the binary hand state for speed adjustments. No additional smoothing or filtering was required during our experiments.

5.2.2 Flight Control

Within the first iteration of our setup, we implemented two supernatural flight control methods (cf. Figure 5.8). The *Superman* method is comparable to steering by pointing with one arm. The more comprehensive *Bird* method requires the use of both arms and additionally movements of the upper body. Both methods control only the direction of movement by changing two rotational degrees of freedom and the forward motion speed. Pitching orientation influences flying upwards and downwards. Yawing orientation influences turning to the left and the right. Both control methods offer the possibility to change the speed by simple hand gestures of the active arms. Opening the hands causes a deceleration and closing hands an acceleration.

5.2.2.1 Method 1: Superman

This control method, inspired by the character Superman's iconic flight with one outstretched fist, uses solely tracking data of the participant's dominant arm. Depending on the pointing direction, pitch and yaw angles of the controller are modified. The control output is relative to a precalibrated neutral position. Afterward, a deviation from the neutral position is interpreted as steering. Moving the fist up results in a rotation around the user's pitch angle, meaning they fly upwards, vice versa.

Moving the fist to the right results in a rotation around the yaw angle, meaning the user turns to the right, vice versa. It is implemented by adding the differential vector to the current Euler-based orientation. Let $v_n \in \mathbb{R}^3$ be the position of the user's hand in neutral position, $v_c \in \mathbb{R}^3$ the current position and $\delta v := v_c - v_n$. Furthermore, let $o_c \in \mathbb{R}^3$ be the orientation of the user's virtual body and $a,b,c \in \mathbb{R}$ factors determining the orientation change. The new orientation is calculated on a frame-by-frame basis $o := \{o_{c_x} + \delta v_y * a, o_{c_y} + \delta v_x * b, v_x * c\}$. We then used the Unity function `Quaternion.RotateTowards` for smoothing, limiting the maximum orientation change between two frames to 2.6°.

5.2.2.2 Method 2: Bird

The Bird method involves more parts of the participant's body. Participants have to lift both arms, imitating a gliding bird or plane. Yawing angles are controlled by the relative position of both hands. If the left hand is lower than the right hand, whereas lower corresponds to being closer to the Kinect, the user's virtual avatar will turn left around the yaw axis, vice versa. The pitch axis is controlled by changing the spinal position up or down. The relative position of both hands controls yawing angles. Again, a neutral position is precalibrated for each participant before starting the flight. Let $v_l \in \mathbb{R}$, $v_r \in \mathbb{R}$ be the distance to the user's left and right hand respectively. Let $s_n \in \mathbb{R}$, be the distance to the user's spine during the calibration and $s_c \in \mathbb{R}$ be the current distance to the user's. Let $hs \in \mathbb{R}$ be the measured distance between both spread out hands, $s_{min}, s_{max} \in \mathbb{R}$ be the calibrated minimal and maximal distance to the spine. The rotation depends on whether the user is leaning forward or backward, as s_{min} and s_{max} values might be nonequal since leaning backward and forwards for humans is not symmetric. If they are leaning forward, $r_z := \frac{|s_c - s_n|}{s_{min} - s_n}$ and if they are leaning backward $r_z := \frac{|s_c - s_n|}{s_{max} - s_n}$, $r_x := \frac{|v_l - v_r|}{hs}$. Furthermore, let $o_c \in \mathbb{R}^3$ be the orientation of the user's virtual body and $a,b,c \in \mathbb{R}$ factors determining the orientation change. The new orientation is calculated on a frame-by-frame basis $o := \{r_z * a, o_{c_y} + r_x * b, r_x * c\}$. We then used the Unity function `Quaternion.RotateTowards` for smoothing, limiting the maximum orientation change between two frames to 2.6°,

5.2.3 Experiment

In this section, we describe the experiment that we conducted to evaluate the steering techniques for supernatural flying in a VR environment.

5.2.3.1 Participants

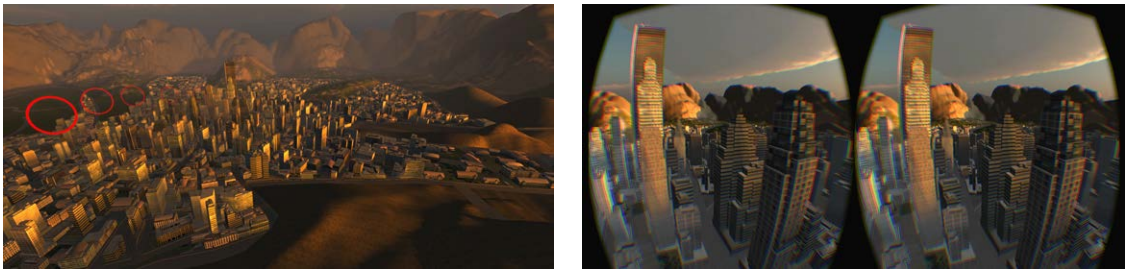
We recruited 13 participants for our experiment, 9 male and 4 female (ages 21–49, $M = 29$). The participants were students or professionals in computer science and employees of the administration office. Participants were naive to the experimental conditions. None of the participants reported known visual or motor disorders.

5.2.3.2 Materials

In addition to the setup described in Section 5.2.1, we used an Intel Core i7-4930K 3.4 GHz workstation with 16GB RAM and Nvidia GeForce GTX 780 Ti graphics card for the real-time rendering and sensor processing. In the experiment scenario, the user started in a hovering state in mid-air above a city with skyscrapers within a desert (cf. Figure 5.9).

5.2.3.3 Protocol

In the experiments, every participant had to pass the same course with both flight control methods, but with an altered order. Tracking of the participants was then calibrated by setting up their neutral position. After a short adaptation phase, during which they were allowed to look about or turn around, the flight is started. The adaptation phase helped them to get familiar with the steering and the influence of body movements on steering. When the flight begins, participants had to follow a



(a) (b)

Figure 5.9: Visual stimulus. The flight course is marked by rings, which are gates that have to be passed. (a) Overview of the track. (b) Users' view that is rendered for displaying with the Oculus Rift.

path marked by red rings. The rings disappeared after being passed. The participants' task was to pass as many rings as possible in preferably short time. After passing all rings, participants were allowed to fly around freely for as long as they desired.

5.2.3.4 Methods

We used repeated measures within-subjects design with two counter-balanced conditions. The independent variables were the control method (bird vs. superman). The dependent variable was the time it took the participants to reach the gates during the experiment, the accuracy how close to the center of the target gates they flew through.
Furthermore, we collected demographic information with a questionnaire before the experiment and measured the participants' task load with the NASA TLX questionnaire (cf. Section 2.4.3), as well as the cybersickness and the degree of presence of the proposed simulator (cf. Section 2.4.2.4). After completion of each of the tasks we collected informal responses from the participants and asked them to provide qualitative feedback related to the tested steering method. Participants had short breaks between the two conditions. During these breaks, the participants would be lowered to the ground and helped out of the setup to answer questionnaires. After the break, the participant had to put on the setup again, and we calibrated the neutral position again. The total time per participant including pre-questionnaires, instructions, experiment, breaks, post-questionnaires, and debriefing was 30 minutes.

5.2.3.5 Results

The SUS-PQ (cf. Section 2.4.2.4) was used to compare the level of presence of both control methods in the proposed system (cf. Figure 5.10(a)). The mean evaluation method for the SUS-PQ was used. In direct comparison, 6 participants gave higher ratings for the Superman method and 7 rated the Bird method with higher scores. In general, both methods had a high mean score (Superman: $M = 4.96$, Bird: $M = 5.11$). There was no significant main effect for presence in comparison of both proposed control methods ($Z = -.153$, $p = .878$). Results of the SSQ (cf. Figure 5.10(b)) revealed that using the system increases the level of cybersickness. Both methods show similar effects. The comparison of both methods showed no significant effect ($t(12) = .184$, $p = .05$). We found a significant effect comparing the cybersickness before ($M = 10.93$) and after the experiment for the Bird condition ($M = 29.63$, $t(12) = 2.616$, $p < .05$) and the Superman condition ($M = 25.70$, $t(12) = 2.572$, $p < .05$).

Figure 5.11 and Table 5.2 show the NASA TLX results. Due to a software error, we had to exclude a participant from the TLX evaluation.

Two participants did not pass all the gates and were removed from the time necessary between evaluation and the deviation evaluation. Figure 5.12 shows the recorded performance and deviation

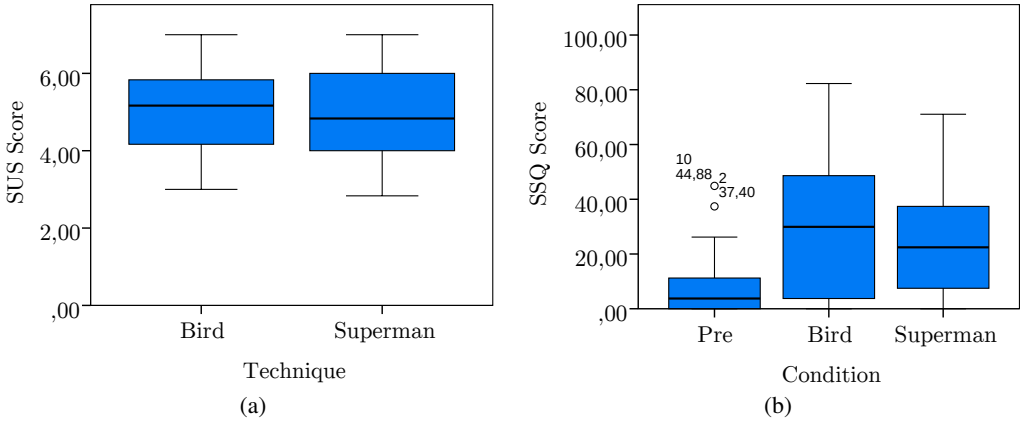


Figure 5.10: (a) Summary of SUS-PQ evaluating the level of presence. (b) Summary of SSQ questionnaire evaluating the influence on cybersickness level.

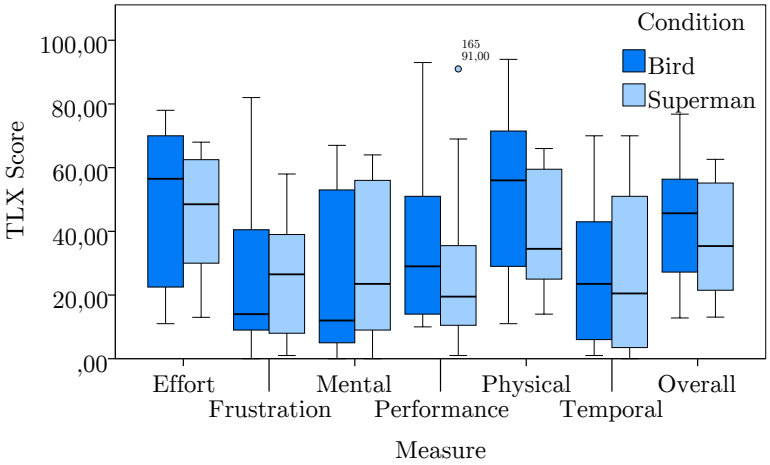


Figure 5.11: Summary of NASA task load index (TLX).

from the center of a checkpoint. We found no significant influence of the technique on the time the subjects needed to get from one gate to another ($F(1) = .679, p = .431, \eta_p^2 = .070$). We found a significant influence of the gate on the time ($F(11) = 2.795,186, p < .05, \eta_p^2 = .237$). We found no significant influence of the technique on the deviation ($F(1) = .264, p = .62, \eta_p^2 = .028$). We found a significant influence of the gate on the deviation ($F(11) = 5.692, p < .001, \eta_p^2 = .387$).

The collected qualitative feedback supports the quantitative results and generally indicates very positive judgments of the ability to steer the self-motion in supernatural flight with both techniques. 8 participants preferred the Bird control method over the Superman control method. 5 of them preferred the Superman control method. Comments of the participants often noted that the bird method was more realistic and comfortable. One participant noted: “With both arms stretched out you felt like you would bump into the buildings.” Increased nausea was reported when flying very close to buildings. During supervision, we noticed participants visibly twitching just before colliding with buildings in the VE.

5.2.3.6 Discussion

The significant results for the gate show that the difficulty in the track was adequately spread out, with some parts being more difficult than others. Some of the participants had problems using one method and did not manage to fly through all gates of the course.

Score	Bird condition		Superman condition		Wilcoxon Signed Rank Test	
	Mean	*SD*	*Mean*	*SD*	*Z*	*p*
Mental Demand	24.91	25.17	31.08	25.31	−.712	.477
Physical Demand	53.00	25.56	39.50	19.30	−1.570	.116
Temporal Demand	26.17	22.81	27.67	26.00	.000	1.000
Performance	37.67	27.50	27.67	27.26	−.786	.432
Effort required	48.33	25.20	45.33	19.55	−.785	.432
Frustration	26.08	24.66	25.50	19.89	.000	1.000
Overall TLX	43.06	19.63	37.23	17.82	−2.197	$< .05$

Table 5.2: TLX results.

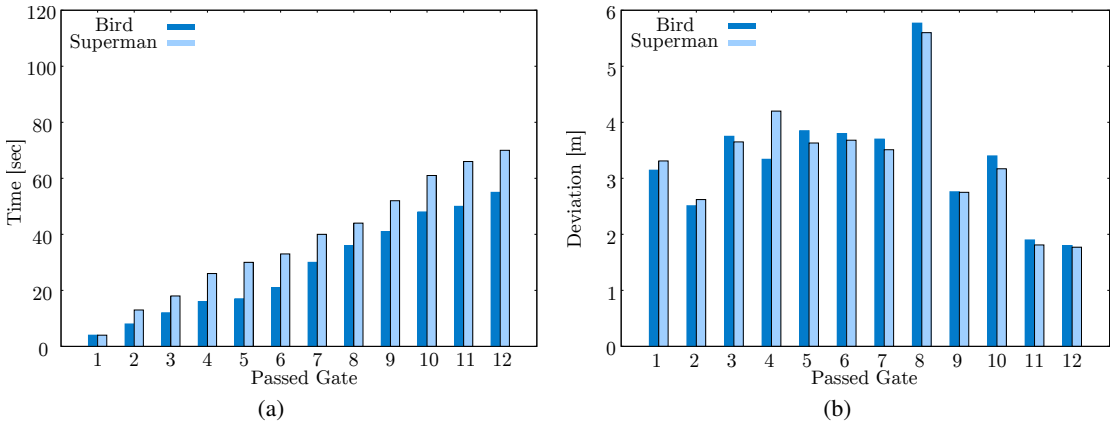


Figure 5.12: Median of flight times (a) and mean deviation from the center of the gate centers (b).

However, we could not find any effect of our techniques, except on the TLX results. The overall TLX scores indicate that the bird technique may be easier to use, although the single TLX subscores did not confirm that. The bird method for controlling the flight has the tendency to be preferred as well. This is supported by informal comments that report that the Bird method feels more natural or real. The presence results for both conditions was high, which indicates that the VE and our setup induced a relatively high sense of presence.

The high level of presence reported by the participants of the study presumably results from the combination of three different aspects: The mounting system, the control metaphors, and the high-quality visualization. The described flying metaphors induce a kind of (super-)realism that is similar to Superman movies, flying birds or airplanes. Then there is the mounting facility, where participants are hanging in mid-air, which induces a strong immersive experience. Finally, the high-quality visualization with effects of illumination, motion blur, and simulated physics completes the system to a very immersive simulator.

Unfortunately, for both techniques, we found an increase in cybersickness, which may still be caused by the old HMD (cf. Section 2.5). Participants reported stronger nausea when flying very close to buildings or objects in general. Participants showed individual preferences for the techniques. However, neither was preferred by all participants. Every untrained participant was at least able to use one of the control methods intuitively. This shows that self-flight control by body posture is sufficiently controllable. In general, the participants liked the setup of the system and enjoyed the flight. It is worth to continue working on both, or even more, control principles for flying in VR environments.

5.2.4 Second Iteration

In a second iteration, we decided to change the system to allow users to stand instead of lying while still suspended in the air. This change was meant to enable more reliable tracking with the Kinect 2, as the setup is more in line with the typical use case of the device. Additionally, we wanted the system to be more comfortable and avoid the physical strain of having to lift your torso. The main inspirations for the second iteration are the Iron Man movies and comic books. Besides 3DUIs, which are similar to Underkoffler's UIs in *Minority Report* [Got11], the Iron Man movies present the eponymous Iron Man suit and it's flight capabilities. The suit works using multiple thrusters that give the character very agile flight capabilities. Whereas current technology does not allow suits with such abilities, VR can be used to simulate it, and use the users' knowledge of the movies to shorten the training period possibly.

5.2.4.1 Iron Man in Science-Fiction

Iron Man's suit has thrusters in the gloves and boots that lift it off the ground and steer his direction of flight. For extra thrust, additional thrusters are in the back of the suit, allowing higher speeds. In addition to basic flying, Iron Man can (semi-) automatically regulate its thrust and correction nozzles to stand in the air. Iron Man uses different flight strategies in different situations. Near the ground and at obstacles, he is standing in the air and leans forward slightly to move forward. Iron Man makes slight sideways movements to move to the right or left or extends his arms forward to move backward. For fast course changes, he turns on the spot on his axis by pointing the hand thrusters in opposite directions. This allows a flight movement in 3 translational DOF and one rotational DOF around the yaw axis. For longer routes without obstacles, especially high up in the sky, Iron Man is usually flying horizontally. With this method, a dynamic dive is possible. We aimed to recreate the agility seen in the movies in our setup.

5.2.4.2 Setup Changes

The main hardware components with the climbing harness hanging from metal trusses were not changed for the second iteration. However we modified the way these components were utilized (cf. Figure 5.13). The main difference to the previous setup is that users are now standing while suspended in the air at the height of about 10 cm to 15 cm. Standing, rather than hanging in the harness, causes less weight to be distributed on the user's chest. This new setup also simplifies putting the setup on since users can easily step into the bands for their feet after putting the harness on. A trained user can even attach the carabiner hooks themselves, so they only need a supervisor to operate the winch. An elastic band was used to prevent the user from spreading their legs too much, imitating the resistance of a suit. Since the new setup is a standing setup, the Kinect 2 skeletal tracking appeared more robust. However no quantitative evaluation was done to prove that.

Flight Control

The flight strategies of Iron Man in the movies (cf. Section 5.2.4.1) can be condensed to two control concepts, which are described in detail in the following. First, based on the ground-level flight strategy, we developed a *translational control* where the user's virtual body is upright all the time, moving right and left, up and down, and forward and backward. The corresponding conceptual model behind this user interface would be that the flight suit would automatically ensure an upright pose through "stabilizing thrusters". Without these, an arbitrary orientation of the body with a *rotational control* would be possible (cf. Figure 5.14, similar to the techniques described in Section 5.2. This is based on the long-range travel strategy of Iron Man, in which the body and flight direction are almost parallel to the ground.

We considered matching the user's body orientation to the one of the virtual body. We found out that getting upright after leaning forward is cumbersome for some users. Thus, a quick swap between both techniques would be impossible. We accordingly decided to only turn the virtual

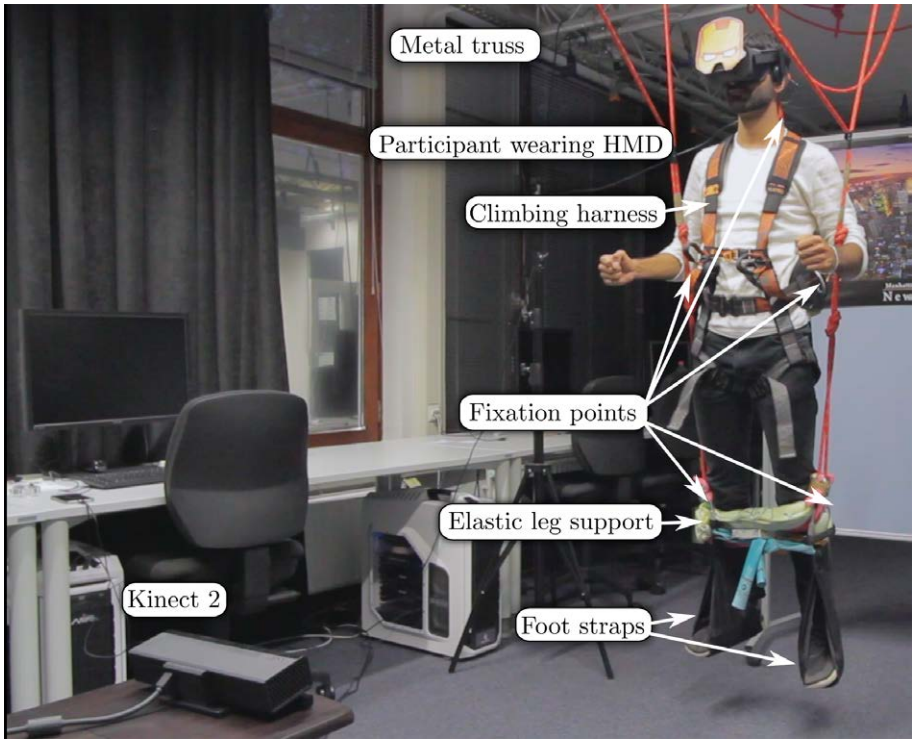


Figure 5.13: Basic setup of the Iron Man Simulator.

Figure 5.14: Illustration of the virtual initial pose for the (left) translational control and (right) rotational control.

body into a horizontal orientation for the rotational control. The user could stay in a vertical pose.

Translational Control

The translational control uses the position of the user's arms, hands, and torso. The user's virtual body stands upright in the virtual space (cf. Figure 5.14). Using the Kinect 2, the position of both wrists and elbows is tracked, resulting in a direction vector ($v_1, v_2 \in \mathbb{R}^3$) from the elbow to the wrist for each arm. The two vectors in the conceptual model correspond to the directions of the thrust of nozzles in the user's palms. We compute the virtual body's translation vector $v = v_1 + v_2$ by

Figure 5.15: Illustrations of the flight controls: (left and middle) User pose for the translational control. (right) Gesture interaction for rotational control.

adding these two thrust vectors. Again, by opening or closing their hands, the user can activate or deactivate thrust. If both hands are open at the same time, the virtual body is moved in direction v. Since we found out that a forward movement with the arm control alone is difficult to accomplish in informal pilot studies, we decided to add the upper body as another control component. The difficulty was partly caused due to a single, front-facing Kinect 2 not reliably detecting arms behind the user (cf. Section 2.5). Thus, the user can initiate a forward movement by leaning forward, which simulates the exhaust nozzle in the back of Iron Man. While leaning forward, opening just one hand causes the virtual body to rotate around the yaw axis. Opening the right hand causes a rotation to the left, vice versa. The further away from the body the open hand is, the faster the rotation. When the hands are closed and the posture is upright, the virtual body slows down and stops.

Rotational Control

In contrast to translational control, the virtual body in is in a horizontal position (cf. Figure 5.14). The virtual body flies in the direction of the central body axis, from the feet to the head. For example, if the virtual body is in an upright pose, it flies upwards.

Once more, opening the hands causes acceleration up to a predefined maximum speed, and closing hands causes deceleration. Leaning forward causes a downwards rotation around the pitch axis, holding both hands in front of their body caused an upwards rotation. To rotate around the yaw axis (turn right / left when the avatar is lying), the user raises the right or left arm as if to assume a half-T pose. The maximum angular momentum is reached when the user's pose matches the ideal rotary poses (cf. Figure 5.15).

To determine the direction and speed of rotation, the angles α_i between the arm vector (direction from the shoulder to the hand) a_{left}, a_{right} and the ideal direction vectors d_i (all vectors are normalized) with $\alpha_i = cos^{-1}(a_j * d_i)$ is calculated with $j \in \{left, right\}$ and $i \in \{pitch, yaw\}$. For each arm vector a_j the values $percent_i = \frac{(totalAmount - \alpha i)}{totalAmount}$ with $totalAmount = \alpha_{pitch} + \alpha_{yaw}$ are determined. If one of the $percent_i > 0.6$, a rotation around the respective axis i is triggered. The 0.6 threshold is designed so that the arm's resting position (arm loosely hanging down) does not result in an inadvertent rotation. As soon as a rotation is to take place, the respective $percent_i$ value is further modified by an S-curve function. This function ensures that a more precise adjustment of the speed is possible at a slow rotational speed than at a high rotational speed. For this purpose, $percent_i \in [0.5; 1]$ to the value range $[-4; +4]$ stretched and then through the S-curve function $f(x) = \frac{x}{\sqrt{1+x^2}}$ with $f(x) \in [-1; 1]$ modulated. The value in the range $[0; 1]$ is transformed, multiplied by the maximum rotation speed and converted as angular momentum.

Head-Up Display

To display additional information on the environment, a head-up display (HUD) is shown. The HUD fills parts of the FOV in the HMD and shows the user information about the current body pose, flight altitude, speed, and collision information about objects that are directly in the direction of flight. An additional map shows the position in the virtual world together with destination information in

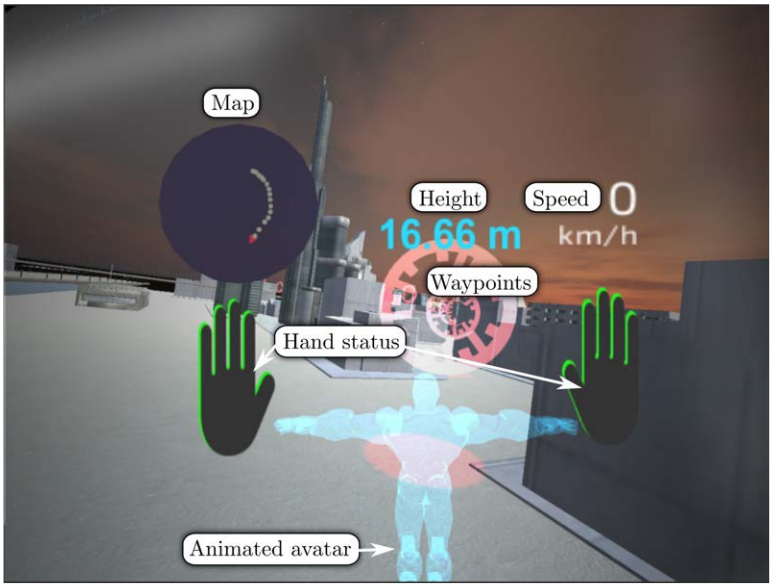


Figure 5.16: Illustration of the HUD.

the manner of a waypoint-based navigation system. Besides, the user receives feedback about the information the Kinect has about the user's body. The hand status (open or closed) is displayed along with a miniature version of the avatar. This miniature body moves according to the arm, head and body movement of the user to provide information about the Kinect-tracked body alignment at all times. This allows the user to see and correct wrong interpretations of the Kinect directly. This is particularly useful for hand recognition that has not always been accurate during development. The color of the hand illustrations changes to green when the hand is open and to gray when the hand is closed. These elements should help the user with orientation in flight. Another reason for integrating a HUD is that an optical anchor in the user's field of vision can potentially reduce the symptoms of movement and cybersickness, and especially nausea (cf. Section 2.4.2.4)

5.2.4.3 Evaluation

We conducted a usability study to evaluate the updated system and compare the two flight controls. We evaluated the following hypotheses:

- $\mathbf{H}_1$ On the long route, the rotational control is easier to operate because the user only has to do slight compensatory movements, while the directional control requires more movement in different directions. The rotational control is comparably good or even better regarding performance than the directional control.
- $\mathbf{H}_2$ At a short route, directional control is better concerning usability and performance, as it can be used to adapt to difficult routes because of its ability to move sideways in confined spaces.

Materials and Methods

We used the setup described in Section 5.2.4.2. We tested both flight controls over a short-distance (route A) and long-distance (route B) scenario. For this purpose, ring-shaped waypoints were displayed in the navigation system (cf. Figure 5.16). The subjects had to fly through 15 consecutive hovering rings. The waypoints in the long route were further apart and required more direction changes up, down, left or right. The route was clear of obstacles and was designed to simulate the environment for Iron Man's fast flight strategy higher up in the sky. The speed was limited on the long route to 54 km/h, while the short route was flown at a maximum of 36 km/h. On

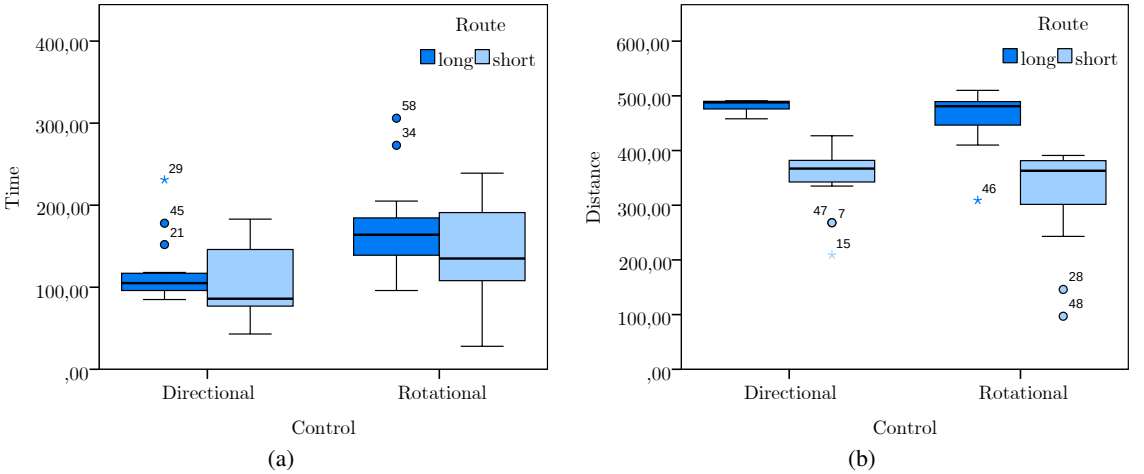


Figure 5.17: Summary of the (a) time required for the different conditions and (b) the precision when flying through a checkpoint.

the long route, the speed was higher to match the scenario of the fast flight strategy. The short, low-speed route required precise navigation in a small space with closely spaced rings and obstacles, such as buildings. Here, it was important that emerging obstacles could be quickly avoided. This environment simulated the application scenario when Iron Man flies upright near the ground.

We used a full-factorial repeated-measures design experiment. Each subject tested all conditions. Overall, each subject flew four times in both routes with both controls in randomized order.

We collected the data using digital questionnaires (cf. Section 2.4.2.4) and automatically collected user behavior in the virtual world in terms of the time and accuracy required to complete the navigation tasks. The number of rings passed correctly and the precision of flying through the individual rings were measured by the distance of the avatar to the center of the ring at the time of traversal. The value 0 would be an ideal traversal, and the value 600 would be the worst possible traversal.

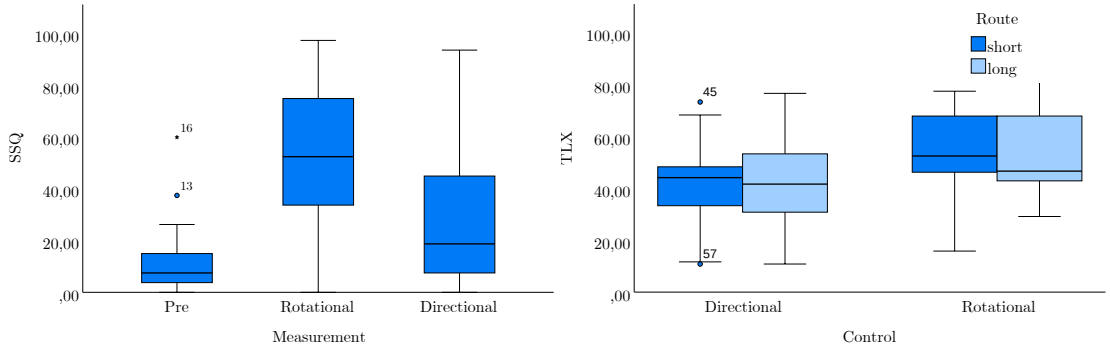
We used the SSQ to measure symptoms of motion and cybersickness before and after the conditions to record the changes due to a flight. Also, we employed the SUS-PQ to determine the subjective level of presence. Besides, to examine the physical and mental effort of subjects in the terms, the NASA TLX was chosen. These questionnaires were answered for each flight control individually to be able to evaluate a part of H_1 (rotational control is easier to operate in free space). Finally, we used a demographic questionnaire.

Participants

24 participants (8 female and 16 male, ages 18–37, $M = 24.74$) took part in the experiment. The experiment took about one hour per participant. 12 people were able to complete the experiment. The rest of the experimental subjects prematurely stopped the trial due to symptoms of cybersickness. The subjects were recruited through intra-university posters and notices. Student participants acquired class credit for their participation.

Results

The average flight time for the directional control (cf. Figure 5.17(a)) was $M = 118.79$ s ($SD = 111.01$ s) and for the rotational control $M = 151.66$ s ($SD = 60.22$ s). Shapiro-Wilk tests of normality showed that the time data was not normally distributed ($p < .05$). Thus, we used a Friedman two-way analysis of variance by ranks and found no differences between the conditions $\chi^2(3) = 5.698, p = .127$.



(a) (b)

Figure 5.18: Summary of the (a) SSQ and (b) TLX scores.

The median count of rings passed for all conditions was 14. A Friedman test on the ordinal count of passed rings and found no differences between the conditions $\chi^2(3) = .728, p = .867$.

The precision data is shown in Figure 5.17(b). Shapiro-Wilk tests of normality showed that the precision data was not normally distributed ($p < .05$). We used a Friedman two-way analysis of variance by ranks $\chi^2(3) = 34.611, p < .001$ and found no influence of the flight control method on the precision results $\chi^2(1) = 0.31, p = .577$, but an influence of the route $\chi^2(1) = 30.0, p < .001$. The mean precision score for the directional control ($M = 415.77, SD = 79.06$) was lower than for the rotational control ($M = 390.70, SD = 101.13$). The mean precision score for the long route ($M = 470.97, SD = 30.10$) was lower than for the short route ($M = 335.50, SD = 76.55$). Shapiro-Wilk tests of normality revealed that our SSQ measurements (cf. Figure 5.18 (a)) before the trial and after. A paired T-Test of the accumulated differences between the SSQ values before and after the conditions yielded a significant difference with $t(16) = -3.5739$, $p < .01$. Subjects have reported stronger symptoms of cybersickness in the rotational control conditions with $M = 5.82$ ($SD = 3.60$) than in conditions with directional control with $M = 1.17$ ($SD = 3.22$).

A Shapiro-Wilk test of normality showed that the TLX data (cf. Figure 5.18 (b)) is normally distributed (all $p > .05$). We analyzed the results with a repeated measures ANOVA and found a significant main effect of control ($F(1,14) = 8.201, p < .05, \eta_p^2 = .369$). The mean TLX score for the directional control ($M = 41.53, SD = 17.79$) was lower than for the rotational control ($M = 53.17, SD = 17.41$). We did not find an effect of the route on the TLX score ($F(1,14) = .027, p = .871, \eta_p^2 = .002$). We did not find any interaction effects.

A paired T-Test on the means of the SUS-PQ presence questionnaire showed no significant difference ($t(14) = .505, p = .621$). The directional control had an average of $M = 4.4$ ($SD = .92$) and the rotational control was $M = 4.2$ ($SD = .84$).

Discussion

In our study, we compared both flight controls and investigated whether they differed in their maneuverability and usability. We did not find any influence of the flight control on the performance for the short route and long route flights, but the performance has generally been good. However, the precision in the long route was significantly higher, which supports our assumption that the routes differed in difficulty. The obstacles on the short route make it harder to pass the rings precisely. We assumed that the rotational control would make the long route easier, however, the results of the TLX show that the contrary is the case. Thus, we have to reject H_1.

While our hypothesis H_2 was one-sided, we did find a difference between the controls. The directional control was more usable, according to the TLX results. However, we did not find a difference in precision or speed. Thus we can only partly confirm H_2,

The SSQ results revealed the rotational control caused more sickness than the directional control. Especially regarding the high drop-out rate in our experiment, we must conclude that this is indeed the most important of all the measured values, even regarding practical application scenarios. The results of the investigation indicate that the directional control is preferable to the rotational control.

Nonetheless, the rotational control should be further investigated, as it may not have been able to demonstrate its full potential under the circumstances of the investigation. Our participants were naive to the experiment conditions and barely had time to adjust to the individual controls. Based on observations, we can hypothesize that the rotational control would behave better with more training. Participants often made very jerky turns at full speed in the test setting with the rotational control and barely used the opportunity to turn more slowly. This observation could account for part of the high dropout rate. For example, some subjects accidentally caused 360° turns, which could be because some participants were inclined to test the system's limits. In the future, it would be possible to change the sensitivity of the rotational control for optimization and, if necessary, carry out another test for optimal sensitivity. Overall, these observations suggest that rotational control would benefit more from a longer training session than directional control, but at the same time that directional control is potentially more intuitive. A total of 82% of our participants were familiar with the character Iron Man. It is also possible that the directional control is more in line with the expectation of flying like Iron Man and the behavior of nozzles on the hands and the higher level of intuitiveness can be explained.

5.2.5 Conclusion

In this section, we presented our suspended flying setup in two iterations. We evaluated two flight controls each. In both setups, users positively responded to the presence and the supernatural travel techniques. Previous research has shown that further improvements to the presence could be achieved through the stimulation of more senses. We could not find performance differences between the techniques in either iteration, but the feedback and measurements indicate that the performance is overall adequate for all techniques. However, for all the setups, we found an increase of cybersickness, which is especially challenging since some participants had to abort the experiment. One of the reasons for that might be the vehicle metaphor used, meaning that the vestibular sense conflicted with the visual sense (cf. Section 2.4.2.4). Nonetheless, current hardware improvements, especially the increased frame rate of modern HMDs may decrease the cybersickness symptoms and allow using suspended flying as a VR travel solution.

Thus, we decided to investigate natural walking and address its limitations as a VR travel technique.

5.3 Natural Walking

Natural walking is regarded as the most natural form of traveling through a VE. In particular, traveling by means of *real walking* has received much attention recently due to results of perceptual and cognitive experiments. These showed that walking has significant advantages over other forms of traveling in terms of the user's sense of feeling present in the VE [Uso+99a], navigational search tasks [RL09], cognitive map buildup [RVB10], and required cognitive resources [Mar+13b] (cf. Section 2.6.2). Nonetheless, when asked to walk in a VE, users tend to walk differently than in real environments [Jan+17a]. Especially distance misperception is often mentioned as a possible reason for this behavior (cf. Section 2.4.2). One of the reasons participants in our experiments mentioned is that they are afraid of walking into real obstacles when wearing a fully-immersive HMD. Another shortcoming of realizing walking VR travel interfaces by a one-to-one mapping from real position to virtual position is the usually limited physical space. Given a VE which is

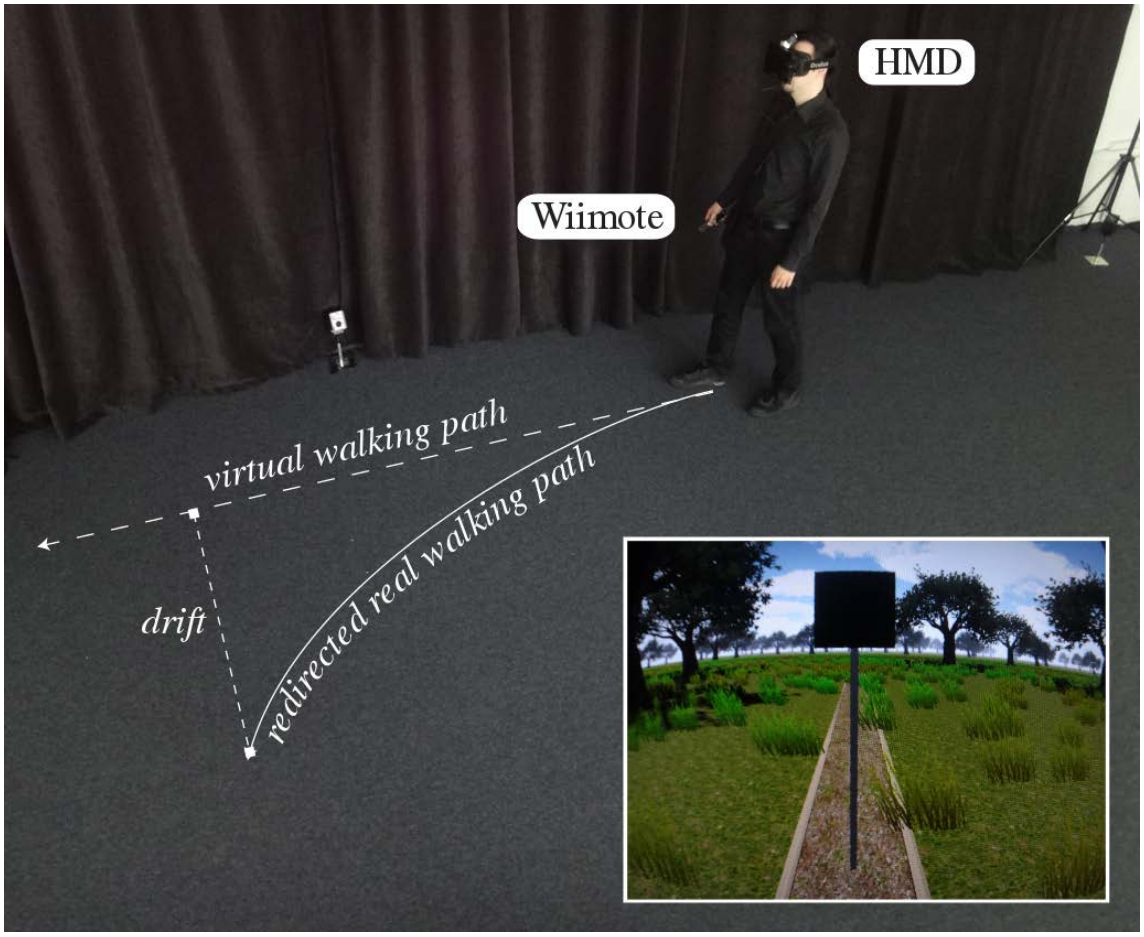


Figure 5.19: Redirected walking scenario: A user walks in the real workspace with an HMD on a curved path in comparison to the visually perceived straight path in the virtual world. The inset shows the user's view on the HMD as used for the experiment.

larger than the physical environment, solutions had to be developed that allow users to explore the complete VE. The developed solutions were discussed in Section 2.6.2. The technique we are considering in this Section is RDW [Raz05], which is using the dominance of the visual sense when the input of the sensory channels conflicts [Ber00; DB78] (cf. Figure 5.19).

First, we present an analysis of the cognitive demands of RDW to assess whether it is cognitively distracting or not [BLS15]. Afterward, we utilize RDW to allow supernatural travel through VEs, but also give users a safe area within which they can explore the VE naturally, without any manipulation [LBS14c].

5.3.1 Cognitive Demands

Redirection techniques have been applied particularly in robotics to control a remote robot by walking [Gro+05]. For such scenarios, much effort has been undertaken to prevent collisions [Gro+05; NHS04]. These techniques guide users on physical paths for which lengths, as well as turning angles, of the visually perceived paths, are maintained. But the users observe the discrepancy between both worlds and have to consciously follow the visual stimuli, which may introduce severe cognitive demands. For a more extensive literature review of natural walking in VR cf. Section 2.6.2. In particular, to our knowledge, no previous work provided insight into the amount of cognitive demand that is induced by redirected walking manipulations of different magnitudes on the walking

user. This section is based on work previously published in [BLS15].

Theoretical Models of Cognitive Resources

In addition to the human cognition models discussed in Section 2.4.3, the human working memory draws from finite cognitive resources, for which several theoretical models have been proposed [GC93]. These models usually distinguish at least between verbal and spatial cognitive resources [BH74]. A well-known theoretic model of cognition and working memory was proposed by Baddeley and Hitch [Bad12; BH74], which considers manipulation and storage of visual and spatial information in a speech-based loop. According to this model, general attention and access to both verbal and spatial working memory are handled by a central executive. If concurrent verbal and spatial tasks interfere with a similar magnitude with a task of interest, then the task is considered to require general attention resources (or may require similar amounts of verbal and spatial resources).

Because of limitations in the sensory feedback and required control actions described above, redirected walking cannot be considered truly natural. The user is required to actively, consciously or subconsciously, compensate for the introduced manipulations. Those aspects may cause users to employ strategies requiring additional cognitive resources. These strategies compete for cognitive resources from the same pools that are utilized for successful completion of a user's primary domain-related tasks.

In particular, in this section we analyze and discuss:

- Effects of redirected walking on verbal and spatial working memory tasks.
- Influence of verbal and spatial tasks on locomotion behavior when using redirection.
- Implications for using redirected walking techniques in IVEs.

To answer these research questions, dual-task studies can be used, which are a widely used method to understand influences of cognitive tasks on traveling or locomotion gait and balance (cf. Woollacott and Shumway-Cook [WS02] for a review). The dual-task method requires users to perform a secondary task while performing a primary task to determine the costs involved in performing the concurrent task [Bea+05], such as performing an additional cognitive task while walking through a virtual world. Various secondary tasks have demonstrated the interaction effects between cognitive demands and locomotion. Different experiments have used speech as the distraction task [GC93], and others have used secondary tasks based on intentional movements involving a motor or muscular component [Mor+00] or even electrical stimulation [Reg+05]. For instance, using a counting-backward task, effects on locomotion were observed in older adults, but not in young adults [Bea+05]. Cognitive costs of locomotion are observed via dual-tasking, for example, by studying changes in speed, cadence, step-length and double support time while engaged in secondary tasks. Observed decrements in locomotion performance are presumed to be caused by a limited attentional capacity depending on the complexity of the concurrent task [Nad+10].

This hypothesis is supported by experiments by Zanbaka et al. [Zan+05], which have shown that using certain locomotion interfaces that are not entirely natural can be cognitively demanding. Furthermore, Nadkarni et al. [Nad+10] have demonstrated that cognitive tasks which activate working memory and spatial attention can have a significant effect on human locomotion. In particular, they found that changes in gait parameters, including gait speed, stride length, and double support time, were affected by cognitive tasks. In experiments described by Marsh et al. [Mar+13b], the dual-task selective-interference paradigm was implemented to analyze the impact of spatial and verbal cognitive tasks on different aspects of locomotion. In particular, they compared the cognitive resource demands of locomotion user interfaces that varied in their naturalness as well as the impact of a restricted FOV on cognitive working memory demands while performing movements in an IVE. Their results suggest that locomotion with a less natural interface increases spatial working memory demands, and they showed that locomotion with a smaller FOV increases general attentional demands.

5.3.1.1 Experiment

As discussed above, applying redirected walking in IVEs may not only be noticeable but can induce cognitive demands on the user that are competing with other tasks for finite cognitive resources. In this section, we describe the experiment in which we analyzed such a mutual influence of redirected walking and two different (spatial and verbal) cognitive tasks.

Participants

16 subjects (11 female and 5 male, ages $19-45$, $M=27.6$) participated in the experiment. The participants were students or members of the local department of computer science, who obtained class credit for their participation. All of our participants had normal or corrected-to-normal vision. Two participants wore glasses, and four participants wore contact lenses during the experiment. None of our participants reported a disorder of equilibrium. One of our participants reported a slight red-green weakness. Our participants reported no other vision disorders. Ten participants had participated in an experiment involving HMDs before. We measured the IPDs of our participants before the experiment [Wil+08] (cf. Section 2.4.2.1). The IPDs of our participants ranged between 5.6 cm to 6.3 cm ($M=6.3$ cm, $SD=0.3$ cm). We used the IPD of each participant to provide a correct perspective on the HMD. Participants were naive to the experimental conditions. The total time per participant, including pre-questionnaires, instructions, experiment, breaks, post-questionnaires, and debriefing, was 1 hour. Participants wore the HMD for approximately 45 minutes. They were allowed to take breaks at any time.

Materials

We performed the experiment in a $12\,\text{m}\times 5\,\text{m}$ darkened laboratory room. As illustrated in Figure 5.19, subjects wore a wireless Oculus Rift DK1 HMD for the stimulus presentation (cf. Section 2.5). We attached an active infrared marker to the HMD and tracked its position within the laboratory with a WorldViz PPT-X4 (cf. Section 2.5). The head orientation was tracked with an InterSense InertiaCubeBT sensor at 180 Hz update rate that we attached to the HMD. We compensated for inertial orientation drift by incorporating the PPT optical heading plugin. The visual stimulus consisted of a simple VE with grass, trees, and pavement (cf. Figure 5.19). We used an Asus WAVI wireless kit to transmit the rendered images at 60 Hz from a rendering computer to the HMD. As claimed by the manufacturers, not more than 22 ms latency is introduced due to the wireless connection. An Anker Astro Pro2 portable battery powered the HMD and wireless transmitter box. The boxes were carried in a small belt bag. For rendering, system control and logging we used an Intel computer with 3.4 GHz Core i7 processor, 16 GB of main memory and two Nvidia GeForce 780Ti SLI graphics cards. To focus participants on the task, no communication between experimenter and subject was performed during the experiment. Task instructions were presented via slides on the HMD during the experiment. Participants performed the cognitive tasks via button presses on a Nintendo Wii remote controller. Participants wore fully-enclosed Sennheiser RS 180 wireless headphones during the experiment. We used the headphones to display forest and nature sounds to minimize the ability of participants to estimate their physical position and orientation in the laboratory via ambient noise. The participants received auditive feedback in the form of a clicking sound when they pressed a button on the Wii remote controller.

Methods

We made use of a $3\times 9\times 2$ dual-task within-subjects experimental design. We tested 3 cognitive conditions (i.e., verbal task, spatial task, and no task), and 9 locomotion conditions (i.e., redirected walking with curvature gains [Ste+10] $g_C\in\{-\frac{1}{2.5},-\frac{1}{5},-\frac{1}{10},-\frac{1}{15},0,\frac{1}{15},\frac{1}{10},\frac{1}{5},\frac{1}{2.5}\}$), with 2 repetitions each. Thus, the experiment conditions included a single-task walking condition, and two dual-task conditions (walking plus either spatial or verbal working memory task). We maintained the fixed above order of the cognitive conditions but randomized the locomotion conditions. This

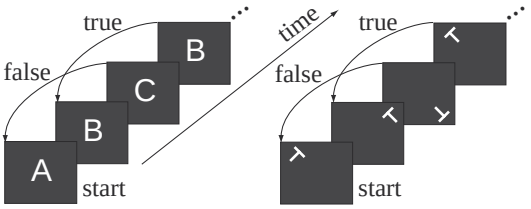


Figure 5.20: Illustration of the cognitive two-back tasks: (left) verbal working memory task and (right) spatial working memory task.

ensured that none of the cognitive tasks would be favored due to potential training effects (cf. Marsh et al. [Mar+13a]).

Before the experiment, all participants filled out an informed consent form and received detailed instructions on how to perform the cognitive tasks. Furthermore, they filled out the SSQ [Ken+93] (cf. Section 2.4.2.4) immediately before and after the experiment, further the SUS-PQ [Uso+99b], (cf. Section 2.4.2.4) and a demographic questionnaire. Every participant practiced each of the cognitive conditions four times before the experiment started, twice while standing in the VE, and twice during redirected walking with randomized gains, which we used to minimize later training effects [Mar+13a]. In total, participants completed 12 training trials.

Locomotion Tasks For each trial, participants were instructed to direct their gaze to a target displayed at 2 m distance in front of them along the virtual pavement (cf. Figure 5.19). The target moved at a speed of $0.55\,\mathrm{m/s}$ along the path in the VE during the experiment trials. In the locomotion conditions, participants were instructed to follow the leading target while maintaining the initial distance of 2 m, similar to the task used by Neth et al. [Net+11]. The total distance walked was 7 m over a duration of 12.6 s, after which the trial ended, and participants were guided to the next start position in the laboratory by aligning two markers in an otherwise blank 2D view. The subsequent trial started once participants reached the new start position and indicated that they were ready to start by pressing a button on the Wii remote controller.

While participants were walking along the virtual pavement, we applied different curvature gains [Ste+10]. These gains exploit the fact that when users walk straight ahead in the virtual world, iterative injections of reasonably small camera rotations to one side force them to walk along a curved path in the opposite direction in the real world to stay on a straight path in the VE. Curvature gains $g_C \in \mathbb{R}$ define the ratio between translations and applied virtual scene rotations, i.e., they describe the bending of the user's path in the real world. The bending is determined by a segment of a circle with radius $r \in \mathbb{R}^+$, as illustrated in Figure 5.19. Curvature gains are defined as $g_C := \frac{1}{r}$, with $g_C = 0$ for real walking with $r = \infty$. If the injected manipulations are reasonably small, users will unknowingly compensate for the virtual rotations and walk along a curved path. Curvature gains $|g_C| \in [0, 0.045]$ are considered undetectable for users (cf. Steinicke et al. [Ste+10]). We tested gains $g_C \in \{-\frac{1}{2.5}, -\frac{1}{5}, -\frac{1}{10}, -\frac{1}{15}, 0, \frac{1}{15}, \frac{1}{10}, \frac{1}{5}, \frac{1}{2.5}\}$, i.e., each curvature was tested both in clockwise and counterclockwise direction. The tested gains correspond to circular radii that fit within laboratories with a walking area of $5\,\mathrm{m} \times 5\,\mathrm{m}$, $10\,\mathrm{m} \times 10\,\mathrm{m}$, $20\,\mathrm{m} \times 20\,\mathrm{m}$, or $30\,\mathrm{m} \times 30\,\mathrm{m}$, respectively.

We used the standard deviation of *lateral head movements* when walking straight ahead along the path in the VE as the dependent variable, which provides indications on how locomotion behavior is affected by redirected walking.

Cognitive Tasks To analyze the cognitive resource demands of redirected walking we considered verbal and spatial working memory tasks. Participants registered their responses on the cognitive tasks (detailed below) by pressing a button on the Wii remote controller. The display duration for every stimulus on the cognitive paradigms was set to 500 ms with a pseudo-randomized interstimulus interval of 1100 ms–1500 ms similar to Baumann et al. [BRK07], thereby allowing for 6 stimuli for every trial with 4 recorded responses. Participants were instructed to perform the cognitive task to the best of their ability while maintaining the distance to the leading target in the locomotion dual-task conditions. Our dependent variable was the *percentage of correct responses* in the cognitive tasks, which indicates how the cognitive tasks are affected by redirected walking.

Verbal two-back working memory task As illustrated in Figure 5.20(left), the verbal working memory task was a letter *two-back task* [GC93] (cf. Section 2.4.3). In every trial, participants were shown a continuous stream of letters that were flashed on a virtual sign in the VE. Participants maintained a fixed 2 m distance to the leading target, which always ensured good readability. Participants were instructed to respond by pressing the button on the Wii remote if a presented letter was the same as the one that came up two stimuli back in the sequence (true condition in Figure 5.20(left)). This task has a high verbal working memory load since it requires continuous on-line monitoring and maintenance of the presented letter until two consecutive letters appeared. If (and only if) the stimulus matched the one that came up two stimuli before it, participants had to press a button on the Wii remote. This task did not require large shifts of spatial attention or memory as the letters appear continuously in the center of the screen.

Spatial two-back working memory task As illustrated in Figure 5.20(right), the spatial working memory task examined covert shifts of spatial memory and attention. In every trial, participants were shown a continuous stream of T-shaped symbols that were flashed on a virtual sign in the VE. The stimulus appeared in one of the four corners of the sign rotated by $\Theta \in \{\frac{1}{4}\pi, \frac{3}{4}\pi, \frac{5}{4}\pi, \frac{7}{4}\pi\}$ radians. The participants always maintained a distance of 2 m to the leading target, which always ensured good readability. Participants were instructed to respond by pressing a button on the Wii remote if a presented symbol was oriented in the same way as the one that came up two stimuli back in the sequence (true condition in Figure 5.20(right)). This task did not require large verbal working memory. The displayed symbols are considered hard to verbalize (cf. [BRK07]).

5.3.1.2 Results

We found no effect of walking clockwise or counterclockwise along a circular path and therefore pooled the data. Figure 5.21 shows the combined results for the tested curvature gains plotted against the performance in the locomotion and cognitive tasks. The vertical bars indicate the standard error of the mean. The colored lines show the results for the verbal task, spatial task, or condition without a cognitive task. The x-axes show the pooled (absolute) curvature gains $|g_C|$. The y-axes show the standard deviation of lateral movements in Figure 5.21(a), and the percentage of correct responses in the cognitive tasks in Figure 5.21(b).

The results were normally distributed according to a Shapiro-Wilk test at the 5% level. We analyzed the results with a repeated-measures ANOVA and Tukey multiple comparisons at the 5% significance level with Bonferroni correction. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated.

Locomotion Performance

We observed low lateral sway without redirected walking manipulations and without cognitive tasks, which was indicated by a standard deviation in lateral movements of $M = 0.042$ m, $SD = 0.0058$ m. Lateral sway increased for both cognitive tasks as well as for larger curvature gains.

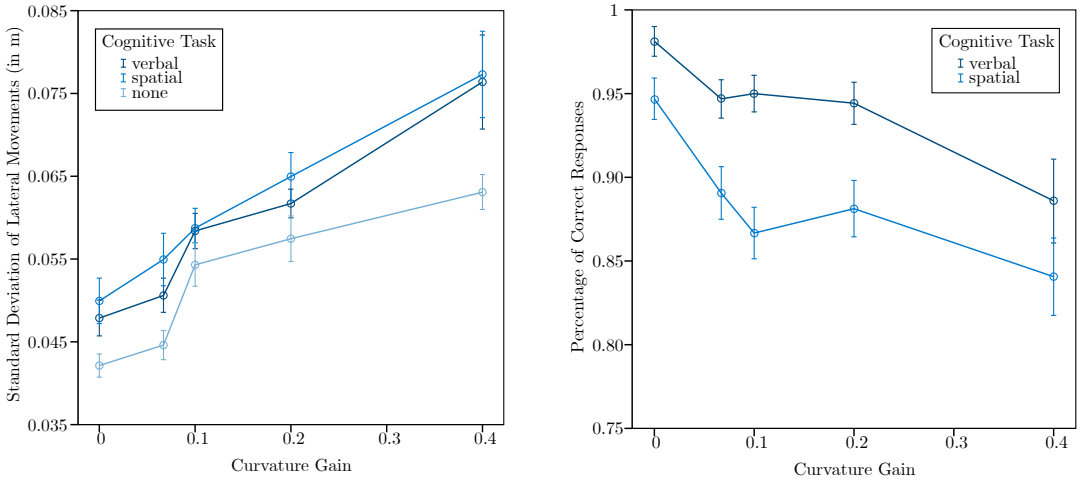


Figure 5.21: Pooled results of the experiment with (a) standard deviation of lateral movements, which indicates corrective movements during forward locomotion, and (b) percentage of correct responses for the cognitive tasks.

We found a significant main effect of curvature gain on the standard deviation of lateral movements ($F(1.487, 22.307) = 34.003$, $p<.001$, $\eta_p^2 = .694$). Post-hoc tests showed that the lateral sway between each two curvature gains was significantly different ($p<.05$), except between $g_C = 0$ and $|g_C| = \frac{1}{15}$ ($p = .66$) as well as between $|g_C| = \frac{1}{10}$ and $|g_C| = \frac{1}{5}$ ($p = .45$).

In addition, we found a significant main effect of cognitive task on the standard deviation of lateral movements ($F(2,30) = 11.993$, $p<.001$, $\eta_p^2 = .444$). Post-hoc tests showed that the results significantly differed between the spatial task and no task ($p < .001$) and between the verbal task and no task ($p < .05$), but not between the spatial task and the verbal task ($p = .621$). Both cognitive tasks exhibited similar effects on lateral sway.

We did not find a significant interaction effect between cognitive task and curvature gain on the standard deviation of lateral movements ($F(2.970, 44.545) = 1.598$, $p = .20$, $\eta_p^2 = .094$).

Cognitive Performance

We observed high task performance without redirected walking manipulations both for the spatial task (M = .947, SD = .050) and the verbal task (M = .981, SD = .036). Task performance was decreased for both cognitive tasks for larger curvature gains.

We found a significant main effect of curvature gain on the percentage of correct responses ($F(2.758, 41.370) = 10.887$, $p<.001$, $\eta_p^2 = .421$). Post-hoc tests showed no significant differences between each two curvature gains ($p<.05$).

Moreover, we compared the percentage of correct responses for the spatial and the verbal task with a paired T-Test. We found a significant main effect of the cognitive task on the percentage of correct responses between the spatial task and the verbal task ($p<.001$). Participants made significantly more errors in the spatial task compared to the verbal task.

We found a significant interaction effect between cognitive task and curvature gain on the percentage of correct responses ($F(3.557, 53.357) = 3.419$, $p<.05$, $\eta_p^2 = .186$). For the verbal task, the results show a significant difference of the largest tested curvature gain $|g_C| = \frac{1}{2.5}$ to all other gains except for $|g_C| = \frac{1}{5}$. In particular, post-hoc tests showed for the verbal task a significant difference between curvature gains $g_C = 0$ and $|g_C| = \frac{1}{2.5}$ ($p<.001$), and between $|g_C| = \frac{1}{15}$ and $|g_C| = \frac{1}{2.5}$ ($p < .05$), as well as between $|g_C| = \frac{1}{10}$ and $|g_C| = \frac{1}{2.5}$ ($p < .05$). For the spatial task, the

results show a significant difference between $g_C = 0$ and $|g_C| = \frac{1}{5}$ ($p < .05$), as well as between $g_C = 0$ and $|g_C| = \frac{1}{15}$ ($p < .001$).

Questionnaires

We measured a mean SSQ-score of M = 13.2 (SD = 15.2) before the experiment, and a mean SSQ-PQ score of M = 47.4 (SD = 60.8) after the experiment. The results indicate a typical increase in cybersickness for extensive walking with an HMD over the time of the experiment. The mean SUS-PQ score for the sense of feeling present in the VE was M = 4.71 (SD = .87), which indicates a high sense of presence [Uso+99b]. The participants judged their fear to collide with physical obstacles during the experiment as comparably low (rating scale, 0 = no fear, 4 = high fear, M = 1.33, SD = 1.23)

5.3.1.3 Discussion

The results of the experiment show a significant influence of redirected walking on verbal as well as spatial working memory tasks, and we also found a significant effect of cognitive tasks on walking behavior. According to [Ste+10] a straight path in the VE can be turned into a circular arc in the real world with a radius of approximately 22 m, while users are still not able to reliably detect the manipulation. This corresponds to a curvature gain of $|g_C| = \frac{1}{22}$. Our experiments showed that a significant increase of lateral sway for both spatial task, as well as verbal task, starts at a gain of $|g_C| = \frac{1}{10}$. Furthermore, we also found that the task performance was decreased for both cognitive tasks again for gains larger than $|g_C| = \frac{1}{10}$. Hence, only at gains where users can detect the manipulation, cognitive task performance for spatial as well as verbal tasks decreases, and besides lateral sway increases when users are challenged with cognitive tasks.

These are important findings for the application of redirected walking techniques. The results show that only large redirected walking manipulations require additional cognitive resources by the user. Also, with increasing amounts of manipulations, the required cognitive resources also increase. Hence, VR developers can apply redirected walking below the detection thresholds of $|g_C| = \frac{1}{22}$ or below $|g_C| = \frac{1}{10}$, and we can assume that the amount of additionally required cognitive resources will not significantly increase.

On the other hand, if manipulation with curvature gains of $|g_C| > \frac{1}{10}$ are applied, users are clearly able to detect the manipulation (cf. [Ste+10]). But more importantly, for such curvature gains, cognitive task performance for spatial as well as verbal tasks decreases, and lateral sway also increases when users are challenged with cognitive tasks. Hence, one cannot recommend applying redirected walking in VR laboratories with a size below 10 m × 10 m in case users have to perform complex cognitive tasks.

5.3.1.4 Conclusion

In this section, we presented an experiment in which we evaluated the mutual influence of redirected walking and verbal as well as spatial working memory tasks in VR laboratories. We analyzed how curvature gains correlate with spatial and verbal working memory demands. The results of the experiment showed a significant influence of redirected walking on verbal as well as spatial working memory tasks, and we also found a significant effect of cognitive tasks on walking behavior. We discussed the implications and provided guidelines for using redirected walking in VR laboratories.

5.3.2 Safety Sphere

This section is based on work previously published in [LBS14c]. With the analysis of the cognitive demand of RDW, we decided to investigate user's safety concerns of RDW. During many of our RDW studies, participants informally commented on not feeling entirely safe while using fully-immersive HMDs. We believe that one critical feature of redirected walking interfaces is *safety*, i.e., inducing the feeling of being safe although the user is walking in the presence of physical obstacles

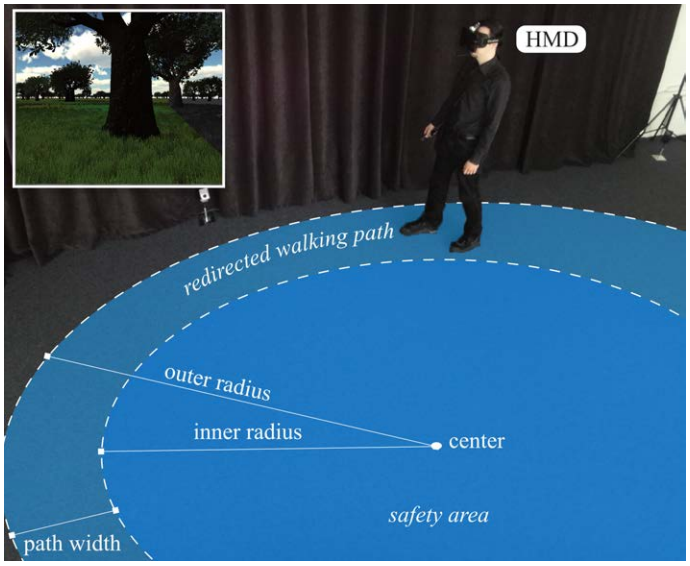


Figure 5.22: User exploring a VE using the Walking-on-Circles interface. The illustration shows the safety area and the redirected walking path on the ground. The inset shows the user's view with a visual barrier to his right side preventing him from walking into the wall.

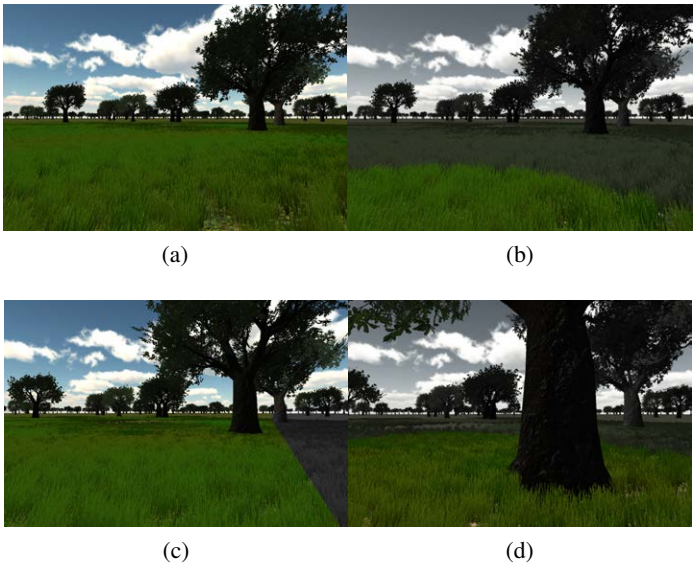
in the VR lab [Ste+13]. Another critical feature is *predictability*, i.e., transparent user interface mechanics that avoid unstable walking behavior and breaks in the user's sense of feeling present in the VE [Sla09].

A more extensive literature review on traveling in VR and especially RDW is described in Section 2.6.2. Different approaches have been presented that can be used to provide an environment safe from collisions with objects in the real world. For instance, interventions may be used, which may take the form of coloring the virtual view red in the presence of danger or providing auditive instructions to stop [Raz05]. Other less intruding approaches are based on dynamically appearing visual barriers that inform users of limits of the physical workspace, such as the magic barrier tape [Cir+09] or deterrents [PFW11]. These barriers or deterrents are objects in the VE that users are instructed to stay away from or not to cross. These barriers fade in as users near the edge of the workspace and fade out as users walk away from the edge. While this approach provides users with a visual cue about the size and orientation of the physical workspace in the VE relative to the user, practical tests usually show that users interpret them as *virtual* objects in the VE or as user interface elements rather than elements from the real world.

In this section, we address these requirements and propose the *Walking-on-Circles* user interface, which supports real walking in a potentially infinite virtual scene while confined to a small physical workspace. The technique relies on a safe movement volume (cf. Figure 5.22), which is visually indicated to users as a semi-transparent half-capsule (cf. Figures 5.23(b), 5.23(d)). While the user is within this safety volume, no RDW manipulations are applied, i.e., an object of interest can be explored without interferences caused by manipulated self-motion behavior. Once the user leaves the safety area, an RDW technique based on circular redirected walking is used to support infinite walking while the user is guided on a circular path with an optimal radius that leads around the inner area (cf. Figure 5.22). The technique is overt, as it clearly shows boundaries and the initial manipulation is noticeable [Sum+12].

5.3.2.1 Walking-on-Circles User Interface

We propose the Walking-on-Circles user interface as an improvement of RDW approaches in small labs. There are two fundamental objectives of the user interface:



(a) (b)

(c) (d)

Figure 5.23: User's view: (a) inside the safety volume, (b) inside (but close to the border of) the safety volume, (c) on the (counterclockwise curved) redirected walking path, and (d) inside the safety volume again. Note that the safety volume is visible in the background, giving the user also feedback about the distance to the border behind them and the size of the safe area.

- Provide a sense of safety by informing the user of the limits of the physical walking area from within the VE compellingly, without breaking the user's sense of presence.
- Provide an integrated locomotion technique that combines an unimpeded sense of walking in a safe area while a tailored RDW approach with transparent mechanics is used to reach any point of interest in the VE.

To address these requirements, we support walking in a safe area that is indicated by a virtual barrier, which represents the safe region of the physical workspace. For long-distance walking, we designed a circular RDW technique around this safe area. We integrated the two concepts, providing an all-walking integrated virtual locomotion interface that improves usability aspects of redirected walking interfaces in small physical workspaces.

Virtual Workspaces

When walking through a VE we distinguish between three stages in the Walking-on-Circles user interface: (i) walking in the safe inner region, (ii) redirected walking around the outer path, and (iii) a transition between these two stages.

Safe Inner Region Inside the safe inner region, movements in the physical workspace are mapped one-to-one to the virtual workspace, i.e., translations and rotations, are coherent between the real and virtual space. The boundaries of the virtual workspace are represented by a half-capsule, which is entirely transparent when the user is at least a step away (based on measured step length) from the boundaries of the safety region (cf. Figure 5.23(a)). The boundaries become progressively visible as the user gets closer (cf. Figure 5.23(b), 5.23(d)). We decided to use a desaturation of the VE as a visualization, which is clearly visible but does not entirely occlude the VE. This approach informs users of the available size of the safe area in which they can freely walk to explore or interact with elements of the VE that reach inside the safety region.

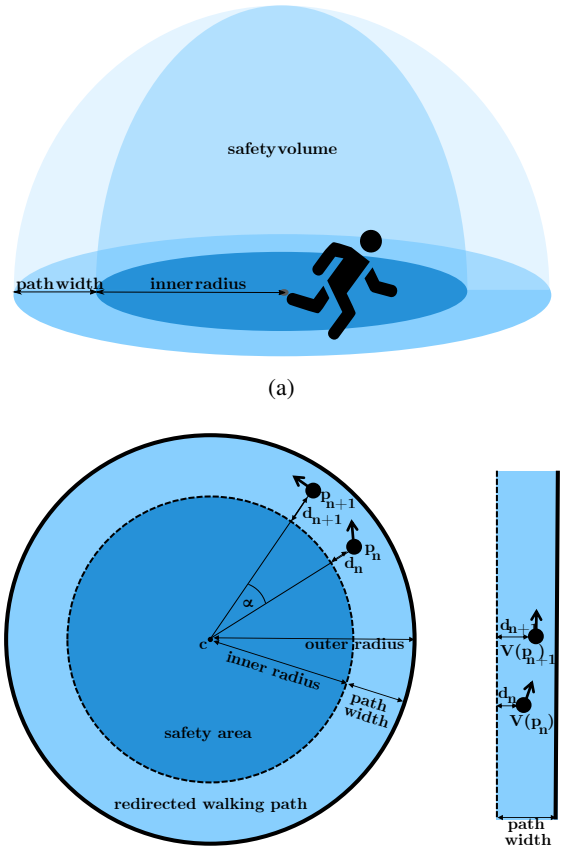


Figure 5.24: (a) Illustration of the safety volume and redirected walking path running around the inner region in the physical workspace, and (b) redirected walking mapping from the walking path in the real world (left) to the virtual straight walking path (right).

Transition When approaching the boundaries of the safe inner region, the desaturation factor of the barrier increases to inform users that they are about to leave the safety volume, and enter the redirected walking path. Right in front of the boundary, the user can still see the VE. However it will be wholly desaturated, meaning that the VE beyond the boundary appears in grayscale. When the user walks through the semi-transparent barrier, we start to apply camera rotations to guide the user on the path that leads around the inner safety area (cf. Figure 5.22, 5.24). Therefore, we compute the minimum angle we have to reorient users onto the path depending on which direction they were heading towards when exiting the safety region. By rotating the virtual camera using a linear transition from the moment users exit the inner region to the distance of half the width of the walking path, they are quickly able to continue walking in the desired direction.

Redirected Walking After users transitioned onto the redirected walking path, visual cues in the form of a virtual barrier are used to inform users that this path is located close to the boundaries of the workspace (cf. Figure 5.22, 5.23(c)). The barrier provides a visual deterrent for users not to move over that barrier, which is essential to avoid collisions considering that they are walking close to physical obstacles in the real world. The straight barrier is visualized as a wall, which desaturates the VE behind it, similarly to the inner sphere boundaries. Since the redirected walking path leads around the safety area, users are always redirected on a circular path with the maximum possible radius in the physical workspace, thus providing a near-constant and predictable magnitude of

manipulations.

The mapping from movements in tracking coordinates to virtual camera motions follows the *curvature gains* approach by Steinicke et al. [Ste+10] with a fixed circle radius. As illustrated in Figure 5.24 when the user's head position changes between frames $n \in \mathbb{N}$ and $n+1 \in \mathbb{N}$ from position $p_n \in \mathbb{R}^2$ to $p_{n+1} \in \mathbb{R}^2$, the mapped virtual position $V(p_{n+1}) \in \mathbb{R}^2$ is computed from the previous position $V(p_n) \in \mathbb{R}^2$ using movement components along the virtual path and in the orthogonal strafe direction.

We compute the movement l_{n+1} along the virtual path using the covered distance on the redirected walking path with the following equation:

$$l_{n+1} = \alpha \cdot r_{path} \cdot g_s \tag{5.1}$$

with r_{path} the circular redirected walking path radius and α the corresponding covered angle (in radians), i.e., the virtual covered distance is computed from the arc length of the physical path (cf. Figure 5.24). Additionally, we introduced a *speed gain* $g_s \in \mathbb{R}^+$, which can be freely chosen to scale translational movements, so that walked physical distances can be transferred to longer or shorter covered distances in the virtual world [Wil+06]. We apply this gain only to scale movements in the main movement direction. Hence, unintended lateral shifts can be prevented (cf. [IRA07]).

Strafe movements d_{n+1} orthogonal to the main walking direction are computed using the distance from the user's position to the center c of the circular workspace minus the radius r_{inner} of the inner region:

$$d_{n+1} = ||p_{n+1} - c||_2 - r_{inner} \tag{5.2}$$

To give the user the impression of walking straight in the VE, we rotate the virtual camera using the angle α that the user walked on the circle. Additionally, we found that introducing an additional time-dependent *steer-to-orbit* gain [Raz05] helped users slightly change their intended movement direction in the VE.

Once a user stops in front of an object of interest in the VE, we slowly start rotating the world around the user using a *steer-to-center* gain [Raz05], such that the region of interest moves into the inner safety region. Once the user takes the last steps towards the object, they can perform tasks within the virtual safe workspace at the new location.

Implementation

As shown in Figure 5.22 we used a wireless Oculus Rift DK1 HMD with an attached active infrared (IR) target. The target was tracked with an optical WorldViz Precision Position Tracking (PPT X4) system with sub-millimeter precision for position and orientation data in a small 6 m×6 m lab room. We used an inertial InertiaCube 4BT sensor for head orientation tracking. We used an Asus WAVI wireless transmitter box to transmit the images at 60 Hz from a rendering computer to the HMD. As claimed by the manufacturers, not more than 22 ms latency was introduced due to the wireless connection. An Anker Astro Pro2 portable battery powered the HMD and wireless transmitter box. The boxes were carried in a small belt bag. The VE was rendered with Unity3D Pro and our software on a MacBook Pro (2014) laptop with an NVIDIA GeForce GT 750M. As illustrated in Figure 5.23 the virtual world consisted of an outdoor scene with landmarks that were spatially separated and randomly distributed over the plain. The safety region was located in the center of our workspace with a 3 m diameter using 0.8 m wide redirected walking paths, which we based on standard shoulder widths (0.535 m [GK97]) plus a safety offset.

Considering that most VR labs have a rectangular workspace, it appears odd that we deliberately chose a circular inner region and outer path for the Walking-on-Circles user interface. However, it comes with specific benefits: First, RDW relies on angular manipulations which guide users on circular paths, i.e., we designed our interface in such a way that the maximum circle could be

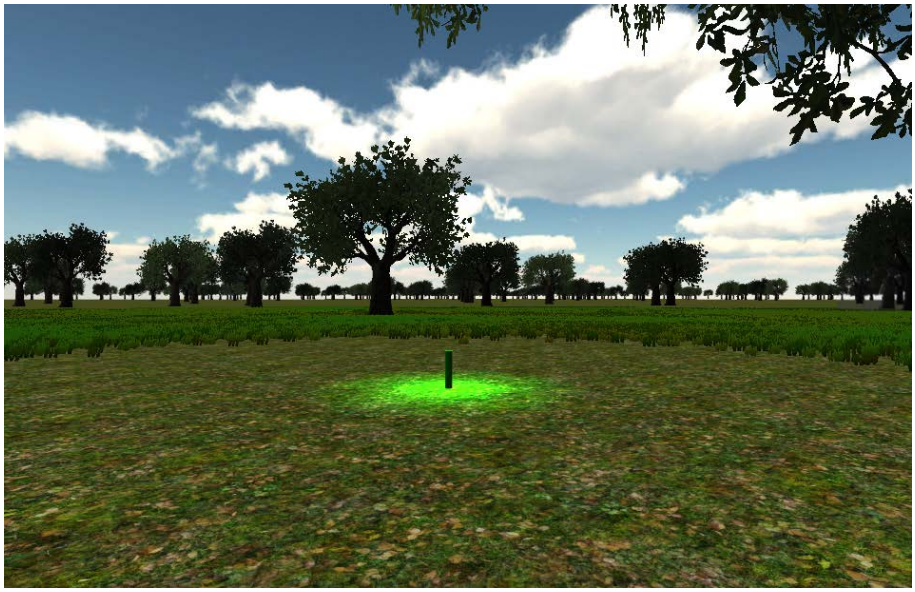


Figure 5.25: VE used during the user study. Users had to walk between points of interest indicated by the illuminated vertical cylinders, which were displayed at 5 m, 15 m, or 25 m distance from the user.

used, thus providing per definition optimal redirected walking performance. Second, the circular design does not require substantial dynamic changes in curvature gains and reduces the magnitude of manipulations during the transition from the inner region to the outer redirected walking path compared to rectangular shapes.

5.3.2.2 Usability Study

We evaluated the Walking-on-Circles user interface using the setup described in Section 5.3.2.1. In this study, we focused on the general usability of our technique rather than on the performance metrics in comparison to other approaches such as joystick-based navigation.

Participants

We recruited 9 participants for the evaluation, 6 male and 3 female (aged 22–45, M = 31.1). The participants were students or professionals of human-computer interaction or computer science. None of the participants reported known vision disorders or displacement of balance. 3 participants had prior experience with RDW with HMDs. The other participants were naive to IVEs.

Protocol

At the beginning of the user study, participants filled out a the SSQ [Ken+93] and a demographics questionnaire and were then immersed in the VE shown in Figure 5.23. Participants were instructed to explore 9 landmarks sequentially, of which only one was visible at any time. We tested 3 landmarks at a distance of 5 m from the user's position, 3 landmarks at a distance of 15 m, as well as 3 landmarks at a distance of 25 m. The landmarks were randomly distributed in the VE and tested in random order. Landmarks were indicated by virtual cylinders of 1 m height that were brightly illuminated by a spotlight to make them easier to spot from a distance (cf. Figure 5.25). The task of the participants was to walk to the currently visible landmark and touch the tip of the cylinder. Afterward, the cylinder vanished, and the next landmark appeared. After participants touched the last cylinder, we asked them to fill out an SSQ (cf. Section 2.4.2.4) and SUS-PQ [Uso+99b] (cf. Section 2.4.2.4), as well as a NASA TLX [Har06] (cf. Section 2.4.3) and AttrakDiff [HBK03] questionnaire.

Afterward, we performed a semi-structured interview with the participants, asking about the three main parts of the user interface mechanics described in Section 5.3.2.1, and giving them the opportunity to comment on the approach. The user study took about 30 minutes per participant.

Results

We measured a mean SSQ-score of M = 10.8 (SD = 16.5) before the user study, and a mean SSQ-score of M = 29.5 (SD = 30.2) after the user study. The increase in cybersickness is in line with results of typical RDW studies over the time of our user study [Ste+13]. The Walking-on-Circles locomotion user interface does not further increase cybersickness symptoms than other RDW techniques.

The mean SUS-PQ score for the reported sense of feeling present in the VE was M = 4.76 (SD = 1.28), which indicates a reasonably high level of presence [Uso+99b]. Participants judged their level of fear to collide with physical obstacles during the user study on a 5-point Likert scale (1 = no fear, 5 = strong fear) with M= 1.78 (SD = 0.97). Their judgments show that they felt reasonably safe in the IVE. Subjective feelings of safety and thus an unimpeded sense of presence in the VE were two of the major goals of the proposed technique.

The results of the NASA TLX questionnaire showed scores for mental demand (M = 49.4), physical demand (M = 53.9), temporal demand (M = 25.0), performance (M = 31.1), effort (M = 49.4), and frustration (M = 35.0) (cf. [Har06]). The results indicate that the mental demand of learning to use the interface was relatively high, as were the required physical demand and effort. On the other hand, the frustration of the participants was relatively low.

The results of the AttrakDiff questionnaire are shown in Figure 5.26. The results show that the *pragmatic* quality, i.e., an indication of the usability, of the user interface was reasonably high (M = 0.5), with room for improvement [HBK03]. The *hedonic* quality, i.e., the extent to which users may identify with the product or its support for subjective development and progress, also shows that the user is stimulated by the product, but shows room for improvement. Moreover, the *attractiveness* of the user interface was judged as M = 1.0, which shows that it was comparably attractive.

The semi-structured interview revealed that all participants judged the safety area approach as very useful once they understood that barriers provided visual feedback about potential collisions. Regarding the transitions between safety area and RDW path, all participants stated that they had to train it a few times before they felt safe transitioning between the areas, but they reported having no trouble doing it at the end of the user study. 4 of the participants stated that they felt that the transitions induced postural instability. For some trials, the participants actively reduced the amount of reorientation by leaving the inner area at an acute angle. When questioned about walking on the outer RDW path, all participants unanimously stated that they felt very safe due to the visual barrier, and only 3 of them even noticed to be redirected. In particular, the three professionals reported that they never walked so fast and felt so safe from collisions with RDW in the past. The virtual barrier enabled them to walk through the VE without having to focus as much on compensating for manipulations as in other RDW implementations.

5.3.2.3 Conclusion

In this section, we introduced the Walking-on-Circles locomotion user interface which integrates a safe workspace approach with RDW in small VR labs. We showed how user movements can be mapped from the physical workspace to the VE to enable natural exploration of regions of interest and redirected walking between locations at arbitrary distances in VEs. We described the design choices and characteristics of the user interface approach and discussed results in the scope of an implementation in our VR laboratory. We reported the results of a qualitative usability study of the user interface. The results suggest that the approach can improve the user's sense of feeling safe in the IVE during walking.

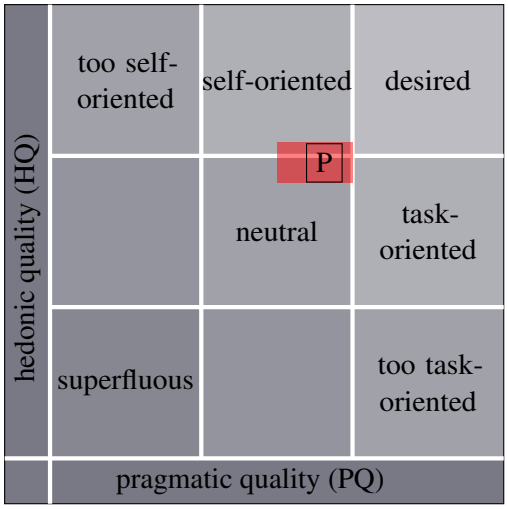


Figure 5.26: Results of the AttrakDiff questionnaire: Medium values for the dimensions pragmatic quality (PQ) and hedonic quality (HQ) and the confidence rectangle.

Since the first results suggest that the user interface may help to reduce the perceptual and cognitive demands of RDW interfaces, we plan to further evaluate the technique in different VEs such as indoor or more cluttered environments. Furthermore, we aim to extend the interface with passive haptic feedback approaches using registered real-world proxy objects in the safety region, such as a table or chair. Moreover, we aim to investigate multi-user scenarios, i.e., multiple users may group within the safety volume, while a user may lead the others to regions of interest via RDW.

5.4 Presence Through Peripheral Stimulation

The limited FOV is a current hardware limitation that is especially relevant for travel techniques. Peripheral vision in IVE is critical for application fields that require high spatial awareness and veridical impressions of three-dimensional spaces (cf. Section 2.4.2.1). However, HMDs use screens and optical elements in front of a user's eyes, which often do not natively support a wide FOV to stimulate the entire human visual field. Such limited visual angles are often identified as causes of reduced navigation performance and sense of presence (cf. Section 2.5).

In this section, we present an approach to extend the visual field of HMDs towards the periphery by incorporating additional optical LED elements structured in an array, which provide additional low-resolution information in the periphery of a user's eyes. We detail our approach, technical realization, and present an experiment, in which we show that such far peripheral stimulation can increase subjective estimates of presence, and has the potential to change user behavior during navigation in a VE.

HMDs can provide visual information of a VE over a considerable portion of the human visual field. For instance, the Oculus Rift DK1 HMD provides a visual field of 90° and the Fakespace Labs Wide5 HMD 150° horizontally. However, the currently used optics and flat display technologies do not support stimulation of human far peripheral vision [Ras+09]. The total horizontal binocular visual field of the human eyes approximates on average 200°, with the vertical field approximating 120° degrees [Pal99]. Restrictions of the FOV in the real world are known to affect human perception and change behavior, such as navigation and maneuvering performance [Has+07] and spatial estimation [PWP06] (cf. Section 2.4.2.1).

In the field of VR, stimulation of peripheral vision has shown to elicit more natural behavior,

perception and performance, i.e., reducing differences between IVEs and the real world [LK03]. However, providing visual information of a VE over the entire human visual field imposes challenges and difficulties not only on HMD display technologies but also on computer graphics rendering techniques based on planar geometric projections [Ras+09]. Moreover, recent display trends suggest that even simple extensions of ambient light in the periphery of screens [SVK07] or HMDs [TF12] might help users feeling more present in VEs.

In this section, we present a peripheral low-resolution visual field extension for immersive HMDs based on Light-Emitting Diodes (LEDs). We describe our approach as well as our hardware and software implementation. We present an evaluation of the effects of such peripheral stimulation and discuss the results.

5.4.1 Related Work

While most early HMD technologies only supported comparably small diagonal FOVs, i.e., about 30° to 60°, and current-state HMDs support medium peripheral stimulation of about 90°, e.g. Oculus Rift DK2, only a few HMDs were built that stimulate more, such as the Fakespace Labs Wide5 with 150° or Arthur's 176° HMD [Art96]. Neither of them encompassed the whole 200° of the human total horizontal visual field (cf. Section 2.5). Far peripheral vision is crucial due to a higher sensitivity to optic flow and motion detection than in foveal or near-peripheral vision, as well as higher detectability of temporal changes and an increased sensitivity to luminance in low-light situations, although visual resolution in such regions is reduced [Pal99] (cf. Section 2.4.2.1).

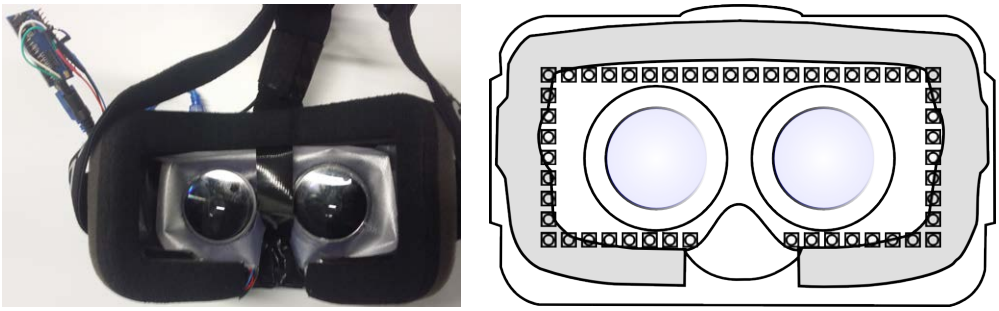
A large body of research has been focused on identifying effects of the FOV of HMDs on spatial perception, task performance, presence and behavior (cf. Section 2.4.2.1 and Section 2.5). For instance, Bolte et al. [Bol+10] found that a reduced FOV can change the user's behavior and increase the amount of necessary head rotation. Moreover, in an experiment conducted by Jones et al. [Jon+12] they found using a Wide5 HMD and emulating different FOVs that a 150° visual field significantly improved distance judgments in a VE compared to a reduced FOV of 60° degrees. They showed that even static peripheral stimulation with white light in the far periphery of HMDs resulted in more accurate distance and size judgments [JSB13]. Investigating how presence is affected by the visual field of virtual stimulation, experiments revealed a correlation between increased FOV and increased presence but also increased cybersickness [Lin+02].

5.4.2 Ambiculus Peripheral Display Extension

In this section, we first detail our conceptual approach of extending the FOV of current state-of-the-art HMDs and then specify our current implementation based on an Oculus Rift DK2. We call our display expansion Ambiculus, as a combination of ambient light and the Oculus Rift.

5.4.2.1 Concept

Our hardware concept for the Ambiculus peripheral display expansion is to mount arrays of RGB LEDs around the central display units of current-state HMDs (cf. [TF12]). Due to the predominance of flat rectangular screens used in recent HMDs, which are similar to or equal to cellphone-screens, we assume that the HMD has a rectangular shape in which optical magnification lenses for the left and right eye are mounted in front of the screen (cf. Section 2.5). This arrangement divides the screen into a left and right side (cf. Section 2.5). We place the LEDs in arrays with a width of 1 to n LEDs in rectangular shapes around the screen area and add diffusers above the LEDs to diffuse the light and to compensate for the spatially discrete light sources. Hence, users wearing the HMD see the VE on the main display in front of their eyes, and they receive light from the LEDs in the far periphery. As the peripheral light is not transmitted through the lenses of the HMD, it is not subject to the optical distortions usually found with lenses [Ras+09](cf. Section 2.5). By incorporating multiple layers of LEDs around the central display, it is possible to increase the



(a) (b)

Figure 5.27: a) The Oculus Rift DK2 with peripheral display extension prototype and diffusing foil. b) Schematic illustrating the LED placement.

peripheral resolution.

Usually, Ambilight setups in the home cinema market intercept the input image using video input signals and repeat visual information from the screen's edges towards the periphery using single LEDs or strips [Die+04]. We tested this approach by attaching LED stripes inside HMDs and noticed that it seemed to suffice to increase presence, which we observed anecdotally during informal pilot tests.

5.4.2.2 Realization

As an exemplary implementation of a low-resolution peripheral display, we decided to utilize RGB LEDs with an integrated WS2812B Controller. We implemented the peripheral light capture as a Unity3D plugin, and we used our protocol to transmit the rendered pixel data through a virtual serial port to an Arduino Nano board. The Arduino Nano board uses the Adafruit Neopixel library [Ada17] to transfer the color data to the LEDs (also called pixels). This allows the transmission of any signal and any color, which, for example, can include a warning light in case a user comes close to a collision with a physical obstacle in a head-tracked IVE (cf. Section 5.3.2).

In our current implementation, we used RGB LEDs from a strip with 144 LEDs per meter at 10 mm width with 5 V and 60 mA at maximum brightness. We included 50 LEDs (cf. Figure 5.27b) into the frame of an Oculus Rift DK2 HMD and used an external power supply to power the LEDs (cf. Figure 5.27a). Additionally, the scripts allow decoupling the LED output from the visual output, giving developers the possibility to display warnings or change the atmosphere of a scene with the illumination.

The current prototype supports a far peripheral visual output with a resolution corresponding to the number of LEDs in the grid (cf. Figure 5.27a), increasing the FOV to approximately 118°. The LEDs were diffused, and color values were optically calibrated to ensure the LED brightness does not surpass the HMD screen brightness.

5.4.3 Experiment

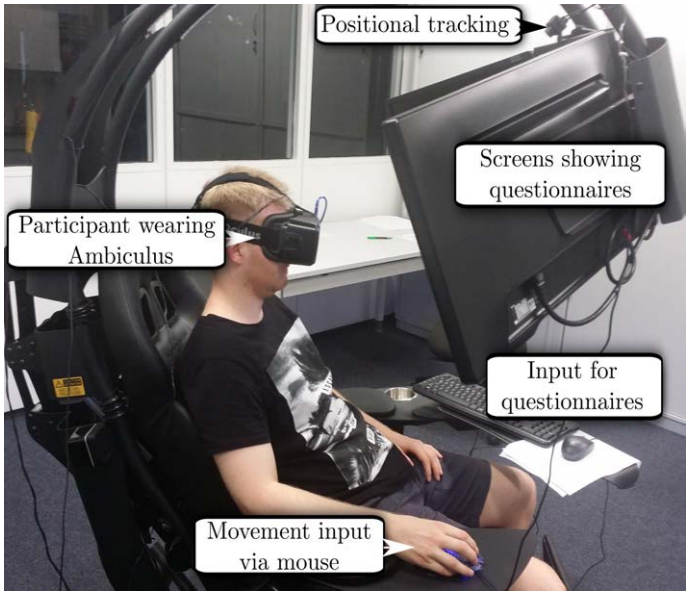
In this subsection, we describe the experiment that we conducted to evaluate the peripheral display extension described in Section 5.4.2. We address the following research questions:

Q$_1$ Can peripheral stimulation increase the user's sense of presence?

Q$_2$ Does light in the periphery change the user's behavior while navigating through a VE?

5.4.3.1 Participants

We recruited fifteen participants from our department for our experiment (2 female, ages 20–52, $M = 30.1$). All of them had normal or corrected vision. Seven subjects reported that they have



(a)

(b)

Figure 5.28: a) Photo of a participant during the experiment with annotations. b) The visual stimulus, including one of the gates the participants were instructed to truss.

participated in an HMD experiment before. The mean duration of the trials wearing the HMD was about ten minutes (30 min with questionnaire and instructions).

5.4.3.2 Materials

As illustrated in Figure 5.28a, users wore an Oculus Rift DK2 HMD using the Oculus DK2 IR camera to provide 6 DOF optical tracking (cf. Section 2.5). The participants were seated in an MWE Lab Emperor 1510 chair and used a computer mouse as input with their dominant hand. We implemented view-directed steering with speed controlled by the mouse. By pressing the left mouse button participants moved forward, the right mouse button moved them backward, and not pressing any button allowed them to stop and look around with head-tracking.

The visual stimulus consisted of a 3D forest scene in bright daylight (cf. Figure 5.28b), which was rendered with Unity3D on an Intel computer with a Core i7 hexacore at 3.5 GHz CPU, 32 GB RAM and two Nvidia GeForce GTX980 in an SLI array. We used the traditional edge-expansion approach to compute the ambient light for the LEDs as described in Section 5.4.2.

	LED	M	SD	$t(11)$	p
time	on	224.13 s	43.74 s	1.958	.076
	off	202.54 s	29.18 s		
distance	on	578.28 m	61.17 m	2.407	$< .05$
	off	541.12 m	14.15 m		
head translation	on	9.07 m	5.87 m	2.223	$< .05$
	off	6.33 m	3.39 m		
head rotation	on	4071.5°	2454.4°	2.009	.070
	off	3026.8°	489.8°		

Table 5.3: Quantitative experiment results.

5.4.3.3 Methods

Every second participant started with the peripheral LEDs turned on and then completed the second part without peripheral LEDs, and vice versa. To complete a trial, the participants were instructed to pass through ten gates in the virtual forest scene. Upon traversing the last gate, the trial was completed. The participants were informed that it was not a requirement to complete the task as quickly as possible. Since the participants were seated and not able to turn 360° comfortably, the gates were placed to allow the users to walk in a forward direction with a maximum of 45° head turning (cf. Figure 5.30). The dependent variables were the time needed to complete the trial by passing the final gate, the distance traveled in the virtual world, the head movements in meters, the overall angle of head rotations.

As a measure of presence and cybersickness, users were asked to fill in a set of subjective questionnaires (cf. Section 2.4.2.4). Before completing the trials, participants were asked to fill in SSQ and ITQ. Afterward, participants were instructed to put the HMD on and follow the on-screen task instructions. After reaching the final gate, participants had to take the HMD off and answer a post SSQ, a SUS-PQ as well as a WS-PQ. Then, participants had to put the HMD on again, and after the second trial, they had to fill in another post-SSQ, SUS-PQ, and WS-PQ, as well as demographic data. We also collected additional subjective impressions and comments.

5.4.3.4 Results

We had to remove one participant due to cybersickness and two participants because they did not understand the task and performed erratic movements since it was, as they commented, their first time wearing an HMD and they wanted to test the device limitations.

Behavioral Measures

We analyzed the results with a Paired-Samples T-Test at the 5% significance level (cf. Table 5.3).

Questionnaires

We found no significant difference in the average increase in SSQ scores during the trials with Ambiculus ($M = 13.09$, $SD = 34.74$) and without Ambiculus ($M = 2.49$, $SD = 8.18$) in the experiment ($Z = .65$, $p = .52$).

We analyzed the WS-PQ, SUS-PQ and SSQ questionnaires with a non-parametric Wilcoxon Signed Rank Test at the 5% significance level. The WS-PQ indicated that there was no statistically significant difference between the Ambiculus enabled ($M = 73.50$, $SD = 12.62$) and disabled ($M = 66.25$, $SD = 12.30$) presence scores, but a trend ($Z = 62$, $p = .07$). The SUS-PQ questionnaire indicated that there was no statistically significant difference between the Ambiculus enabled ($M = 4.94$, $SD = 1.23$) and disabled ($M = 3.99$, $SD = .96$) presence scores, but a trend ($Z = 62$, $p = .07$). We analyzed the ITQ with a non-parametric Mann-Whitney U Test but found no significant difference in ITQ scores (start without Ambiculus = 61, $SD = 11.40$; with Ambiculus = 57.86

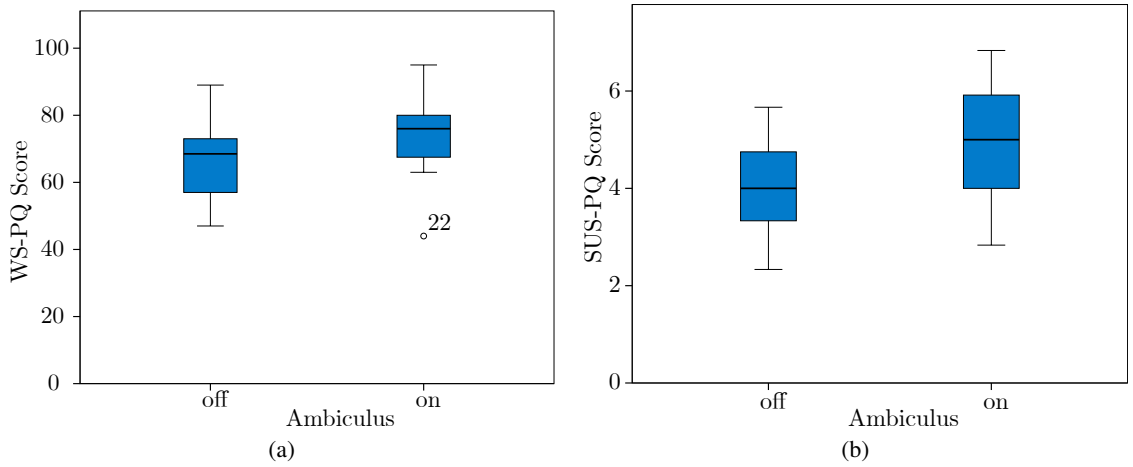


Figure 5.29: Results of the presence questionnaires: The y-axis shows the presence score and the x-axis shows the conditions. a) Witmer-Singer PQ. b) Slater-Usoh-Steed PQ.

Figure 5.30: The paths the participants walked in the experiment. (red) Ambiculus deactivated. (Green) Ambiculus active.

$SD = 23.34$; $U = 17.00$, $n_1 = 5$, $n_2 = 7$, $p = 1.00$).

5.4.3.5 Discussion

We collected additional informal comments, which overall indicated that participants enjoyed moving through the virtual forest more if the peripheral light was turned on and less if it was turned off. The quantitative behavioral results support this notion, in particular, considering the time spent in the virtual world and distance traveled and participants performed more movements with their head in the real world, i.e., translations and rotations. This is an indicator that the users' behavior is influenced by the added peripheral stimuli, as illustrated in Figure 5.30. When asked to reflect upon their behavior and indicate reasons after the experiment, some participants commented that they had a higher sense that the virtual world spanned entirely around their head such that it felt more natural to move around in it. Another participant mentioned that the peripheral light encouraged to look around and especially the difference between bright and dark places was very present and caused ambiance.

We assume that by adding even low resolution peripheral visual stimuli, the participants' desire to explore the scene was stimulated, likely due to the increased sense of presence. The presence questionnaire results support these comments and show a trend that the participants felt more present in the Ambiculus condition.

During the debriefing, two participants mentioned that the additional light in the periphery increased cybersickness symptoms. The SSQ questionnaire results support this statement, although the differences between both conditions were not significant, and have to be considered in future work. Shortly after this publication was presented, another work was published, wherein multiple layers of LEDs were included in an HMD, providing optic flow stimuli in the far periphery [XB16]. This work confirmed our results, showing that peripheral stimulation can increase presence. They also found subjective preference that a countervection shown in the periphery may reduce cybersickness.

5.4.4 Conclusion

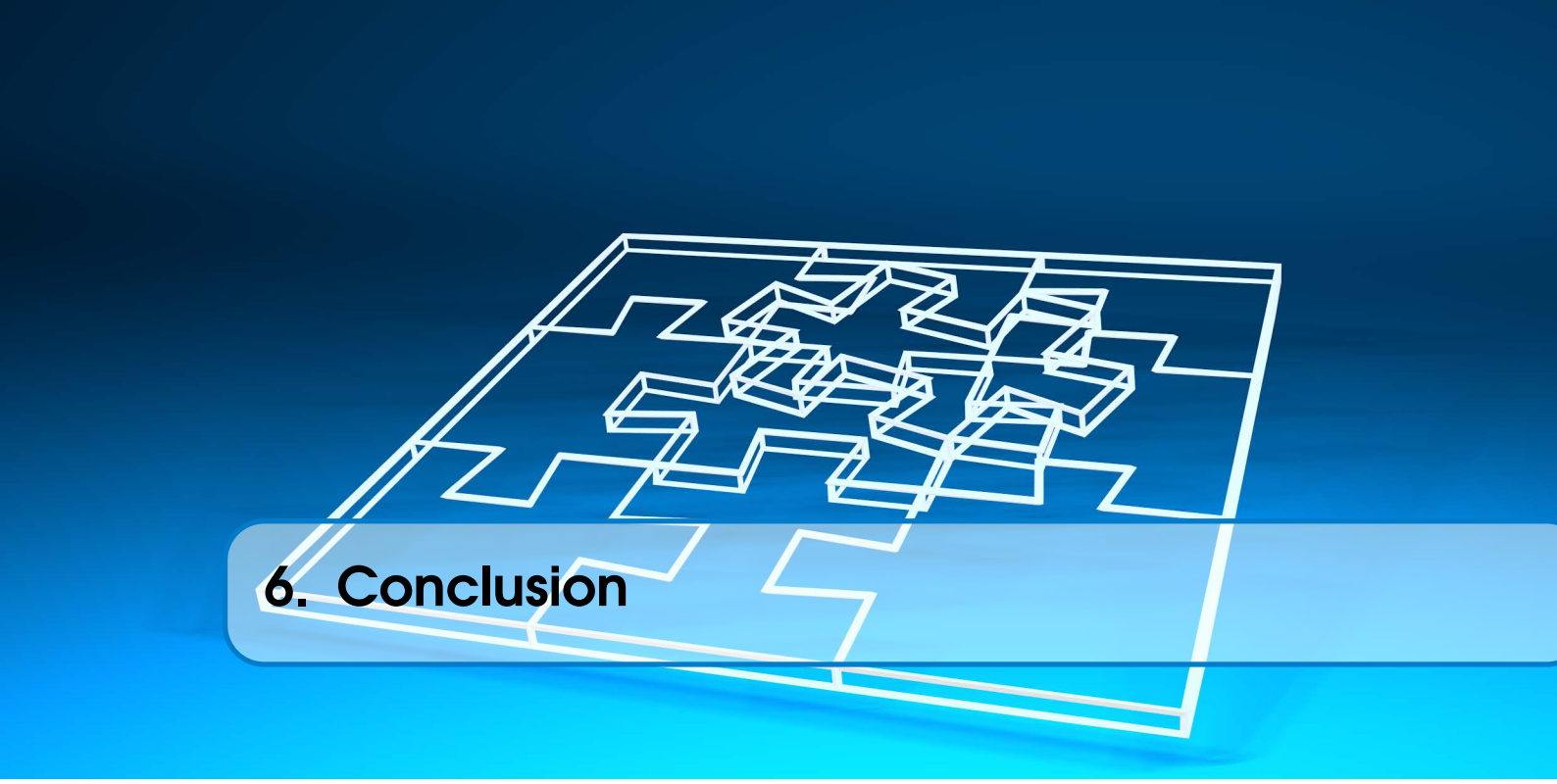
In this section, we presented an approach to provide peripheral visual stimulation for HMDs, which is based on RGB LEDs that are structured in an array around the main HMD screen. The implemented solution is intended to be open-source, allowing interested parties to use the plugin and sketches to replicate the device. Although the approach does not support a high resolution comparable to that in the central fields of an HMD, we found in an experiment that it nevertheless had benefits for subjective estimates of presence. Moreover, we observed that the peripheral light changed the behavior of users while navigating through a realistic virtual 3D scene.

In the described experiment we computed the peripheral light by repeating the border colors of the displayed images towards the periphery. The results are encouraging and we believe that low-resolution peripheral stimulation may be used for hints and to subtly lead people in IVEs.

5.5 Summary

In this chapter, we describe the travel techniques developed throughout this thesis. We start in Section 5.1 by a leaning-based virtual flight travel technique, the Dragon-Rider technique, which was inspired by supernatural creatures. The technique was evaluated positively, and we found out that the Kinect 2 was the preferable input device compared to the Wii Balance Board. Afterward, in Section 5.2, we present the development process of a flight technique and setup that requires users to be suspended in the air. In both iterations, users reported high levels of presence, however there were no clear preferences for regarding the steering methods. Considering all flying techniques caused simulator sickness, we investigated natural walking as a travel technique, as described in Section 5.3. All in all, natural walking techniques are often limited to the available tracking space, leading us to investigate redirected walking. We evaluate the cognitive demands of redirected walking, and then develop a travel technique that uses redirected walking to allow unlimited exploration of an environment. Within a central safe region, no redirection is applied, allowing users to explore naturally. Finally, in Section 5.4, we present our peripheral HMD extension, Ambiculus. In our experiment, we show that Ambiculus lets users explore VEs more naturally since their peripheral vision is stimulated using a low-resolution display.

Overall, the works presented in this chapter expand the knowledge of natural and supernatural travel techniques for fully-immersive VEs. We show that flight techniques can offer a high degree of presence, but possibly at a cost of increased simulator sickness. Whereas redirected walking techniques can cognitively distract users, unless done overtly.



6. Conclusion

While previous research in the field of HCI often focused on interfaces inspired by the way how humans interact with their environment and with each other, this thesis takes a step further by developing SNUIs and basic 3D interaction. However, the SNUIs introduced in this thesis are still inspired by natural interactions. Hence, they intentionally enable interaction with virtual objects in a way impossible within physical environments.

After establishing a common vocabulary, we placed the supernatural and basic interactions developed within this thesis into a historical context in Chapter 2. Following the historical background, we discussed the relevant fundamental knowledge in the field of Human Factors, presented the utilized hardware and finally discussed related research in the 3DUI field. We identified critical aspects of basic and supernatural 3D interaction that have not been sufficiently researched before.

In Chapter 3, we analyzed selection performance in semi-immersive and fully immersive VEs. Users are familiar with hover interactions from 2D UIs with a mouse. However, little has been known about hover behavior in stereoscopic, head-tracked environments. Thus, we developed perceptually-inspired hover volumes, called HoverSpace, allowing users of stereoscopic, optionally head-tracked displays to utilize hover interactions. Afterward, we analyzed fully-immersive environments and direct interaction with virtual objects to identify the optimal placement for virtual, interactive elements. Results have shown that distance misperception is still a serious problem, requiring placement of objects to be relatively close to the user's eyes. As a solution, we suggest ellipsoid selection volumes, similar to those used in the HoverSpace. However, during the experiments, we noticed fatigue caused by typical placement of 3DUI elements in front of the user's eyes.

Within Chapter 4, we presented our research that showed that fatigue is a relevant concern for 3DUIs and introduced solutions for this challenge. Therefore, we first analyzed the user's interaction space and compared their mid-air selection performance with and without an armrest as physical support. We showed that prolonged use of mid-air interactions causes degradation of performance and is easily alleviated by providing physical support. This result indicates that ergonomics need to be taken into account for the development of 3DUIs. Following that, we introduced Joint-Centered User Interfaces (JCUIs), a first step to position interface elements based on the biomechanics of the

human arm. In another experiment, we showed that JCUIs are comfortable, yet offer comparable performance to standard placement of UI elements. Finally, we described another solution using multiple hand pairs. We found that switching hand pairs to reduce the distance to a target can be a solid alternative to direct interaction, potentially decreasing fatigue.

Traveling may be necessary to reach interactive elements within a VE. In Chapter 5, we evaluated various steering metaphors for fully-immersive VR. We achieved high levels of presence but also showed the risk of increased simulator sickness, due to sensory conflicts. We investigated redirected walking (RDW), as it promises a more natural kind of movement within a VE. We found that RDW allows natural exploration, but causes significantly higher cognitive demand than walking without redirection. During the walking experiments, we noticed a mismatch between how people walk when fully-immersed compared to natural walking, possibly due to fear of colliding with virtual objects. Therefore, we presented the safe-and-round technique, an overt redirection technique that shows safe, redirection-free areas directly to the user. Finally, we investigated the hardware FOV limitations of contemporary VR HMDs and found out that peripheral, low-resolution stimulation is a way to increase presence in VR.

3DUI Design Recommendations

It is challenging to define golden rules or guidelines that apply to all 3DUIs, considering the full range of potential input and output devices that can be used for 3DUIs. In the following, we suggest recommendations for the design of 3DUIs, which we derived from the results of the research done within the scope of this thesis:

- A simple translation of 2D interactions, such as hovering, into 3D is not possible since mental models differ between users. If the technical means to support both mental models are available, offer them choices (cf. Chapter 3).
- Distance misperception is still a major challenge in IVEs. Interfaces should be designed without requiring users to correctly estimate the depth of an object, for example, by appropriately designing the depth of your interactive elements. Alternatively, use selection volumes with more depth, such as ellipsoid selection volumes, or other image-based approaches (cf. Chapter 3).
- When working with 3DUIs, interactions need not perfectly mimic interactions that humans do in a physical environment (cf. Chapter 4).
- Performance should not be the only metric to evaluate 3DUIs. In particular perceived exertion should be taken into account when developing applications for prolonged use (cf. Chapter 4). As such, an ergonomic work environment should be provided, e.g., by offering a seated position and armrests. The UI should incorporate these physical supports, e.g. by using JCUIs (cf. Chapter 4).
- Decoupling the virtual hands from the real hands, resulting in shorter distances between selections, can also allow users to stay in a more comfortable position and is preferred by users, despite higher cognitive demand and potentially slower selections (cf. Chapter 4).
- For short-term use, such as for demonstrations at exhibits, 3DUIs mimicking human interaction are a good choice, as they require only a short learning period (cf. Chapter 3).
- Leaning-based and suspension-based travel techniques can offer high presence but are often coupled with increased cybersickness. They should be used with caution (cf. Chapter 5).
- RDW techniques which redirect users on a circle with a radius less than 10 meters cause lateral sway and have a negative influence on cognitive tasks. Thus, such gains should be avoided (cf. Chapter 5).
- Having a central area without redirection, combined with an ideal redirection in a circle around the safe area, makes users walk less reluctantly (cf. Chapter 5).

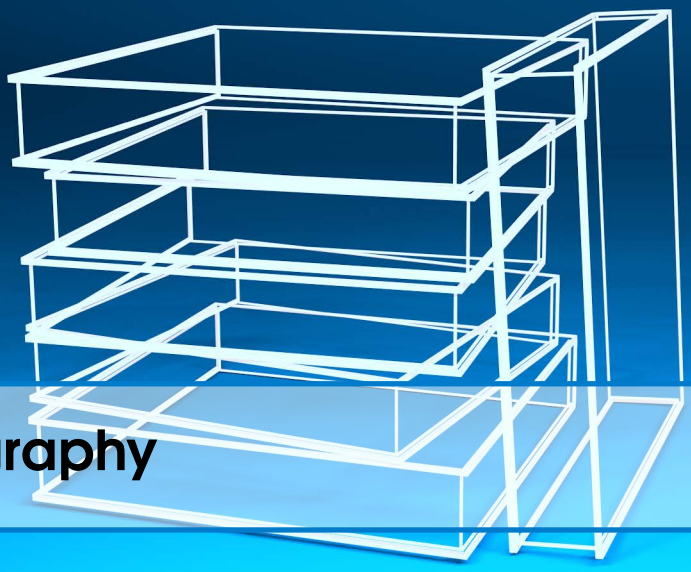
The focus of this thesis is selection in, and exploration of 3DUIs. Basic interactions in 3DUIs are analyzed, problems such as fatigue and distance misperception identified and natural and supernatural solutions are presented to solve them. Various natural and supernatural techniques of travel through are developed and evaluated, showing the limitations of steering techniques and redirected walking, as well as solutions to overcome these. The work in this thesis was mainly focused on (semi-)immersive VR and was evaluated as such.

Next Steps

However, the question arises whether these results can be used for other scenarios along Milgram's mixed reality continuum [MK94]. AR presents different challenges, and it has yet to be shown whether the findings in this thesis, especially concerning the supernatural exploration of virtual worlds, can be transferred to AR. The concept of exploring AR in the same way as VR seems unlikely, however, at least partially changing the virtual content in AR could be feasible.

The results of this thesis are based on the premise that humans have to move their physical body to interact with digital content, whether by leaning, moving their arm or a hand-held device. In the future, with the advent of brain-computer interfaces (BCIs), many of these concepts may have to be revisited [WW12]. For instance, if BCIs allowed a user to select specific virtual objects with a thought, it would create another supernatural selection method. Afterwards, the object would have to be manipulated, and it is unclear whether this will be possible using BCIs, or whether humans will prefer using their hands to manipulate objects. As such, SNUIs are likely to stay, similar to CLIs, which are still being used nowadays (cf. Chapter 2).

Besides that, using AI for UIs shows potential. The advancements in AI can be seen in voice assistance systems that manage to use the user's voice commands more and more reliably to control computer systems. Previous research has shown how voice recognition systems can be used in VR systems [Bol80]. UIs which learn from how their users use them and adapt accordingly, perhaps learning to automate repetitive or physically demanding tasks, could improve future 3DUIs. Further uses of AI for HCI could be evaluated in future research.



Bibliography

Articles

[Akk93] Robert A. Akka. “Utilizing 6D head-tracking data for stereoscopic computer-graphics perspective transformations”. In: volume 1915. 1993. DOI: 10.1117/12.157033. URL: http://dx.doi.org/10.1117/12.157033 (cited on page 46).

[Ann+11] Michelle Annett, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. “Medusa: A Proximity-aware Multi-touch Tabletop”. In: *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. UIST ’11. Santa Barbara, California, USA: ACM, 2011, pages 337–346. ISBN: 978-1-4503-0716-1. DOI: 10.1145/2047196.2047240. URL: http://doi.acm.org/10.1145/2047196.2047240 (cited on page 55).

[Aok+09] Takafumi Aoki, Hironori Mitake, Shoichi Hasegawa, and Makoto Sato. “Haptic Ring: Touching Virtual Creatures in Mixed Reality Environments”. In: *SIGGRAPH ’09: Posters*. SIGGRAPH ’09. New Orleans, Louisiana: ACM, 2009, 100:1–100:1. DOI: 10.1145/1599301.1599401. URL: http://doi.acm.org/10.1145/1599301.1599401 (cited on page 195).

[AA13] Ferran Argelaguet and Carlos Andujar. “A survey of 3D object selection techniques for virtual environments”. In: *Computers & Graphics* 37.3 (2013), pages 121–136. ISSN: 0097-8493. DOI: http://dx.doi.org/10.1016/j.cag.2012.12.003. URL: http://www.sciencedirect.com/science/article/pii/S0097849312001793 (cited on pages 51, 53, 78, 104).

[ALS15] Oscar Javier Ariza Nunez, **Paul Lubos**, and Frank Steinicke. “HapRing: A Wearable Haptic Device for 3D Interaction”. In: *Proceedings of Mensch und Computer*. Edited by Sarah Diefenbach, Niels Henze, and Martin Pielot. Berlin: De Gruyter Oldenbourg, 2015, pages 421–424 (cited on pages 41, 112, 114, 191, 195).

[Art96] Kevin Arthur. “Effects of Field of View on Task Performance with Head-mounted Displays”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’96. Vancouver, British Columbia, Canada: ACM, 1996, pages 29–30. ISBN: 0-89791-832-0. DOI: 10.1145/257089.257116. URL: http://doi.acm.org/10.1145/257089.257116 (cited on page 161).

[Azm+15] Mahdi Azmandian, Timofey Grechkin, Mark Bolas, and Evan Suma. “Physical Space Requirements for Redirected Walking: How Size and Shape Affect Performance”. In: *Proceedings of the 25th International Conference on Artificial Reality and Telexistence and 20th Eurographics Symposium on Virtual Environments*. ICAT - EGVE ’15. Kyoto, Japan: Eurographics Association, 2015, pages 93–100. ISBN: 978-3-905674-84-2. DOI: 10.2312/egve.20151315. URL: http://dx.doi.org/10.2312/egve.20151315 (cited on page 63).

[Azm+16] Mahdi Azmandian, Mark S. Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D. Wilson. “Haptic Retargeting: Dynamic Repurposing of Passive Haptics for Enhanced Virtual Reality Experiences”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’16. San Jose, California, USA: ACM, 2016, pages 1968–1979. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858226. URL: http://doi.acm.org/10.1145/2858036.2858226 (cited on page 34).

[Bad12] Alan D. Baddeley. “Working Memory: Theories, models, and controversies. ,” in: *Annual Review of Psychology* 63.1 (2012), pages 1–29 (cited on page 148).

[BH74] Alan D. Baddeley and Graham Hitch. “Working memory”. In: *Psychology of learning and motivation* 8 (1974), pages 47–89 (cited on page 148).

[BM97] Ravin Balakrishnan and I. Scott MacKenzie. “Performance Differences in the Fingers, Wrist, and Forearm in Computer Input Control”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’97. Atlanta, Georgia, USA: ACM, 1997, pages 303–310. ISBN: 0-89791-802-9. DOI: 10.1145/258549.258764. URL: http://doi.acm.org/10.1145/258549.258764 (cited on page 102).

[Bar09] Linsey Marinn Barker. “Measuring and modeling the effects of fatigue on performance: Specific application to the nursing profession”. PhD thesis. 2009 (cited on page 40).

[Bat+04] Lukas M. Batteau, Alan Liu, J. B. Antoine Maintz, Yogendra Bhasin, and Mark W. Bowyer. “A Study on the Perception of Haptics in Surgical Simulation”. In: *Medical Simulation: International Symposium, ISMS 2004, Cambridge, MA, USA, June 17-18, 2004. Proceedings*. Edited by Stéphane Cotin and Dimitris Metaxas. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pages 185–192. ISBN: 978-3-540-25968-8. DOI: 10.1007/978-3-540-25968-8_21. URL: https://doi.org/10.1007/978-3-540-25968-8_21 (cited on page 196).

[BRK07] Martin RK Baumann, Diana Rösler, and Josef F Krems. “Situation Awareness and Secondary Task Performance While Driving”. In: *Engineering Psychology and Cognitive Ergonomics: Lecture Notes in Computer Science (LNCS)* 4562 (2007), pages 256–263 (cited on page 151).

[Bea+05] Olivier Beauchet, Véronique Dubost, Régis Gonthier, and Reto W Kressig. “Dual-task-related gait changes in the elderly: does the type of cognitive task matter?” In: *Journal of Motor Behavior* 37 (2005), pages 259–264 (cited on page 148).

[BF07] Hrvoje Benko and Steve Feiner. “Balloon Selection: A Multi-Finger Technique for Accurate Low-Fatigue 3D Selection”. In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. Mar. 2007. DOI: 10.1109/3DUI.2007.340778 (cited on page 54).

[BWB06] Hrvoje Benko, Andrew D. Wilson, and Patrick Baudisch. “Precise selection techniques for multi-touch screens”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM. 2006, pages 1263–1272 (cited on pages 55, 191).

[Bér+09] François Bérard, Jessica Ip, Mitchel Benovoy, Dalia El-Shimy, Jeffrey R. Blum, and Jeremy R. Cooperstock. “Did “Minority Report” Get It Wrong? Superiority of the Mouse over 3D Input Devices in a 3D Placement Task”. In: *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT)*. Edited by Tom Gross, Jan Gulliksen, Paula Kotzé, Lars Oestreicher, Philippe Palanque, Raquel Oliveira Prates, and Marco Winckler. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pages 400–414. ISBN: 978-3-642-03658-3 (cited on pages 26, 93).

[Ber+16] Jeffrey Traer Bernstein, David T. Amm, Omar Leung, Christopher Tenzin Mullens, Brian Michael King, Brian Richards Land, and Reese T. Cutler. “Touch and hover sensing”. US Patent 9,323,398. Apr. 2016 (cited on page 55).

[Ber00] Alain Berthoz. *The Brain’s Sense of Movement*. Cambridge, Massachusetts: Harvard University Press, 2000 (cited on pages 34, 147).

[BIL00] René Bertin, Isabelle Israël, and Markus Lappe. “Perception of two-dimensional, simulated ego-motion trajectories from optic flow”. In: *Vision Research* 40.21 (2000), pages 2951–2971 (cited on page 34).

[BR03] Alan Blackwell and Kerry Rodden. *Preface to Sketchpad: A man-machine graphical communication system*. Technical report. University of Cambridge, 2003 (cited on page 21).

[Bli78] James F. Blinn. “Simulation of Wrinkled Surfaces”. In: SIGGRAPH ’78 (1978), pages 286–292. DOI: 10.1145/800248.507101. URL: http://doi.acm.org/10.1145/800248.507101 (cited on page 24).

[Bol80] Richard Bolt. “Put-That-There”. In: *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. 1980, pages 262–270 (cited on pages 52, 169).

[Bol+10] Benjamin Bolte, Gerd Bruder, Frank Steinicke, Klaus H. Hinrichs, and Markus Lappe. “Augmentation Techniques for Efficient Exploration in Head-Mounted Display Environments”. In: *Proceedings of ACM Symposium on Virtual Reality Software and Technology (VRST)*. 2010, pages 11–18 (cited on pages 14, 161).

[BSB11] Benjamin Bolte, Frank Steinicke, and Gerd Bruder. “The Jumper Metaphor: an Effective Navigation Technique for Immersive Display Setups”. In: *Proceedings of the ACM Virtual Reality International Conference (VRIC)*. 2011 (cited on page 63).

[Bor82] Gunnar A. Borg. “Psychophysical bases of perceived exertion”. In: *Med sci sports exerc* 14.5 (1982), pages 377–381 (cited on pages 40, 113, 200).

[Bor98] Gunnar A. Borg. *Borg’s perceived exertion and pain scales*. Human kinetics, 1998 (cited on page 40).

[Bou+02] Laroussi Bouguila, Makoto Sato, Shoichi Hasegawa, Hashimoto Naoki, Naoki Matsumoto, Atsushi Toyama, Jelel Ezzine, and Dalel Maghrebi. “A New Step-in-Place Locomotion Interface for Virtual Environment with Large Display System”. In: *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. ACM, 2002, pages 197–207 (cited on page 61).

[Bow99] Doug A. Bowman. “Interaction Techniques for Common Tasks in Immersive Virtual Environments: Design, Evaluation, and Application”. PhD thesis. Georgia Institute of Technology, 1999 (cited on pages 78, 102, 104).

[BH97] Doug A. Bowman and Larry F. Hodges. “An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments”. In: *ACM SIGGRAPH Symposium on Interactive 3D Graphics*. 1997, pages 35–38 (cited on pages 51, 52).

[BH99] Doug A. Bowman and Larry F. Hodges. “Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments”. In: *Journal of Visual Languages and Computing* 10.1 (1999), pages 37–53 (cited on pages 47, 49, 65).

[BKH97] Doug A. Bowman, David Koller, and Larry F. Hodges. “Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control Techniques”. In: *Proceedings of IEEE Virtual Reality Annual International Symposium (VRAIS)*. Volume 7. IEEE, 1997, pages 45–52 (cited on pages 56, 57, 61).

[Bow+04] Doug A. Bowman, Ernst Kruijff, Joseph J LaViola Jr, and Ivan P Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional, 2004 (cited on page 24).

[Bow+01] Doug A. Bowman, Chadwick Wingrave, Joshua Campbell, and Vinh Ly. “Using pinch gloves (tm) for both natural and abstract interaction techniques in virtual environments”. In: (2001) (cited on pages 39, 95).

[Boz+16] Evren Bozgeyikli, Andrew Raij, Srinivas Katkoori, and Rajiv Dubey. “Point & Teleport Locomotion Technique for Virtual Reality”. In: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. CHI PLAY ’16. Austin, Texas, USA: ACM, 2016, pages 205–216. ISBN: 978-1-4503-4456-2. DOI: 10.1145/2967934.2968105. URL: http://doi.acm.org/10.1145/2967934.2968105 (cited on pages 58, 61).

[BR96] Mark F. Bradshaw and Brian J. Rogers. “The Interaction of Binocular Disparity and Motion Parallax in the Computation of Depth”. In: *Vision Research* 36.21 (1996), pages 3457–3468. ISSN: 0042-6989. DOI: http://dx.doi.org/10.1016/0042-6989(96)00072-7. URL: http://www.sciencedirect.com/science/article/pii/0042698996000727 (cited on page 31).

[Bro+96] John Brooke et al. “SUS-A quick and dirty usability scale”. In: *Usability evaluation in industry* 189.194 (1996), pages 4–7 (cited on pages 113, 120).

[Bro94] Frederick P. Brooks Jr. “Is There Any Real Virtue In Virtual Reality?” In: *public lecture cosponsored by the Royal Academy of Engineering and the British Computer Society, London* 30 (1994) (cited on pages 12, 20).

[Bro99] Frederick P. Brooks Jr. “What’s real about virtual reality?” In: *IEEE Computer graphics and applications* 19.6 (1999), pages 16–27 (cited on pages 21, 26, 46).

[Bro+90] Frederick P. Brooks Jr., Ming Ouh-Young, James J. Batter, and P. Jerome Kilpatrick. “Project GROPEHaptic Displays for Scientific Visualization”. In: *SIGGRAPH Computer Graphics* 24.4 (Sept. 1990), pages 177–185. ISSN: 0097-8930. DOI: 10.1145/97880.97899. URL: http://doi.acm.org/10.1145/97880.97899 (cited on page 12).

[Bro+10] Johnell O. Brooks, Richard R. Goodenough, Matthew C. Crisler, Nathan D. Klein, Rebecca L. Alley, Beatrice L. Koon, William C. Logan, Jennifer H. Ogle, Richard A. Tyrrell, and Rebekkah F. Wills. “Simulator sickness during driving simulation studies”. In: *Accident Analysis & Prevention* 42.3 (2010), pages 788–796 (cited on page 35).

[Bro58] John Brown. “Some tests of the decay theory of immediate memory”. In: *Quarterly Journal of Experimental Psychology* 10.1 (1958), pages 12–21 (cited on page 36).

[BLS15] Gerd Bruder, **Paul Lubos**, and Frank Steinicke. “Cognitive Resource Demands of Redirected Walking”. In: *IEEE Transactions on Visualization and Computer Graphics* 21.4 (Apr. 2015), pages 539–544. ISSN: 1077-2626. DOI: 10.1109/TVCG.2015.2391864 (cited on pages 13, 15, 127, 147, 148).

[BSH09] Gerd Bruder, Frank Steinicke, and Klaus H. Hinrichs. “Arch-Explore: A Natural User Interface for Immersive Architectural Walkthroughs”. In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. 2009, pages 75–82 (cited on page 63).

[BSS13a] Gerd Bruder, Frank Steinicke, and Wolfgang Stuerzlinger. “Effects of Visual Conflicts on 3D Selection Task Performance in Stereoscopic Display Environments”. In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. 2013, pages 115–118 (cited on pages 12, 13, 52, 78, 81, 86, 95).

[BSS13b] Gerd Bruder, Frank Steinicke, and Wolfgang Stuerzlinger. “To Touch or not to Touch? Comparing 2D Touch and 3D Mid-Air Interaction on Stereoscopic Tabletop Surfaces”. In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. ACM Press, 2013, pages 1–8 (cited on pages 65, 66, 69).

[BSS13c] Gerd Bruder, Frank Steinicke, and Wolfgang Stuerzlinger. “Touching the Void Revisited: Analyses of Touch Behavior On and Above Tabletop Surfaces”. In: *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT)*. 2013 (cited on pages 12, 13, 52–54, 65, 66, 78, 81, 86, 95, 192, 195).

[BSW11] Gerd Bruder, Frank Steinicke, and Phil Wieland. “Self-motion illusions in immersive virtual reality environments”. In: *Proceedings of the IEEE Virtual Reality conference (VR)*. Mar. 2011, pages 39–46. DOI: 10.1109/VR.2011.5759434 (cited on pages 33, 61).

[Bry99] Steve Bryson. “Virtual reality: A definition history”. In: *Omnibus Lexicon Definition Supplement* (1999) (cited on pages 20, 21).

[BKB97] Hans-Jörg Bullinger, Peter Kern, and Martin Braun. “Controls”. In: *Handbook of Human Factors and Ergonomics* (1997), pages 697–728 (cited on page 39).

[BC03] Grigore C. Burdea and Phillipe Coiffet. *Virtual Reality Technology*. Wiley-IEEE Press, 2003 (cited on pages 21, 40).

[Bur+05] Eric Burns, Sharif Razzaque, Abigail T. Panter, Mary C. Whitton, Matthew R. McCallus, and Frederick P. Brooks Jr. “The Hand is Slower than the Eye: A quantitative exploration of visual dominance over proprioception”. In: *Proceedings of the IEEE Virtual Reality conference (VR)*. IEEE, 2005, pages 3–10 (cited on page 34).

[Bux90] William A. S. Buxton. “A Three-state Model of Graphical Input”. In: *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT)*. INTERACT ’90. Amsterdam, The Netherlands, The Netherlands: North-Holland Publishing Co., 1990, pages 449–456. ISBN: 0-444-88817-9. URL: http://dl.acm.org/citation.cfm?id=647402.725582 (cited on pages 54, 55, 66).

[Bux10] William A. S. Buxton. “A Touching Story: A Personal Perspective on the History of Touch Interfaces Past and Future”. In: *SID Symposium Digest of Technical Papers* 41.1 (2010), pages 444–448. ISSN: 2168-0159. DOI: 10.1889/1.3500488. URL: http://dx.doi.org/10.1889/1.3500488 (cited on page 23).

[Can+82] David Canfield Smith, Charles Irby, Ralph Kimball, and Eric Harslem. “Designing the Star user interface”. In: *Byte* 7 (1982), pages 242–282 (cited on page 22).

[CMR91] Stuart K. Card, Jock D. Mackinlay, and George G. Robertson. “A Morphological Analysis of the Design Space of Input Devices”. In: *ACM Transactions Inf. Syst.* 9.2 (Apr. 1991), pages 99–122. ISSN: 1046-8188. DOI: 10.1145/123078.128726. URL: http://doi.acm.org/10.1145/123078.128726 (cited on page 101).

[CMN86] Stuart K. Card, Thomas P. Moran, and Allen Newell. “The Model Human Processor- An Engineering Model of Human Performance”. In: *Handbook of perception and human performance.* 2 (1986), pages 45–1 (cited on pages 27, 28).

[Cat74] Edwin Catmull. *A subdivision algorithm for computer display of curved surfaces*. Technical report. Utah University Salt Lake City School of Computing, 1974 (cited on page 24).

[C+99] Don B. Chaffin, Gunnar Andersson, Bernard J. Martin, et al. *Occupational biomechanics*. Wiley New York, 1999 (cited on page 38).

[Cha+10] Li-Wei Chan, Hui-Shan Kao, Mike Y Chen, Ming-Sui Lee, Jane Hsu, and Yi-Ping Hung. “Touching the Void: Direct-Touch Interaction for Intangible Displays”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. 2010, pages 2625–2634 (cited on pages 12, 13, 52, 78, 195).

[CH02] Yam San Chee and Chit Meng Hooi. “C-VISions: Socialized Learning Through Collaborative, Virtual, Interactive Simulations”. In: *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community*. CSCL ’02. Boulder, Colorado: International Society of the Learning Sciences, 2002, pages 687–696. URL: http://dl.acm.org/citation.cfm?id=1658616.1658789 (cited on page 59).

[CMS88] Michael Chen, S. Joy Mountford, and Abigail Sellen. “A Study in Interactive 3-D Rotation Using 2-D Control Devices”. In: *SIGGRAPH Computer Graphics* 22.4 (June 1988), pages 121–129. ISSN: 0097-8930. DOI: 10.1145/378456.378497. URL: http://doi.acm.org/10.1145/378456.378497 (cited on page 25).

[CC91] Roger W. Cholewiak and Amy A. Collins. “Sensory and physiological bases of touch”. In: *The psychology of touch* (1991), pages 23–60 (cited on page 196).

[Chu93] James Che-Ming Chung. “Intuitive Navigation in the Targeting of Radiation Therapy Treatment Beams”. UMI Order No. GAX94-15297. PhD thesis. Chapel Hill, NC, USA: University of North Carolina at Chapel Hill, 1993 (cited on page 58).

[Cir+09] Gabriel Cirio, Maud Marchal, Tony Regia-Corte, and Anatole Lécuyer. “The Magic Barrier Tape: A Novel Metaphor for Infinite Navigation in Virtual Worlds with a Restricted Walking Workspace”. In: *Proceedings of ACM Symposium on Virtual Reality Software and Technology (VRST)*. ACM Press, 2009, pages 155–162 (cited on page 154).

[Cob+99] Sue V.G. Cobb, Sarah Nichols, Amanda Ramsey, and John R. Wilson. “Virtual reality-induced symptoms and effects (VRISE)”. In: *Presence: teleoperators and virtual environments* 8.2 (1999), pages 169–186 (cited on page 35).

[Col96] Thomas S. Collett. “Vision: Simple Stereopsis”. In: *Current Biology* 6.11 (1996), pages 1392–1395 (cited on page 31).

[Cre+15] Sarah H. Creem-Regehr, Jeanine K. Stefanucci, William B. Thompson, Nathan Nash, and Michael McCardell. “Egocentric Distance Perception in the Oculus Rift (DK2)”. In: *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*. SAP ’15. Tübingen, Germany: ACM, 2015, pages 47–50. ISBN: 978-1-4503-3812-7. DOI: 10.1145/2804408.2804422. URL: http://doi.acm.org/10.1145/2804408.2804422 (cited on page 44).

[Cru+92] Carolina Cruz-Neira, Daniel J Sandin, Thomas A DeFanti, Robert V Kenyon, and John C Hart. “The CAVE, Audio Visual Experience Automatic Virtual Environment”. In: *Communications of the ACM* 35.6 (1992), pages 64–72 (cited on pages 12, 33).

[Cue17] Eduardo Cuervo. “BEYOND REALITY: Head-Mounted Displays for Mobile Systems Researchers”. In: *GetMobile: Mobile Comp. and Comm.* 21.2 (Aug. 2017), pages 9–15. ISSN: 2375-0529. DOI: 10.1145/3131214.3131218. URL: http://doi.acm.org/10.1145/3131214.3131218 (cited on page 41).

[CJ84] Bruce G. Cumming and Stuart J. Judge. “Neural Mechanisms of Convergence and Accommodation”. In: *Advances in Psychology* 22 (1984), pages 429–437 (cited on page 32).

[CV95] James E. Cutting and Peter M. Vishton. “Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth”. In: *W. Epstein & S. Rogers (Eds.)* Vol 5; Perception of Space and Motion (1995), pages 69–117 (cited on page 33).

[DFK12] Florian Daiber, Eric Falk, and Antonio Krüger. “Balloon Selection Revisited: Multi-touch Selection Techniques for Stereoscopic Data”. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*. AVI ’12. Capri Island, Italy: ACM, 2012, pages 441–444. ISBN: 978-1-4503-1287-5. DOI: 10.1145/2254556.2254641. URL: http://doi.acm.org/10.1145/2254556.2254641 (cited on page 54).

[DMR12] Marvin Dainoff, Wayne Maynard, and Michelle Robertson. “Office Ergonomics”. In: *Handbook of Human Factors and Ergonomics*. Edited by Gavriel Salvendy. John Wiley & Sons, 2012, pages 192–242 (cited on page 23).

[Dee98] Michael F. Deering. “The limits of human vision”. In: *2nd International Immersive Projection Technology Workshop*. Volume 2. 1998 (cited on pages 30, 43).

[Deu11] Deutsche Gesetzliche Unfallversicherung e.V. *DGUV Regel 112-198 Benutzung von persönlichen Schutzausrüstungen gegen Absturz*. 2011 (cited on page 133).

[Dib+16] Christian Dibbern, Julian Hettwer, Josephine Höltermann, Daniel Sinn, Sukhpreet Singh, **Paul Lubos**, and Gerd Bruder. “Virtual Reality Flight Interfaces inspired by Iron Man”. In: *Proceedings of the GI Workshop on Virtual and Augmented Reality (GI VR/AR)*. Edited by Thies Pfeiffer, Julia Fröhlich, and Rolf Kruse(Hg.) Aachen: Shaker Verlag, 2016, pages 97–108 (cited on pages 13, 15, 127, 132).

[DB78] Johannes Dichgans and Thomas Brandt. “Visual Vestibular Interaction: Effects on Self-Motion Perception and Postural Control”. In: *Perception. Handbook of Sensory Physiology*. Edited by R. Held, H.W. Leibowitz, and H.L. Teuber. Volume 8. Berlin, Heidelberg, New York: Springer, 1978, pages 755–804 (cited on pages 34, 147).

[Die+04] Elmo M. A. Diederiks, Erwin R. Meinders, Edwin Van Lier, and Ralph H. Peters. “Method of and system for controlling an ambient light and lighting unit”. WO Patent App. PCT/IB2003/002,875. Jan. 2004 (cited on page 162).

[Dix09] Alan Dix. “Human-Computer Interaction”. In: *Encyclopedia of Database Systems*. Edited by LING LIU and M. TAMER ÖZSU. Boston, MA: Springer US, 2009, pages 1327–1331. ISBN: 978-0-387-39940-9. DOI: 10.1007/978-0-387-39940-9_192. URL: https://doi.org/10.1007/978-0-387-39940-9_192 (cited on page 19).

[Dix+00] Melissa W. Dixon, Maryjane Wraga, Dennis R Proffitt, and George C. Williams. “Eye height scaling of absolute size in immersive and nonimmersive displays.” In: *Journal of Experimental Psychology: Human Perception and Performance* 26.2 (2000), page 582 (cited on page 46).

[Dja98] Johan Partomo Djajadiningrat. “Cubby: what you see is where you act. Interlacing the display and manipulation spaces.” In: (1998) (cited on page 52).

[Dör+14] Ralf Dörner, Wolfgang Broll, Paul Grimm, and Bernhard Jung. *Virtual und augmented reality (VR/AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität*. Springer-Verlag, 2014 (cited on pages 30, 31, 33).

[DKM99] Sarah A Douglas, Arthur E Kirkpatrick, and I Scott MacKenzie. “Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM. 1999, pages 215–222 (cited on page 78).

[DKK07] Assaf Y. Dvorkin, Robert V. Kenyon, and Emily A. Keshner. “Reaching within a dynamic virtual environment”. In: *Journal of NeuroEngineering and Rehabilitation* 4.23 (2007). DOI: doi:10.1186/1743-0003-4-23 (cited on pages 50, 81, 95).

[EHK08] Florian Echtler, Manuel Huber, and Gudrun Klinker. “Shadow tracking on multi-touch tables”. In: *Proceedings of the working conference on Advanced visual interfaces*. ACM. 2008, pages 388–391 (cited on page 55).

[End12] Mica R Endsley. “Situation Awareness”. In: *Handbook of Human Factors and Ergonomics*. Edited by Gavriel Salvendy. John Wiley & Sons, 2012, pages 553–568 (cited on page 36).

[Eng88] D. Engelbart. “The Augmented Knowledge Workshop”. In: *A History of Personal Workstations*. Edited by Adele Goldberg. New York, NY, USA: ACM, 1988. Chapter The Augmented Knowledge Workshop, pages 185–248. ISBN: 0-201-11259-0. DOI: 10.1145/61975.66918. URL: http://doi.acm.org/10.1145/61975.66918 (cited on page 21).

[EEB67] William K. English, Douglas C. Engelbart, and Melvyn L. Berman. “Display-Selection Techniques for Text Manipulation”. In: *IEEE Transactions on Human Factors in Electronics* HFE-8.1 (1967), pages 5–15 (cited on pages 21, 22).

[Far94] Sadeg M. Faris. “Novel 3D stereoscopic imaging technology”. In: *Stereoscopic Displays and Virtual Reality Systems*. Volume 2177. 1994. DOI: 10.1117/12.173875. URL: http://dx.doi.org/10.1117/12.173875 (cited on page 44).

[Fis+12] Martin Fischbach, Marc E. Latoschik, Gerd Bruder, and Frank Steinicke. “smARTbox: out-of-the-box technologies for interactive art and exhibition”. In: *Proceedings of the ACM Virtual Reality International Conference (VRIC)*. Laval, France: ACM, 2012, 19:1–19:7. ISBN: 978-1-4503-1243-1. DOI: 10.1145/2331714.2331737. URL: http://doi.acm.org/10.1145/2331714.2331737 (cited on pages 47, 191).

[Fis+87] Scott S Fisher, Micheal McGreevy, James Humphries, and Warren Robinett. “Virtual Environment Display System”. In: *Proceedings of the 1986 Workshop on Interactive 3D Graphics*. I3D ’86. Chapel Hill, North Carolina, USA: ACM, 1987, pages 77–87. ISBN: 0-89791-228-4. DOI: 10.1145/319120.319127. URL: http://doi.acm.org/10.1145/319120.319127 (cited on pages 11, 45).

[Fit54] Paul Morris Fitts. “The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement”. In: *Journal of Experimental Psychology: General* 47.6 (1954), pages 381–391 (cited on pages 49, 50, 86, 116, 118).

[FP67] Paul Morris Fitts and Michael I Posner. “Human performance.” In: (1967) (cited on page 29).

[FK86] Edna B. Foa and Michael J. Kozak. “Emotional processing of fear: exposure to corrective information.” In: *Psychological bulletin* 99.1 (1986), page 20 (cited on page 28).

[FHZ96] Andrew Forsberg, Kenneth Herndon, and Robert Zeleznik. “Aperture based Selection for Immersive Virtual Environments”. In: *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. 1996, pages 95–96 (cited on pages 52, 53).

[Fre+07] Harald Frenz, Markus Lappe, Marina Kolesnik, and Thomas Bührmann. “Estimation of Travel Distance from Visual Motion in Virtual Environments”. In: *ACM Transactions on Applied Perception (TAP)* 4.1 (3) (2007), pages 419–428 (cited on page 33).

[Frö+99] Bernd Fröhlich, Stephen Barrass, Björn Zehner, John Plate, and Martin Göbel. “Exploring Geo-Scientific Data in Virtual Environments”. In: *IEEE Proceedings of Visualization*. 1999, pages 169–173 (cited on page 26).

[Frö+95] Bernd Fröhlich, Gernoth Grunst, Wolfgang Krüger, and Gerold Wesche. “The Responsive Workbench: A Virtual Working Environment For Physicians”. In: *Computers in biology and medicine* 25.2 (1995), pages 301–308 (cited on page 46).

[Frö+05] Bernd Fröhlich, Jan Hochstrate, Jörg Hoffmann, Karsten Klüger, Roland Blach, Matthias Bues, and Oliver Stefani. “Implementing Multi-Viewer Stereo Displays”. In: *13th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision* 13 (2005), pages 139–146 (cited on page 192).

[Frö+06] Bernd Fröhlich, Jan Hochstrate, Verena Skuk, and Anke Huckauf. “The GlobeFish and the GlobeMouse: Two New Six Degree of Freedom Input Devices for Graphics Applications”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’06. 3D input devices explanation isotonic elastic isometric - position and rate control. Montréal, Québec, Canada: ACM, 2006, pages 191–199. ISBN: 1-59593-372-7. DOI: 10.1145/1124772.1124802. URL: http://doi.acm.org/10.1145/1124772.1124802 (cited on pages 25, 26).

[Frö+00] Bernd Fröhlich, John Plate, Jürgen Wind, Gerold Wesche, and M Gobel. “Cubic-Mouse-Based Interaction in Virtual Environments”. In: *IEEE Computer Graphics and Applications* 20.4 (2000), pages 12–15 (cited on pages 12, 26).

[Fur86] Thomas A. Furness III. “The Super Cockpit and its Human Factors Challenges”. In: *Proceedings of the Human Factors Society Annual Meeting* 30.1 (1986), pages 48–52. DOI: 10.1177/154193128603000112. URL: https://doi.org/10.1177/154193128603000112 (cited on page 12).

[GC93] Alan S. Gevins and Brian A. Cutillo. “Neuroelectric evidence for distributed processing in human working memory”. In: *Electroencephalography and Clinical Neurophysiology* 87 (1993), pages 128–143 (cited on pages 37, 148, 151).

[GVH14] Alexander Giesler, Dimitar Valkov, and Klaus H. Hinrichs. “Void Shadows: Multi-touch Interaction with Stereoscopic Objects on the Tabletop”. In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI ’14. Honolulu, Hawaii, USA: ACM, 2014, pages 104–112. ISBN: 978-1-4503-2820-3. DOI: 10.1145/2659766.2659779. URL: http://doi.acm.org/10.1145/2659766.2659779 (cited on page 54).

[Glo+14] Daniel Glomberg, Daniel Kirchhof, Okan Köse, Fabian Schöndorff, Marcel Tiator, Roman Wiche, and Christian Geiger. “ZeroGravity - eine virtuelle Nutzererfahrung in Luft und Wasser”. In: *Mensch & Computer 2014 - Workshopband*. Edited by Andreas Butz, Michael Koch, and Johann Schlichter. Berlin: De Gruyter Oldenbourg, 2014, pages 049–052 (cited on page 133).

[Got11] Gregory Goth. “Brave NUI World”. In: *Communications of the ACM* 54.12 (Dec. 2011), pages 14–16. ISSN: 0001-0782. DOI: 10.1145/2043174.2043181. URL: http://doi.acm.org/10.1145/2043174.2043181 (cited on page 140).

[GK97] Etienne Grandjean and Karl H. E. Kroemer. *Fitting the task to the human: a textbook of occupational ergonomics*. CRC press, 1997 (cited on pages 38, 87, 157).

[Gro+05] Henning Groenda, Fabian Nowak, Patrick Rößler, and Uwe D Hanebeck. “Telepresence Techniques for Controlling Avatar Motion in First Person Games”. In: *Intelligent Technologies for Interactive Entertainment (INTETAIN 2005)*. 2005, pages 44–53 (cited on page 147).

[GB04] Tovi Grossman and Ravin Balakrishnan. “Pointing at trivariate targets in 3D environments”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM. 2004, pages 447–454 (cited on page 51).

[Gro+06] Tovi Grossman, Ken Hinckley, Patrick Baudisch, Maneesh Agrawala, and Ravin Balakrishnan. “Hover widgets: using the tracking state to extend the capabilities of pen-operated devices”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM. 2006, pages 861–870 (cited on page 55).

[Gui87] Yves Guiard. “Asymmetric Division of Labor in Human Skilled Bimanual Action”. In: *Journal of Motor Behavior* 19.4 (1987). PMID: 15136274, pages 486–517. DOI: 10.1080/00222895.1987.10735426. URL: https://doi.org/10.1080/00222895.1987.10735426 (cited on page 39).

[Gui+15] Darren Guinness, Alvin Jude, G. Michael Poor, and Ashley Dover. “Models for Rested Touchless Gestural Interaction”. In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI ’15. Los Angeles, California, USA: ACM, 2015, pages 34–43. ISBN: 978-1-4503-3703-8. DOI: 10.1145/2788940.2788948. URL: http://doi.acm.org/10.1145/2788940.2788948 (cited on pages 102, 108).

[Gun+14] Jože Guna, Grega Jakus, Matevž Pogačnik, Sašo Tomažič, and Jaka Sodnik. “An Analysis of the Precision and Reliability of the Leap Motion Sensor and Its Suitability for Static and Dynamic Tracking”. In: *Sensors* 14.2 (2014), pages 3702–3720. ISSN: 1424-8220. DOI: 10.3390/s140203702. URL: http://www.mdpi.com/1424-8220/14/2/3702 (cited on page 42).

[Gup+13] Sidhant Gupta, Dan Morris, Shwetak N. Patel, and Desney Tan. “AirWave: Non-contact Haptic Feedback Using Air Vortex Rings”. In: *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp ’13. Zurich, Switzerland: ACM, 2013, pages 419–428. ISBN: 978-1-4503-1770-2. DOI: 10.1145/2493432.2493463. URL: http://doi.acm.org/10.1145/2493432.2493463 (cited on page 195).

[HW10] Taejin Ha and Woontack Woo. “An empirical evaluation of virtual hand techniques for 3D object manipulation in a tangible augmented reality environment”. In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2010, pages 91–98 (cited on page 51).

[Hac+11] Martin Hachet, Benoit Bossavit, Aurélie Cohé, and Jean-Baptiste de la Rivière. “Toucheo: multitouch and stereo combined in a seamless workspace”. In: *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. ACM, 2011, pages 587–592 (cited on page 191).

[Hae+16] Steffen Haesler, Karen Obernesser, Tino Raupp, Christoph Jahnke, Jonathan Stapf, Julia Bräker, **Paul Lubos**, Gerd Bruder, and Frank Steinicke. “Edutainment & Engagement at Exhibitions: A Case Study of Gamification in the Historic Hammaburg Model”. In: *Mensch und Computer 2016-Tagungsband* (2016) (cited on page 59).

[Hal15] John E Hall. *Guyton and Hall Textbook of Medical Physiology E-Book*. Elsevier Health Sciences, 2015 (cited on page 37).

[Hal97] Michael Halle. “Autostereoscopic displays and computer graphics”. In: *ACM SIGGRAPH Computer Graphics* 31.2 (1997), pages 58–62 (cited on page 38).

[HP12] Seungju Han and Joonah Park. “A study on touch & hover based interaction for zooming”. In: *CHI’12 Extended Abstracts on Human Factors in Computing Systems*. ACM. 2012, pages 2183–2188 (cited on page 55).

[HCC07] Mark S. Hancock, Sheelagh Carpendale, and Andy Cockburn. “Shallow-depth 3D Interaction: Design and Evaluation of One-, Two- and Three-touch Techniques”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’07. San Jose, California, USA: ACM, 2007, pages 1147–1156. ISBN: 978-1-59593-593-9. DOI: 10.1145/1240624.1240798. URL: http://doi.acm.org/10.1145/1240624.1240798 (cited on page 53).

[Han+06] Mark S. Hancock, Frédéric D. Vernier, Daniel Wigdor, Sheelagh Carpendale, and Chia Shen. “Rotation and translation mechanisms for tabletop interaction”. In: *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP ’06)*. Jan. 2006. DOI: 10.1109/TABLETOP.2006.26 (cited on pages 24, 53).

[Han+17] Jeffrey T. Hansberger, Chao Peng, Shannon L. Mathis, Vaidyanath Areyur Shanthakumar, Sarah C. Meacham, Lizhou Cao, and Victoria R. Blakely. “Dispelling the Gorilla Arm Syndrome: The Viability of Prolonged Gesture Interactions”. In: *Virtual, Augmented and Mixed Reality*. Edited by Stephanie Lackey and Jessie Chen. Cham: Springer International Publishing, 2017, pages 505–520. ISBN: 978-3-319-57987-0 (cited on page 101).

[Har06] Sandra G. Hart. “NASA-Task Load Index (NASA-TLX) 20 years later”. In: *Proceedings of Human Factors and Ergonomics Society Annual Meeting*. 2006, pages 904–908 (cited on pages 37, 113, 120, 158, 159, 199).

[Has+07] Shirin E. Hassan, John C. Hicks, Hao Lei, and Kathleen A. Turano. “What is the minimum field of view required for efficient navigation?” In: *Vision Research* 47.16 (2007), pages 2115–2123 (cited on page 160).

[HBK03] Marc Hassenzahl, Michael Burmester, and Franz Koller. “AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität”. In: *Proceedings of Mensch & Computer*. Springer, 2003, pages 187–196 (cited on pages 74, 158, 159).

[Hei62] Morton L. Heilig. “Sensorama Simulator.(1962)”. In: *United States Patent and Trade Office, Virginia, USA, US-3,050,870* (1962) (cited on page 26).

[Hv12] Susan J. Hespos and Kristy van Marle. “Physics for infants: Characterizing the origins of knowledge about objects, substances, and number”. In: *Wiley Interdisciplinary Reviews: Cognitive Science* 3.1 (Jan. 2012), pages 19–27. ISSN: 1939-5078. DOI: 10.1002/wcs.157 (cited on page 12).

[Hil+09a] Otmar Hilliges, Shahram Izadi, Andrew D. Wilson, Steve Hodges, Armando Garcia-Mendoza, and Andreas Butz. “Interactions in the air: adding further depth to interactive tabletops”. In: *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. ACM. 2009, pages 139–148 (cited on pages 55, 66).

[Hil+09b] Andre Hilsendeger, Stephan Brandauer, Julia Tolksdorf, and Christian Fröhlich. “Navigation in virtual reality with the Wii balance board”. In: *Proceedings of the GI Workshop on Virtual and Augmented Reality (GI VR/AR)*. 2009, pages 1–10 (cited on page 60).

[Hin+14] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. “Consumed Endurance: A Metric to Quantify Arm Fatigue of Mid-air Interactions”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’14. Toronto, Ontario, Canada: ACM, 2014, pages 1063–1072. ISBN: 978-1-4503-2473-1. DOI: 10.1145/2556288.2557130. URL: http://doi.acm.org/10.1145/2556288.2557130 (cited on page 40).

[HH93] Deborah Hix and H. Rex Hartson. *Developing User Interfaces: Ensuring Usability Through Product and Process*. New York, NY, USA: John Wiley & Sons, Inc., 1993. ISBN: 0-471-57813-4 (cited on page 19).

[Hof+08] David M Hoffman, Ahna R Girshick, Kurt Akeley, and Martin S Banks. “Vergence–Accommodation Conflicts Hinder Visual Performance and Cause Visual Fatigue”. In: *Journal of vision* 8.3 (2008), pages 33–33 (cited on pages 33, 89).

[Hof+13] Marina Hofmann, Ronja Bürger, Ninja Frost, Julia Karremann, Jule Keller-Bacher, Stefanie Kraft, Gerd Bruder, and Frank Steinicke. “Comparing 3D Interaction Performance in Comfortable and Uncomfortable Regions”. In: *Proceedings of the GI-Workshop VR/AR, 3-14*. 2013 (cited on pages 52, 53, 90, 92, 93, 95).

[Höl+99] Tobias Höllerer, Steven Feiner, Tachio Terauchi, Gus Rashid, and Drexel Hallaway. "Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system". In: *Computers & Graphics* 23.6 (1999), pages 779–785 (cited on page 60).

[How91] Ian P. Howard. "Spatial vision within egocentric and exocentric frames of reference". In: edited by Stephen Ellis. Taylor & Francis, Inc., 1991, pages 338–358 (cited on page 37).

[How12] Ian P. Howard. *Perceiving in Depth: Other mechanisms of depth perception*. Oxford psychology series. Oxford University Press, 2012. ISBN: 9780199764167. URL: https://books.google.de/books?id=SFgfgFrdB%5C_oC (cited on pages 30–32).

[HCW15] Fu-Chung Huang, Kevin Chen, and Gordon Wetzstein. "The light field stereoscope: immersive computer graphics via factored near-eye light field displays with focus cues". In: *ACM Transactions on Graphics (TOG)* 34.4 (2015), page 60 (cited on page 43).

[Hur+89] Jan Hurst, Michael S. Mahoney, John T. Gilmore, Lawrence G. Roberts, and R. Forrest. "Retrospectives I: the early years in computer graphics at MIT, Lincoln Lab, and Harvard". In: *ACM SIGGRAPH Computer Graphics* 23.5 (1989), pages 19–38 (cited on page 24).

[HHN85] Edwin L. Hutchins, James D. Hollan, and Donald A. Norman. "Direct Manipulation Interfaces". In: *Human-Computer Interaction* 1.4 (Dec. 1985), pages 311–338. ISSN: 0737-0024. DOI: 10.1207/s15327051hci0104_2. URL: http://dx.doi.org/10.1207/s15327051hci0104_2 (cited on page 20).

[IRA06] Victoria Interrante, Brian Ries, and Lee Anderson. "Distance Perception in Immersive Virtual Environments, revisited". In: *Proceedings of the IEEE Virtual Reality conference (VR)*. IEEE, 2006, pages 3–10 (cited on pages 33, 37).

[IRA07] Victoria Interrante, Brian Ries, and Lee Anderson. "Seven League Boots: A New Metaphor for Augmented Locomotion through Moderately Large Scale Immersive Virtual Environments". In: *Proceedings of IEEE Symposium on 3D User Interfaces*. IEEE, 2007, pages 167–170 (cited on pages 61, 62, 157).

[Jae+02] Philip M. Jaekl, Robert S. Allison, Laurence R. Harris, Urszula T. Jasiobedzka, Heather L. Jenkin, Michael R. Jenkin, James E. Zacher, and Daniel C. Zikovitz. "Perceptual stability during head movement in virtual reality". In: *Proceedings of the IEEE Virtual Reality conference (VR)*. IEEE, 2002, pages 149–155 (cited on page 34).

[Jan+17a] Omar Janeh, Eike Langbehn, Frank Steinicke, Gerd Bruder, Alessandro Gulberti, and Monika Poetter-Nerger. "Walking in Virtual Reality: Effects of Manipulated Visual Self-Motion on Walking Biomechanics". In: *ACM Transactions Applied Perception* 14.2 (Jan. 2017), 12:1–12:15. ISSN: 1544-3558. DOI: 10.1145/3022731. URL: http://doi.acm.org/10.1145/3022731 (cited on page 146).

[Jan+17b] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. "Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion". In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI '17. Denver, Colorado, USA: ACM, 2017, pages 3328–3339. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025523. URL: http://doi.acm.org/10.1145/3025453.3025523 (cited on page 14).

[Jav02] Bahram Javidi. *Three-Dimensional Television, Video and Display Technology*. Edited by Fumio Okano. Berlin, Heidelberg: Springer-Verlag, 2002. ISBN: 3540435492 (cited on page 46).

[Jer+08] Jason Jerald, Tabitha Peck, Frank Steinicke, and Mary Whitton. "Sensitivity to Scene Motion for Phases of Head Yaws". In: *Proceedings of Applied Perception in Graphics and Visualization*. ACM, 2008, pages 155–162 (cited on pages 34, 35).

[Joh67] Eric Arthur Johnson. "Touch displays: A programmed man-machine interface". In: *Ergonomics* 10.2 (1967), pages 271–277 (cited on page 23).

[Joh+89] Jeff Johnson, Teresa L. Roberts, William Verplank, David Canfield Smith, Charles H. Irby, Marian Beard, and Kevin Mackey. "The Xerox Star: A Retrospective". In: *Computer* 22.9 (1989), pages 11–26 (cited on page 22).

[Jon+12] James Adam Jones, Evan A. Suma, David M. Krum, and Mark Bolas. "Comparability of Narrow and Wide Field-Of-View Head-Mounted Displays for Medium-Field Distance Judgments". In: *Proceedings of Symposium on Applied Perception (SAP)*. ACM, 2012, pages 119–119 (cited on pages 12, 14, 30, 44, 161).

[JSB13] James Adam Jones, J. Edward Swan, II, and Mark Bolas. "Peripheral Stimulation and its Effect on Perceived Spatial Scale in Virtual Environments". In: *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 19.4 (2013), pages 701–710 (cited on pages 14, 45, 161).

[Jon+08] James Adam Jones, J. Edward Swan,II, Gurjot Singh, Eric Kolstad, and Stephen R. Ellis. "The Effects of Virtual Reality, Augmented Reality, and Motion Parallax on Egocentric Depth Perception". In: *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*. APGV '08. Los Angeles, California: ACM, 2008, pages 9–14. ISBN: 978-1-59593-981-4. DOI: 10.1145/1394281.1394283. URL: http://doi.acm.org/10.1145/1394281.1394283 (cited on page 33).

[JPG14] Alvin Jude, G. Michael Poor, and Darren Guinness. "Personal Space: User Defined Gesture Space for GUI Interaction". In: *Proceedings of the Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI EA '14. Toronto, Ontario, Canada: ACM, 2014, pages 1615–1620. ISBN: 978-1-4503-2474-8. DOI: 10.1145/2559206.2581242. URL: http://doi.acm.org/10.1145/2559206.2581242 (cited on page 102).

[Kel+12] Falko Kellner, Benjamin Bolte, Gerd Bruder, Ulrich Rautenberg, Frank Steinicke, Markus Lappe, and Reinhard Koch. "Geometric Calibration of Head-Mounted Displays and its Effects on Distance Estimation". In: *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 18.4 (2012), pages 589–596 (cited on page 30).

[KCS17] Jonathan W. Kelly, Lucia A. Cherep, and Zachary D. Siegel. “Perceived Space in the HTC Vive”. In: *ACM Transactions Applied Perception* 15.1 (July 2017), 2:1–2:16. ISSN: 1544-3558. DOI: 10.1145/3106155. URL: http://doi.acm.org/10.1145/3106155 (cited on page 44).

[Ken+93] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. “Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness”. In: *The International Journal of Aviation Psychology* 3.3 (1993), pages 203–220 (cited on pages 35, 82, 96, 106, 120, 150, 158).

[Kir58] Wayne K. Kirchner. “Age differences in short-term retention of rapidly changing information.” In: *Journal of experimental psychology* 55.4 (1958), page 352 (cited on page 37).

[Kir52] Lincoln Kirstein. *The Classic Ballet: Basic Technique and Terminology*. University Press of Florida, 1952 (cited on pages 38, 39).

[KI08] Masatomo Kobayashi and Takeo Igarashi. “Ninja cursors: using multiple cursors to assist target acquisition on large screens”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’08. Florence, Italy: ACM, 2008, pages 949–958. ISBN: 978-1-60558-011-1. DOI: 10.1145/1357054.1357201. URL: http://doi.acm.org/10.1145/1357054.1357201 (cited on page 115).

[Kop+10] Regis Kopper, Doug A. Bowman, Mara G. Silva, and Ryan P. McMahan. “A human motor behavior model for distal pointing tasks”. In: *International journal of human-computer studies* 68.10 (2010), pages 603–615 (cited on pages 50, 52).

[KGH85] Myron W. Krueger, Thomas Gionfriddo, and Katrin Hinrichsen. “VIDEOPLACE - an artificial reality”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. 1985, pages 35–40 (cited on page 12).

[Krü+95] Wolfgang Krüger, Christian A. Bohn, Bernd Fröhlich, Heinrich Schüth, Wolfgang Strauss, and Gerold Wesche. “The Responsive Workbench: A Virtual Work Environment”. In: *IEEE Computer Society* 28.7 (1995), pages 42–48 (cited on page 46).

[Kru+16] Ernst Kruijff, Alexander Marquardt, Christina Trepkowski, Robert W. Lindeman, Andre Hinkenjann, Jens Maiero, and Bernhard E. Riecke. “On Your Feet!: Enhancing Vection in Leaning-Based Interfaces Through Multisensory Stimuli”. In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI ’16. Tokyo, Japan: ACM, 2016, pages 149–158. ISBN: 978-1-4503-4068-7. DOI: 10.1145/2983310.2985759. URL: http://doi.acm.org/10.1145/2983310.2985759 (cited on pages 34, 60).

[Kru+15a] Ernst Kruijff, Bernhard E. Riecke, Christina Trekowski, and Alexandra Kitson. “Upper Body Leaning Can Affect Forward Self-Motion Perception in Virtual Environments”. In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI ’15. Los Angeles, California, USA: ACM, 2015, pages 103–112. ISBN: 978-1-4503-3703-8. DOI: 10.1145/2788940.2788943. URL: http://doi.acm.org/10.1145/2788940.2788943 (cited on pages 34, 60, 127).

[Kru+15b] Dennis Krupke, **Paul Lubos**, Lena Demski, Jonas Brinkhoff, Gregor Weber, Fabian Willke, and Frank Steinicke. “Evaluation of Control Methods in a Supernatural Zero-Gravity Flight Simulator”. In: *Proceedings of the GI-Workshop VR/AR*. 2015 (cited on pages 13, 15, 132).

[Kur93] Gordon Paul Kurtenbach. “The design and evaluation of marking menus.” PhD thesis. University of Toronto, 1993 (cited on page 111).

[Kur+97] Gordon Paul Kurtenbach, George Fitzmaurice, Thomas Baudel, and William A. S. Buxton. “The design of a GUI paradigm based on tablets, two-hands, and transparency”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM. 1997, pages 35–42 (cited on page 23).

[Kus14] David Kushner. “Virtual reality’s moment”. In: *IEEE Spectrum* 51.1 (Jan. 2014), pages 34–37. ISSN: 0018-9235. DOI: 10.1109/MSPEC.2014.6701429 (cited on page 12).

[Lan+15] Eike Langbehn, Tobias Eichler, Sobin Ghose, Kai von Luck, Gerd Bruder, and Frank Steinicke. “Evaluation of an Omnidirectional Walking-in-Place User Interface with Virtual Locomotion Speed Scaled by Forward Leaning Angle”. In: *Proceedings of the GI Workshop on Virtual and Augmented Reality (GI VR/AR)*. 2015, pages 149–160 (cited on page 61).

[Lan+17] Eike Langbehn, **Paul Lubos**, Gerd Bruder, and Frank Steinicke. “Bending the Curve: Sensitivity to Bending of Curved Paths and Application in Room-Scale VR”. In: *IEEE Transactions on Visualization and Computer Graphics* 23.4 (Apr. 2017), pages 1389–1398. ISSN: 1077-2626. DOI: 10.1109/TVCG.2017.2657220 (cited on page 63).

[LS18] Eike Langbehn and Frank Steinicke. “Redirected Walking in Virtual Reality”. In: *Encyclopedia of Computer Graphics and Games*. Edited by Niels C. Nilsson. Springer International Publishing, 2018 (cited on page 62).

[LL13] Douglas Lanman and David Luebke. “Near-eye light field displays”. In: *ACM Transactions on Graphics (TOG)* 32.6 (2013), page 220 (cited on page 87).

[LBV99] Markus Lappe, Frank Bremmer, and A. V. Van den Berg. “Perception of self-motion from visual flow”. In: *Trends in Cognitive Sciences* 3.9 (1999), pages 329–336 (cited on page 34).

[LaV00] Joseph J. LaViola Jr. “A Discussion of Cybersickness in Virtual Environments”. In: *SIGCHI Bull.* 32.1 (Jan. 2000), pages 47–56. ISSN: 0736-6906. DOI: 10.1145/333329.333344. URL: http://doi.acm.org/10.1145/333329.333344 (cited on pages 34, 35, 43).

[LaV+01] Joseph J. LaViola Jr., Daniel Acevedo Feliz, Daniel F. Keefe, and Robert C. Zeleznik. "Hands-free Multi-scale Navigation in Virtual Environments". In: *Proceedings of the 2001 Symposium on Interactive 3D Graphics*. I3D '01. New York, NY, USA: ACM, 2001, pages 9–15. ISBN: 1-58113-292-1. DOI: 10.1145/364338.364339. URL: http://doi.acm.org/10.1145/364338.364339 (cited on page 56).

[LaV+17] Joseph J. LaViola Jr., Ernst Kruijff, Ryan P. McMahan, Doug A. Bowman, and Ivan P. Poupyrev. *3D user interfaces: Theory and practice*. Addison-Wesley Professional, 2017 (cited on pages 12–14, 19, 20, 24, 25, 27–29, 31–34, 36–40, 43, 46–48, 51, 52, 56–60, 63, 78, 89, 90, 127).

[LW13] Michael D. Lee and Eric-Jan Wagenmakers. *Bayesian modelling for cognitive science: a practical course*. Cambridge University Press, 2013 (cited on page 123).

[LNY12] Mark R. Lehto, Fiona Fui-Hoon Nah, and Ji Soo Yi. "Decision-making models, decision support, and problem solving". In: *Handbook of Human Factors and Ergonomics*. Edited by Gavriel Salvendy. John Wiley & Sons, 2012, pages 192–242 (cited on page 28).

[LR05] Simon Lessels and Roy A Ruddle. "Movement around real and virtual cluttered environments". In: *Presence: Teleoperators and Virtual Environments* 14.5 (2005), pages 580–596 (cited on pages 30, 61).

[Li+05] Yang Li, Ken Hinckley, Zhiwei Guan, and James A Landay. "Experimental analysis of mode switching techniques in pen-based user interfaces". In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM. 2005, pages 461–470 (cited on pages 51, 81).

[LG94] Jiandong Liang and Mark Green. "JDCAD: A Highly Interactive 3D Modeling System". In: *Computers & Graphics* 18.4 (1994), pages 499–506 (cited on page 52).

[Lin+02] James J.-W. Lin, Henry Been-Lirn Duh, Donald E. Parker, Habib Abi-Rached, and Thomas A. Furness. "Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment". In: *Proceedings of the IEEE Virtual Reality conference (VR)*. 2002, pages 164–171 (cited on pages 14, 45, 161).

[Lin+12] Agneta Lindegård, Jens Wahlström, Mats Hagberg, Rebecka Vilhelmsson, Allan Toomingas, and Ewa Wigaeus Tornqvist. "Perceived exertion, comfort and working technique in professional computer users and associations with the incidence of neck and upper extremity symptoms". In: *BMC musculoskeletal disorders* 13.1 (2012), page 38 (cited on pages 40, 120).

[LCE08] Geniva Liu, Romeo Chua, and James T. Enns. "Attention for perception and action: task interference for action planning, but not for online control". In: *Exp. Brain Res.* 185 (2008), pages 709–717 (cited on page 49).

[LK03] Jack M. Loomis and Joshua M. Knapp. "Visual Perception of Egocentric Distance in Real and Virtual Environments". In: *Virtual and Adaptive Environments*. Edited by L. Hettinger and M. Haas. Erlbaum, 2003, pages 21–46 (cited on pages 33, 53, 78, 161).

[LL86] Jack M. Loomis and Susan J. Lederman. "Tactual Perception". In: edited by Kenneth R. Boff, Lloyd Kaufman, and James P. Thomas. Volume 2. 1986, pages 31–40 (cited on page 34).

[Lub+15] **Paul Lubos**, Oscar Ariza, Gerd Bruder, Florian Daiber, Frank Steinicke, and Antonio Krüger. "HoverSpace". In: *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT)*. Edited by Julio Abascal, Simone Barbosa, Mirko Fetter, Tom Gross, Philippe Palanque, and Marco Winckler. Cham: Springer International Publishing, 2015, pages 259–277. ISBN: 978-3-319-22698-9. DOI: 10.1007/978-3-319-22698-9_17. URL: http://dx.doi.org/10.1007/978-3-319-22698-9_17 (cited on pages 15, 65).

[Lub+14a] **Paul Lubos**, Rüdiger Beimler, Markus Lammers, and Frank Steinicke. "Touching the Cloud: Bimanual annotation of immersive point clouds". In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. Mar. 2014, pages 191–192. DOI: 10.1109/3DUI.2014.6798885 (cited on pages 42, 49, 193).

[Lub+16a] **Paul Lubos**, Gerd Bruder, Oscar Ariza, and Frank Steinicke. "Ambiculus: LED-based Low-resolution Peripheral Display Extension for Immersive Head-mounted Displays". In: *Proceedings of the ACM Virtual Reality International Conference (VRIC)*. VRIC '16. Laval, France: ACM, 2016, 13:1–13:4. ISBN: 978-1-4503-4180-6. DOI: 10.1145/2927929.2927939. URL: http://doi.acm.org/10.1145/2927929.2927939 (cited on pages 14, 15, 59, 127).

[Lub+16b] **Paul Lubos**, Gerd Bruder, Oscar Ariza, and Frank Steinicke. "Touching the Sphere: Leveraging Joint-Centered Kinespheres for Spatial User Interaction". In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI '16. Tokyo, Japan: ACM, 2016, pages 13–22. ISBN: 978-1-4503-4068-7. DOI: 10.1145/2983310.2985753. URL: http://doi.acm.org/10.1145/2983310.2985753 (cited on pages 15, 34, 101).

[LBS14a] **Paul Lubos**, Gerd Bruder, and Frank Steinicke. "Analysis of Direct Selection in Head-Mounted Display Environments". In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2014, pages 1–8 (cited on pages 15, 78, 95, 100–102, 114, 118).

[LBS14b] **Paul Lubos**, Gerd Bruder, and Frank Steinicke. "Are 4 Hands Better Than 2?: Bimanual Interaction for Quadmanual User Interfaces". In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI '14. Honolulu, Hawaii, USA: ACM, 2014, pages 123–126. ISBN: 978-1-4503-2820-3. DOI: 10.1145/2659766.2659782. URL: http://doi.acm.org/10.1145/2659766.2659782 (cited on pages 15, 115, 119, 124).

[LBS14c] **Paul Lubos**, Gerd Bruder, and Frank Steinicke. "Safe-&-round: Bringing Redirected Walking to Small Virtual Reality Laboratories". In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI '14. Honolulu, Hawaii, USA: ACM, 2014, pages 154–154. ISBN: 978-1-4503-2820-3. DOI: 10.1145/2659766.2661219. URL: http://doi.acm.org/10.1145/2659766.2661219 (cited on pages 13, 15, 127, 147, 153).

[LBS15] **Paul Lubos**, Gerd Bruder, and Frank Steinicke. “Influence of Comfort on 3D Selection Task Performance in Immersive Desktop Setups”. In: *Journal of Virtual Reality and Broadcasting* 12(2015).2 (2015). ISSN: 1860-2037. URL: http://nbn-resolving.de/urn:nbn:de:0009-6-41261 (cited on pages 15, 40, 89, 90, 108, 114).

[Lub+14b] **Paul Lubos**, Carina Garber, Anjuly Hoffert, Ina Reis, and Frank Steinicke. “The Interactive Spatial Surface - Blended Interaction on a Stereoscopic Multi-Touch Surface”. In: *Mensch & Computer 2014 - Workshopband*. Edited by Andreas Butz, Michael Koch, and Johann Schlichter. Berlin: De Gruyter Oldenbourg, 2014, pages 343–346 (cited on pages 15, 42, 46, 47, 53, 67, 191, 194).

[Mac94] Carl Machover. “Four decades of computer graphics”. In: *IEEE Computer Graphics and Applications* 14.6 (1994), pages 14–19 (cited on pages 22, 24).

[Mac+87] Christine L. MacKenzie, R. G. Marteniuk, C. Dugas, D. Liske, and B. Eickmeier. “Three-dimensional movement trajectories in Fitts’ task: Implications for control”. In: *Q.J. Exp. Psychology-A* 34.4 (1987), pages 629–647 (cited on pages 50, 78, 81, 86, 95).

[Mac92] I. Scott MacKenzie. “Fitts’ Law As a Research and Design Tool in Human-computer Interaction”. In: *Hum.-Comput. Interact.* 7.1 (Mar. 1992), pages 91–139. ISSN: 0737-0024. DOI: 10.1207/s15327051hci0701_3. URL: http://dx.doi.org/10.1207/s15327051hci0701_3 (cited on page 50).

[MI08] I. Scott MacKenzie and Poika Isokoski. “Fitts’ Throughput and the Speed-accuracy Tradeoff”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’08. Florence, Italy: ACM, 2008, pages 1633–1636. ISBN: 978-1-60558-011-1. DOI: 10.1145/1357054.1357308. URL: http://doi.acm.org/10.1145/1357054.1357308 (cited on pages 50, 81, 95).

[Man01] Steve Mann. *Intelligent Image Processing*. John Wiley and Sons, 2001 (cited on page 101).

[MOB03] Alistair P. Mapp, Hiroshi Ono, and Raphael Barbeito. “What Does the Dominant Eye Dominate? A Brief and Somewhat Contentious Review”. In: *Perception & Psychophysics* 65.2 (2003), pages 310–317 (cited on pages 30, 80, 93).

[MPL11] Maud Marchal, Julien Pettré, and Anatole Lécuyer. “Joyman: A human-scale joystick for navigating in virtual worlds”. In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. 2011, pages 19–26 (cited on page 60).

[MN12] Nicolas Marmaras and Dimitris Nathanael. “Workplace Design”. In: *Handbook of Human Factors and Ergonomics*. Edited by Gavriel Salvendy. John Wiley & Sons, 2012, pages 192–242 (cited on page 23).

[Mar12] William S. Marras. “Basic Biomechanics and Workstation Design”. In: *Handbook of Human Factors and Ergonomics*. Edited by Gavriel Salvendy. John Wiley & Sons, 2012, pages 192–242 (cited on page 38).

[Mar+13a] William E. Marsh, Tim Hantel, Christoph Zetzsche, and Kerstin Schill. “Is the user trained? Assessing performance and cognitive resource demands in the Virtusphere”. In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. 2013, pages 15–22 (cited on page 150).

[Mar+13b] William E. Marsh, Jonathan W. Kelly, Veronica J. Dark, and James H. Oliver. “Cognitive demands of semi-natural virtual locomotion”. In: *Presence: Teleoperators and Virtual Environments* 22.3 (2013), pages 216–234 (cited on pages 146, 148).

[MP07] John McLester and Peter St Pierre. *Applied Biomechanics: Concepts and Connections*. Cengage Learning, 2007 (cited on pages 105, 108).

[Mei+15] Tobias Meilinger, Julia Frankenstein, Katsumi Watanabe, Heinrich H. Bülthoff, and Christoph Hölscher. “Reference frames in learning from maps and navigation”. In: *Psychological Research* 79.6 (Nov. 2015), pages 1000–1008. ISSN: 1430-2772. DOI: 10.1007/s00426-014-0629-6. URL: https://doi.org/10.1007/s00426-014-0629-6 (cited on page 37).

[Mie+16] Philipp Miermeister, Maria Lächele, Rainer Boss, Carlo Masone, Christian Schenk, Joachim Tesch, Michael Kerger, Harald Teufel, Andreas Pott, and Heinrich H Bülthoff. “The CableRobot simulator large scale motion platform based on cable robot technology”. In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Oct. 2016, pages 3024–3029. DOI: 10.1109/IROS.2016.7759468 (cited on page 133).

[MK94] Paul Milgram and Fumio Kishino. “A Taxonomy of Mixed Reality Visual Displays”. In: *IEICE Transactions on Information and Systems, Special issue on Networked Reality*. 1994 (cited on pages 20, 21, 169).

[Mil56] George A. Miller. “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information”. In: *Psychological Review* 63.2 (1956), pages 81–97 (cited on page 36).

[Mil14] Michel Millodot. *Dictionary of Optometry and Visual Science E-Book*. Elsevier Health Sciences, 2014 (cited on page 30).

[MPT10] Kouta Minamizawa, Domenico Prattichizzo, and Susumu Tachi. “Simplified design of haptic display by extending one-point kinesthetic feedback to multipoint tactile feedback”. In: *2010 IEEE Haptics Symposium*. Mar. 2010, pages 257–260. DOI: 10.1109/HAPTIC.2010.5444646 (cited on page 195).

[Min95] Mark R. Mine. *Virtual Environments Interaction Technqiues*. Technical report TR95-018. UNC Chapel Hill Computer Science, 1995 (cited on pages 52, 59).

[MBS97] Mark R. Mine, Frederick P. Brooks Jr., and Carlo H. Sequin. “Moving Objects in Space: Exploiting Proprioception in Virtual-Environment interaction”. In: *Proceedings of the ACM SIGGRAPH conference on Computer graphics and interactive techniques (SIGGRAPH)*. ACM Press/Addison-Wesley Publishing Co. 1997, pages 19–26 (cited on pages 13, 34, 51, 52, 59).

[Mon+14] Yasuaki Monnai, Keisuke Hasegawa, Masahiro Fujiwara, Kazuma Yoshino, Seki Inoue, and Hiroyuki Shinoda. "HaptoMime: Mid-air Haptic Interaction with a Floating Virtual Screen". In: *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. UIST '14. Honolulu, Hawaii, USA: ACM, 2014, pages 663–667. ISBN: 978-1-4503-3069-5. DOI: 10.1145/2642918.2647407. URL: http://doi.acm.org/10.1145/2642918.2647407 (cited on page 195).

[Mor+00] Meg Morris, Robert Iansek, Fiona Smithson, and Frances Huxham. "Postural instability in Parkinson's disease: A comparison with and without an added task". In: *Gait Posture* 12 (2000), pages 205–216 (cited on page 148).

[MI01] Atsuo Murata and Hirokazu Iwase. "Extending Fitts' Law to a three-dimensional pointing task". In: *Human Movement Science* 20 (2001), pages 791–805 (cited on page 51).

[MZC12] Carlo Dal Mutto, Pietro Zanuttigh, and Guido M. Cortelazzo. *Time-of-Flight Cameras and Microsoft Kinect*. Springer, 2012 (cited on pages 42, 135).

[Nad+10] Neelesh K. Nadkarni, Karl Zabjek, Betty Lee, William E. McIlroy, and Sandra E. Black. "Effect of working memory and spatial attention tasks on gait in healthy young and older adults". In: *Motor Control* 14.2 (2010), pages 195–210 (cited on page 148).

[Nai06] Michael Naimark. "Aspen the Verb: Musings on Heritage and Virtuality". In: *Presence: Teleoperators and Virtual Environments* 15.3 (2006), pages 330–335. DOI: 10.1162/pres.15.3.330. URL: https://doi.org/10.1162/pres.15.3.330 (cited on page 11).

[Net+11] Christian T. Neth, Jan L. Souman, David Engel, Uwe Kloos, Heinrich H. Bülthoff, and Betty J. Mohler. "Velocity-dependent dynamic curvature gain for redirected walking". In: *Proceedings of the IEEE Virtual Reality conference (VR)*. 2011, pages 151–158 (cited on page 150).

[NLL17] Diederick C. Niehorster, Li Li, and Markus Lappe. "The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research". In: *i-Perception* 8.3 (2017). DOI: 10.1177/2041669517708205. URL: http://dx.doi.org/10.1177/2041669517708205 (cited on pages 26, 41).

[Nie94] Jakob Nielsen. *Usability Engineering*. Interactive Technologies. Elsevier Science, 1994. ISBN: 9780080520292 (cited on pages 21–23).

[NHS04] Norbert Nitzsche, Uwe D. Hanebeck, and Günther Schmidt. "Motion Compression for Telepresent Walking in Large Target Environments". In: *Presence*. Volume 13. 1. 2004, pages 44–60 (cited on page 147).

[Nol71] A. Michael Noll. "Scanned-display Computer Graphics". In: *Communications of the ACM* 14.3 (Mar. 1971), pages 143–150. ISSN: 0001-0782. DOI: 10.1145/362566.362567. URL: http://doi.acm.org/10.1145/362566.362567 (cited on page 22).

[Nor98] Donald A. Norman. *The Design of Everyday Things*. The MIT Press, 1998 (cited on page 119).

[Nor10] Donald A. Norman. "Natural User Interfaces Are Not Natural". In: *interactions* 17.3 (May 2010), pages 6–10. ISSN: 1072-5520. DOI: 10.1145/1744161.1744163. URL: http://doi.acm.org/10.1145/1744161.1744163 (cited on pages 12, 20).

[Ols+11] J. Logan Olson, David M. Krum, Evan A. Suma, and Mark Bolas. "A design for a smartphone-based head mounted display". In: *Proceedings of the IEEE Virtual Reality conference (VR)*. IEEE. 2011, pages 233–234 (cited on pages 12, 45).

[OF03] Alex Olwal and Steve Feiner. "The Flexible Pointer: An Interaction Technique for Selection in Augmented and Virtual Reality". In: *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. 2003, pages 81–82 (cited on page 52).

[Ort+16] Francisco R. Ortega, Fatemeh Abyarjoo, Armando Barreto, Naphtali Rishe, and Malek Adjouadi. *Interaction Design for 3D User Interfaces: The World of Modern Input Devices for Research, Applications, and Game Development*. CRC Press, 2016 (cited on page 12).

[Pal99] Stephen E. Palmer. *Vision Science: Photons to Phenomenology*. A Bradford Book, 1999 (cited on pages 14, 30, 160, 161).

[PWP06] Robert Patterson, Marc D. Winterbottom, and Byron J. Pierce. "Perceptual Issues in the Use of Head-Mounted Visual Displays". In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 48.3 (2006), pages 555–573 (cited on page 160).

[PFW11] Tabitha C. Peck, Henry Fuchs, and Mary C. Whitton. "An Evaluation of Navigational Ability Comparing Redirected Free Exploration with Distractors to Walking-in-Place and Joystick Locomotion Interfaces". In: *Proceedings of the IEEE Virtual Reality conference (VR)*. IEEE, 2011, pages 56–62 (cited on pages 61, 63, 154).

[PR50] Wilder Penfield and Theodore Rasmussen. "The cerebral cortex of man; a clinical study of localization of function." In: (1950) (cited on page 39).

[PS15] Max Pfeiffer and Wolfgang Stuerzlinger. "3D virtual hand pointing with EMS and vibration feedback". In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. Mar. 2015, pages 117–120. DOI: 10.1109/3DUI.2015.7131735 (cited on page 51).

[Pie+97] Jeffrey S. Pierce, Andrew S. Forsberg, Matthew J. Conway, Seung Hong, Robert C. Zeleznik, and Mark R. Mine. "Image Plane Interaction Techniques in 3D Immersive Environments". In: *ACM SIGGRAPH Symposium on Interactive 3D Graphics*. 1997, pages 39–44 (cited on page 58).

[PNK76] Michael I. Posner, Mary J. Nissen, and Raymond M. Klein. "Visual dominance: An information-processing account of its origins and significance." In: *Psychological review* 83.2 (1976), page 157 (cited on page 34).

[Pou+96] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. “The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR”. In: *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. 1996, pages 79–80 (cited on pages 13, 51).

[Pou+98] Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. “Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques”. In: *Computers & Graphics* 17.3 (1998), pages 41–52 (cited on page 52).

[Pou+97] Ivan Poupyrev, Suzanne Weghorst, Mark Billinghurst, and Tadao Ichikawa. “A framework and testbed for studying manipulation techniques for immersive VR”. In: *Proceedings of the ACM symposium on Virtual reality software and technology*. ACM. 1997, pages 21–28 (cited on page 50).

[PHH11] Dmitry Pyryeskin, Mark S. Hancock, and Jesse Hoey. “Extending interactions into hoverspace using reflected light”. In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ACM. 2011, pages 262–263 (cited on page 55).

[Ras+09] Clarence E. Rash, Michael B. Russo, Tomasz R. Letowski, and Elmar T. Schmeisser. *Helmet-mounted Displays: Sensation, Perception, and Cognition Issues*. U.S. Army Aeromedical Research; 1st edition, 2009 (cited on pages 30, 160, 161).

[Rau92] Matthias Rauterberg. “An empirical comparison of menu-selection (CUI) and desktop (GUI) computer programs carried out by beginners and experts”. In: *Behaviour & Information Technology* 11.4 (1992), pages 227–236 (cited on page 23).

[Raz05] Sharif Razzaque. “Redirected Walking”. PhD thesis. University of North Carolina, Chapel Hill, 2005 (cited on pages 13, 62, 147, 154, 157).

[RKW01] Sharif Razzaque, Zachariah Kohn, and Mary C. Whitton. “Redirected Walking”. In: *Proceedings of Eurographics*. ACM, 2001, pages 289–294 (cited on pages 34, 61).

[Reg+05] Jean-Philippe Regnaux, B. David, Olivier Daniel, D. Ben Smail, M. Combeaud, and Bernard Bussel. “Evidence for cognitive processes involved in the control of steady state of walking in healthy subjects and after cerebral damage”. In: *Neurorehabilitation and Neural Repair* 19.2 (2005), pages 125–132 (cited on page 148).

[Rey08] August de los Reyes. “Predicting the past”. In: *Web Directions South 2008. Sydney Convention Centre, Australia, September*. Volume 25. 2008 (cited on page 20).

[Rhe+15] Max Rheiner, Thomas Tobler, Fabian Troxler, Seki Inoue, Keisuke Hasegawa, Yasuaki Monnai, Yasutoshi Makino, Hiroyuki Shinoda, Jules Françoise, Norbert Schnell, Riccardo Borghesi, Frédéric Bevilacqua, Tuncay Cakmak, and Holger Hager. “Demo Hour”. In: *interactions* 22.2 (Feb. 2015), pages 6–9. ISSN: 1072-5520. DOI: 10.1145/2730891. URL: http://doi.acm.org/10.1145/2730891 (cited on pages 132, 133).

[RB04] Bernhard E. Riecke and Heinrich H. Bülthoff. “Spatial Updating in Real and Virtual Environments: Contribution and Interaction of Visual and Vestibular Cues”. In: *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*. APGV ’04. Los Angeles, California, USA: ACM, 2004, pages 9–17. ISBN: 1-58113-914-4. DOI: 10.1145/1012551.1012553. URL: http://doi.acm.org/10.1145/1012551.1012553 (cited on page 30).

[Rie+06] Kai Riege, Thorsten Holtkamper, Gerold Wesche, and Bernd Fröhlich. “The Bent Pick Ray: An Extended Pointing Technique for Multi-User Interaction”. In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. Mar. 2006, pages 62–65. DOI: 10.1109/VR.2006.127 (cited on page 52).

[Riv+10] Jean Baptiste de la Rivière, Nicolas Dittlo, Emmanuel Orvain, Cédric Kervégant, Mathieu Courtois, and Toni Da Luz. “iliGHT 3d touch: a multiview multitouch surface for 3d content visualization and viewpoint sharing”. In: *ACM International Conference on Interactive Tabletops and Surfaces*. ACM. 2010, pages 312–312 (cited on page 191).

[Riz+14] Albert Rizzo, Arno Hartholt, Mario Grimani, Andrew Leeds, and Matt Liewer. “Virtual reality exposure therapy for combat-related posttraumatic stress disorder”. In: *Computer* 47.7 (2014), pages 31–37 (cited on page 29).

[RH95] Warren Robinett and Richard Holloway. “The Visual Display Transformation for Virtual Reality”. In: *Presence: Teleoperators and Virtual Environments* 4.1 (1995), pages 1–23. DOI: 10.1162/pres.1995.4.1.1. URL: https://doi.org/10.1162/pres.1995.4.1.1 (cited on page 45).

[RK76] John A. Roese and Aida S. Khalafalla. “Stereoscopic viewing with PLZT ceramics”. In: *Ferroelectrics* 10.1 (1976), pages 47–51. DOI: 10.1080/00150197608241949. URL: http://www.tandfonline.com/doi/abs/10.1080/00150197608241949 (cited on page 26).

[RPZ15] Thijs Roumen, Simon T. Perrault, and Shengdong Zhao. “NotiRing: A Comparative Study of Notification Channels for Wearable Interactive Rings”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’15. Seoul, Republic of Korea: ACM, 2015, pages 2497–2500. ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702350. URL: http://doi.acm.org/10.1145/2702123.2702350 (cited on page 195).

[RL06] Roy A. Ruddle and Simon Lessels. “For efficient navigational search, humans require full physical movement, but not a rich visual scene”. In: *Psychological Science* 17.6 (2006), pages 460–465 (cited on page 61).

[RL09] Roy A. Ruddle and Simon Lessels. “The Benefits of Using a Walking Interface to Navigate Virtual Environments”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 16.1 (2009), 5:1–5:18 (cited on pages 61, 146).

[RVB10] Roy A. Ruddle, Ekaterina Volkova, and Heinrich H. Bülthoff. “Walking improves your cognitive map in environments that are large-scale and large in extent”. In: *ACM Transactions on Computer-Human Interaction* 18.2 (2010), 10:1–10:22 (cited on page 146).

[Sal12] Gavriel Salvendy. *Handbook of human factors and ergonomics*. John Wiley & Sons, 2012 (cited on page 27).

[SM87] Mark S. Sanders and Ernest J. McCormick. *Human factors in engineering and design*. McGRAW-HILL book company, 1987 (cited on pages 102, 108, 122).

[SS17] Johannes Scholl and Michael Schmidt-Gabriel. "Device for carrying out movements by shifting the center of gravity and/or actuating muscles of a human body". US20170326412A1. US Patent App. 15/527,348. Nov. 16, 2017 (cited on page 132).

[Sch+09a] Johannes Schöning, Florian Daiber, Antonio Krüger, and Michael Rohs. "Using Hands and Feet to Navigate and Manipulate Spatial Data". In: *CHI '09 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '09. Boston, MA, USA: ACM, 2009, pages 4663–4668. ISBN: 978-1-60558-247-4. DOI: 10.1145/1520340.1520717. URL: http://doi.acm.org/10.1145/1520340.1520717 (cited on page 39).

[Sch+10] Johannes Schöning, Jonathan Hook, Tom Bartindale, Dominik Schmidt, Patrick Oliver, Florian Echtler, Nima Motamedi, Peter Brandl, and Ulrich von Zadow. "Building Interactive Multi-touch Surfaces". English. In: *Tabletops - Horizontal Interactive Displays*. Edited by Christian Müller-Tomfelde. Human-Computer Interaction Series. Springer London, 2010, pages 27–49. ISBN: 978-1-84996-112-7. DOI: 10.1007/978-1-84996-113-4_2. URL: http://dx.doi.org/10.1007/978-1-84996-113-4_2 (cited on page 47).

[Sch+09b] Johannes Schöning, Frank Steinicke, Antonio Krüger, Klaus H. Hinrichs, and Dimitar Valkov. "Bimanual interaction with interscopic multi-touch surfaces". In: *IFIP Conference on Human-Computer Interaction*. Springer. 2009, pages 40–53 (cited on pages 54, 66).

[Sch+16] Britta Schulte, Pauline Bimberg, Julia Hertel, Franziska Neu, Philipp Heidenreich, Gerd Bruder, and **Paul Lubos**. "Evaluation of Two Leaning-Based Control Mechanisms in a Dragon Riding Interface". In: *Proceedings of the GI Workshop on Virtual and Augmented Reality (GI VR/AR)*. Edited by Thies Pfeiffer, Julia Fröhlich, and Rolf Kruse(Hg.) Aachen: Shaker Verlag, 2016, pages 109–116 (cited on pages 13, 15, 59, 127).

[Sch+12] Udo Schultheis, Jason Jerald, Fernando Toledo, Arun Yoganandan, and Paul Mlyniec. "Comparison of a two-handed interface to a wand interface and a mouse interface for fundamental 3D tasks". In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. Mar. 2012, pages 117–124. DOI: 10.1109/3DUI.2012.6184195 (cited on pages 13, 26, 115).

[SVK07] Pieter Seuntiens, Ingrid Vogels, and Arnold van Keersop. "Visual experience of 3D-TV with pixelated ambilight". In: *Proceedings of PRESENCE* (2007) (cited on pages 45, 161).

[Sha98] Christopher D. Shaw. "Pain and Fatigue in Desktop VR". In: *Graphics Interface*. 1998, pages 185–192 (cited on page 100).

[Shi+14] Roy Shilkrot, Jochen Huber, Connie Liu, Pattie Maes, and Suranga Chandima Nanayakkara. "FingerReader: A Wearable Device to Support Text Reading on the Go". In: *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '14. Toronto, Ontario, Canada: ACM, 2014, pages 2359–2364. ISBN: 978-1-4503-2474-8. DOI: 10.1145/2559206.2581220. URL: http://doi.acm.org/10.1145/2559206.2581220 (cited on page 195).

[Shn81] Ben Shneiderman. "Direct manipulation: A step beyond programming languages". In: *ACM SIGSOC Bulletin*. Volume 13. 2-3. ACM. 1981, page 143 (cited on pages 22, 23).

[Sho92] Ken Shoemake. "ARCBALL: a user interface for specifying three-dimensional orientation using a mouse". In: *Graphics Interface*. Volume 92. 1992, pages 151–156 (cited on page 25).

[Sho+11] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. "Real-time human pose recognition in parts from single depth images". In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. Ieee. 2011, pages 1297–1304 (cited on page 42).

[SW75] Alexander W. Siegel and Sheldon H. White. "The Development of Spatial Representations of Large-Scale Environments". In: edited by Hayne W. Reese. Volume 10. Advances in Child Development and Behavior Supplement C. JAI, 1975, pages 9–55. DOI: https://doi.org/10.1016/S0065-2407(08)60007-5. URL: http://www.sciencedirect.com/science/article/pii/S0065240708600075 (cited on page 36).

[SL97] Daniel J. Simons and Daniel T. Levin. "Change blindness". In: *Trends in Cognitive Sciences* 1.7 (1997), pages 261–267. ISSN: 1364-6613. DOI: https://doi.org/10.1016/S1364-6613(97)01080-2. URL: http://www.sciencedirect.com/science/article/pii/S1364661397010802 (cited on page 28).

[Sla09] Mel Slater. "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments". In: *Philosophical Transactions of the Royal Society Biological Science* 364.1535 (2009), pages 3549–3557 (cited on page 154).

[SUS95] Mel Slater, Martin Usoh, and Anthony Steed. "Taking steps: the influence of a walking technique on presence in virtual reality". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 2.3 (1995), pages 201–219 (cited on page 60).

[SW95] Mel Slater and Sylvia Wilbur. "Through the looking glass world of presence: A framework for immersive virtual environments". In: *FIVE*. Volume 95. 1995, pages 1–20 (cited on page 35).

[SK13] Edward E Smith and Stephen M Kosslyn. *Cognitive Psychology: Pearson New International Edition: Mind and Brain*. Pearson Higher Ed, 2013 (cited on page 36).

[Son+12] Peng Song, Wooi Boon Goh, William Hutama, Chi-Wing Fu, and Xiaopei Liu. “A Handle Bar Metaphor for Virtual Object Manipulation with Mid-air Interaction”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’12. Austin, Texas, USA: ACM, 2012, pages 1297–1306. ISBN: 978-1-4503-1015-4. DOI: 10.1145/2207676.2208585. URL: http://doi.acm.org/10.1145/2207676.2208585 (cited on page 193).

[Sou+11] Jan L. Souman, P. Robuffo Giordano, Martin Schwaiger, Ilja Frissen, Thomas Thümmel, Heinz Ulbrich, A. De Luca, Heinrich H. Bülthoff, and Marc O. Ernst. “CyberWalk: Enabling unconstrained omnidirectional walking through virtual environments”. In: *ACM Transactions on Applied Perception (TAP)* 8.4 (2011), pages 1–22 (cited on page 61).

[SK00] Nancy Staggers and David Kobus. “Comparing response time, errors, and satisfaction between text-based and graphical user interfaces during nursing order tasks”. In: *Journal of the American Medical Informatics Association* 7.2 (2000), pages 164–176 (cited on page 23).

[Sta12] International Organization for Standardization. “9241–411 Ergonomics of human-system interaction–Part 411: Evaluation methods for the design of physical input devices”. In: *International Organization for Standardization* (2012) (cited on page 40).

[SKD97] Kay M. Stanney, Robert S. Kennedy, and Julie M. Drexler. “Cybersickness is Not Simulator Sickness”. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 41.2 (1997), pages 1138–1142. DOI: 10.1177/107118139704100292. URL: https://doi.org/10.1177/107118139704100292 (cited on page 35).

[SF87] Robin R. Steeves and Clive S. Fraser. “Statistical Post-Analysis of Least-Squares Adjustment Results”. In: *CISM Adjustment and Analysis Seminars, EJ Krakiwsky (Ed.)* 1987 (cited on page 87).

[Ste+08a] F. Steinicke, G. Bruder, L. Kohli, J. Jerald, and Klaus H. Hinrichs. “Taxonomy and Implementation of Redirection Techniques for Ubiquitous Passive Haptic Feedback”. In: *Cyberworlds*. IEEE Press, 2008, pages 217–223 (cited on page 62).

[Ste16] Frank Steinicke. *Being Really Virtual - Immersive Natives and the Future of Virtual Reality*. Springer, 2016 (cited on pages 11, 12, 21).

[Ste17] Frank Steinicke. “Fooling Your Senses: (Super-)Natural User Interfaces for the Ultimate Display”. In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI ’17. Brighton, United Kingdom: ACM, 2017, pages 1–2. ISBN: 978-1-4503-5486-8. DOI: 10.1145/3131277.3143321. URL: http://doi.acm.org/10.1145/3131277.3143321 (cited on page 13).

[Ste+12] Frank Steinicke, Hrvoje Benko, Antonio Krüger, Daniel Keefe, Jean-Baptiste de la Riviére, Ken Anderson, Jonna Häkkilä, Leena Arhippainen, and Minna Pakanen. “The 3rd Dimension of CHI (3DCHI): Touching and Designing 3D User Interfaces”. In: *CHI ’12 Extended Abstracts on Human Factors in Computing Systems*. CHI EA ’12. Austin, Texas, USA: ACM, 2012, pages 2695–2698. ISBN: 978-1-4503-1016-1. DOI: 10.1145/2212776.2212698. URL: http://doi.acm.org/10.1145/2212776.2212698 (cited on page 90).

[SBH07] Frank Steinicke, Gerd Bruder, and Klaus H. Hinrichs. “Hybrid Traveling in Fully-immersive Large-scale Geographic Environments”. In: *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology*. VRST ’07. Newport Beach, California: ACM, 2007, pages 229–230. ISBN: 978-1-59593-863-3. DOI: 10.1145/1315184.1315234. URL: http://doi.acm.org/10.1145/1315184.1315234 (cited on page 59).

[Ste+09] Frank Steinicke, Gerd Bruder, Klaus H. Hinrichs, Markus Lappe, Brian Ries, and Victoria Interrante. “Transitional Environments Enhance Distance Perception in Immersive Virtual Reality Systems”. In: *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*. APGV ’09. Chania, Crete, Greece: ACM, 2009, pages 19–26. ISBN: 978-1-60558-743-1. DOI: 10.1145/1620993.1620998. URL: http://doi.acm.org/10.1145/1620993.1620998 (cited on pages 12, 78).

[Ste+10] Frank Steinicke, Gerd Bruder, Jason Jerald, Harald Frenz, and Markus Lappe. “Estimation of Detection Thresholds for Redirected Walking Techniques”. In: *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 16.1 (2010), pages 17–27 (cited on pages 33, 61, 62, 149, 150, 153, 157).

[Ste+08b] Frank Steinicke, Gerd Bruder, Timo Ropinski, and Klaus H. Hinrichs. “Moving Towards Generally Applicable Redirected Walking”. In: *Proceedings of the ACM Virtual Reality International Conference (VRIC)*. IEEE Press, 2008, pages 15–24 (cited on pages 34, 35).

[Ste+07] Frank Steinicke, Christian P. Jansen, Klaus H. Hinrichs, Jan Vahrenhold, and Bernd Schwald. “Generating optimized marker-based rigid bodies for optical tracking systems”. In: *VISAPP*. 2007 (cited on page 41).

[SRH04] Frank Steinicke, Timo Ropinski, and Klaus H. Hinrichs. “Object Selection in Virtual Environments with an Improved Virtual Pointer Metaphor”. In: *Proceedings of International Conference on Computer Vision and Graphics (ICCVG)*. 2004, pages 320–326 (cited on pages 48, 52, 78).

[Ste+13] Frank Steinicke, Yon Visell, Jennifer Campos, and Anatole Lécuyer. *Human Walking in Virtual Environments: Perception, Technology, and Applications*. Springer Verlag, 2013 (cited on pages 60, 154, 159).

[SCP95] Richard Stoakley, Matthew J. Conway, and Randy Pausch. “Virtual Reality on a WIM: Interactive Worlds in Miniature”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’95. Denver, Colorado, USA: ACM Press/Addison-Wesley Publishing Co., 1995, pages 265–272. ISBN: 0-201-84705-1. DOI: 10.1145/223904.223938. URL: http://dx.doi.org/10.1145/223904.223938 (cited on pages 26, 51, 56, 58).

[SRJ11] Hans Strasburger, Ingo Rentschler, and Martin Jüttner. "Peripheral vision and pattern recognition: A review". In: *Journal of Vision* 11.5 (2011), page 13. DOI: 10.1167/11.5.13. URL: http://dx.doi.org/10.1167/11.5.13 (cited on page 30).

[SVH11] Sven Strothoff, Dimitar Valkov, and Klaus H. Hinrichs. "Triangle Cursor: Interactions with Objects Above the Tabletop". In: *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*. ITS '11. Kobe, Japan: ACM, 2011, pages 111–119. ISBN: 978-1-4503-0871-7. DOI: 10.1145/2076354.2076377. URL: http://doi.acm.org/10.1145/2076354.2076377 (cited on page 54).

[Sum+12] Evan A. Suma, Gerd Bruder, Frank Steinicke, David M. Krum, and Mark Bolas. "A Taxonomy for Deploying Redirection Techniques in Immersive Virtual Environments". In: *Proceedings of the IEEE Virtual Reality conference (VR)*. 2012, pages 43–46 (cited on pages 61, 62, 154).

[Sum+11] Evan A Suma, Seth Clark, David Krum, Samantha Finkelstein, Mark Bolas, and Zachary Warte. "Leveraging change blindness for redirection in virtual environments". In: *Proceedings of the IEEE Virtual Reality conference (VR)*. Mar. 2011, pages 159–166. DOI: 10.1109/VR.2011.5759455 (cited on pages 28, 61).

[Sut63] Ivan Edward Sutherland. "Sketchpad: A Man-Machine Graphical Communications System". PhD thesis. M.I.T., 1963 (cited on page 21).

[Sut65] Ivan Edward Sutherland. "The Ultimate Display". In: *Proceedings of IFIP Congress 2*. 1965, pages 506–509 (cited on page 11).

[Sut68] Ivan Edward Sutherland. "A Head-Mounted Three Dimensional Display". In: *Proceedings of the AFIPS Fall Joint Computer Conference*. Volume 33. 1968, pages 757–764 (cited on pages 11, 26, 43).

[Tay+01] Russell M. Taylor II, Thomas C. Hudson, Adam Seeger, Hans Weber, Jeffrey Juliano, and Aron T. Helser. "VRPN: A Device-independent, Network-transparent VR Peripheral System". In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. VRST '01. Baniff, Alberta, Canada: ACM, 2001, pages 55–61. ISBN: 1-58113-427-4. DOI: 10.1145/505008.505019. URL: http://doi.acm.org/10.1145/505008.505019 (cited on page 40).

[TS08] Robert J. Teather and Wolfgang Stuerzlinger. "Assessing the effects of orientation and device on (constrained) 3D movement techniques". In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE. 2008, pages 43–50 (cited on page 13).

[TS11] Robert J. Teather and Wolfgang Stuerzlinger. "Pointing at 3D targets in a stereo head-tracked virtual environment". In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. 2011, pages 87–94 (cited on pages 51, 86).

[TS14] Robert J. Teather and Wolfgang Stuerzlinger. "Visual Aids in 3D Point Selection Experiments". In: *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*. SUI '14. Honolulu, Hawaii, USA: ACM, 2014, pages 127–136. ISBN: 978-1-4503-2820-3. DOI: 10.1145/2659766.2659770. URL: http://doi.acm.org/10.1145/2659766.2659770 (cited on page 121).

[Tei90] Michael A. Teitel. "The Eyephone: a head-mounted stereo display". In: volume 1256. 1990. DOI: 10.1117/12.19902. URL: http://dx.doi.org/10.1117/12.19902 (cited on page 12).

[Tho83] James A. Thomson. "Is continuous visual monitoring necessary in visually guided locomotion?" In: *J. Exper. Psych. Hum. Percep. Perform.* 9.3 (1983), pages 427–443 (cited on page 53).

[Uso+99a] Martin Usoh, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Frederick P. Brooks Jr. "Walking > Walking-in-Place > Flying, in Virtual Environments". In: *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*. 1999, pages 359–364 (cited on pages 13, 60, 146).

[Uso+99b] Martin Usoh, Ernest Catena, Sima Arman, and Mel Slater. "Using Presence Questionaires in Reality". In: *Presence: Teleoperators & Virtual Environments* 9.5 (1999), pages 497–503 (cited on pages 35, 82, 96, 106, 107, 150, 153, 158, 159, 201).

[VGH14] Dimitar Valkov, Alexander Giesler, and Klaus H. Hinrichs. "Imperceptible depth shifts for touch interaction with stereoscopic objects". In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM. 2014, pages 227–236 (cited on page 66).

[Val+10a] Dimitar Valkov, Frank Steinicke, Gerd Bruder, and Klaus H. Hinrichs. "Traveling in 3D Virtual Environments with Foot Gestures and a Multi-Touch enabled WIM". In: *Proceedings of the ACM Virtual Reality International Conference (VRIC)*. 2010, pages 171–180 (cited on page 60).

[Val+11] Dimitar Valkov, Frank Steinicke, Gerd Bruder, and Klaus H. Hinrichs. "2D touching of 3D stereoscopic objects". In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM, 2011, pages 1353–1362 (cited on pages 53, 78, 191).

[Val+10b] Dimitar Valkov, Frank Steinicke, Gerd Bruder, Klaus H. Hinrichs, Johannes Schöning, Florian Daiber, and Antonio Krüger. "Touching floating objects in projection-based virtual reality environments". In: *Proceedings of Joint Virtual Reality Conference*. 2010, pages 17–24 (cited on pages 78, 191).

[VI04] Joy Van Baren and Wijnand IJsselsteijn. "Measuring presence: A guide to current measurement approaches". In: *Deliverable of the OmniPres project IST-2001-39237* (2004) (cited on page 35).

[Vet+11] Sebastian Vetter, Jennifer Bützler, Nicole Jochems, and Christopher M. Schlick. “Fitts’ Law in Bivariate Pointing on Large Touch Screens: Age-Differentiated Analysis of Motion Angle Effects on Movement Times and Error Rates”. In: *Universal Access in Human-Computer Interaction. Users Diversity*. Edited by Constantine Stephanidis. Volume 6766. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, pages 620–628. ISBN: 978-3-642-21662-6. DOI: 10.1007/978-3-642-21663-3_67. URL: http://dx.doi.org/10.1007/978-3-642-21663-3_67 (cited on page 80).

[Wad87] Nicholas J. Wade. “On the Late Invention of the Stereoscope”. In: *Perception* 16.6 (1987). PMID: 3331425, pages 785–818. DOI: 10.1068/p160785. URL: https://doi.org/10.1068/p160785 (cited on page 26).

[Wal12] Geoff Walker. “A review of technologies for sensing contact location on the surface of a display”. In: *Journal of the Society for Information Display* 20.8 (2012), pages 413–440. ISSN: 1938-3657. DOI: 10.1002/jsid.100. URL: http://dx.doi.org/10.1002/jsid.100 (cited on pages 23, 42, 66).

[Wan95] Brian A. Wandell. “Useful quantities in vision science”. In: *Foundations of Vision, Sinauer Associates Inc, Sunderland, Massachusetts* (1995) (cited on page 30).

[WL11] Jia Wang and Robert W Lindeman. “Silver Surfer: A system to compare isometric and elastic board interfaces for locomotion in VR”. In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. Mar. 2011, pages 121–122. DOI: 10.1109/3DUI.2011.5759235 (cited on pages 60, 132).

[WM99] Yanqing Wang and Christine L. MacKenzie. “Effects of orientation disparity between haptic and graphic displays of objects in virtual environments”. In: *Proceedings of the IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT)*. Volume 99. 1999, pages 391–398 (cited on pages 52, 78).

[WAB93] Colin Ware, Kevin Arthur, and Kellogg S. Booth. “Fish Tank Virtual Reality”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. 1993, pages 37–42 (cited on page 47).

[WB94] Colin Ware and Ravin Balakrishnan. “Reaching for objects in VR displays: lag and frame rate”. In: *ACM Transactions Compututer-Human Interaction* 1.4 (Dec. 1994), pages 331–356. ISSN: 1073-0516. DOI: 10.1145/198425.198426. URL: http://doi.acm.org/10.1145/198425.198426 (cited on page 51).

[WO90] Colin Ware and Steven Osborne. “Exploration and Virtual Camera Control in Virtual Three Dimensional Environments”. In: *Proceedings of the 1990 Symposium on Interactive 3D Graphics*. I3D ’90. Snowbird, Utah, USA: ACM, 1990, pages 175–183. ISBN: 0-89791-351-5. DOI: 10.1145/91385.91442. URL: http://doi.acm.org/10.1145/91385.91442 (cited on page 58).

[WC15] Rustin Webster and Alex Clark. “Turn-Key Solutions: Virtual Reality”. In: *ASME 2015 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers. 2015, V01BT02A052–V01BT02A052 (cited on page 20).

[Whi+05] Mary C. Whitton, Joseph V. Cohn, Jeff Feasel, Paul Zimmons, Sharif Razzaque, Sarah J. Poulton, Brandi McLeod, and Frederick P. Brooks, Jr. “Comparing VE Locomotion Interfaces”. In: *Proceedings of the IEEE Virtual Reality conference (VR)*. IEEE, 2005, pages 123–130 (cited on page 60).

[Wic+15] Christopher D. Wickens, Justin G. Hollands, Simon Banbury, and Raja Parasuraman. *Engineering psychology & human performance*. Psychology Press, 2015 (cited on pages 27, 28).

[Wic+09] Christopher D. Wickens, Becky L. Hooey, Brian F. Gore, Angelia Sebok, and Corey S. Koenicke. “Identifying Black Swans in NextGen: Predicting Human Performance in Off-Nominal Conditions”. In: *Human Factors* 51.5 (2009). PMID: 20196290, pages 638–651. DOI: 10.1177/0018720809349709. URL: https://doi.org/10.1177/0018720809349709 (cited on pages 27, 28).

[WW11] Daniel Wigdor and Dennis Wixon. *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier, 2011 (cited on pages 12, 27).

[Wil+09] Peter Willemsen, Mark B. Colton, Sarah H. Creem-Regehr, and William B. Thompson. “The Effects of Head-Mounted Display Mechanical Properties and Field-of-View on Distance Judgments in Virtual Environments”. In: *ACM Transactions on Applied Perception (TAP)* 2.6 (2009), pages 1–14 (cited on page 53).

[Wil+08] Peter Willemsen, Amy A. Gooch, William B. Thompson, and Sarah H. Creem-Regehr. “Effects of Stereo Viewing Conditions on Distance Perception in Virtual Environments”. In: *Presence: Teleoperators & Virtual Environments* 17.1 (2008), pages 91–101 (cited on pages 30, 53, 67, 80, 93, 149).

[Wil+11] Betsy Williams, Stephen Bailey, Gayathri Narasimham, Muqun Li, and Bobby Bodenheimer. “Evaluation of Walking in Place on a Wii Balance Board to Explore a Virtual Environment”. In: *ACM Transactions Applied Perception* 8.3 (Aug. 2011), 19:1–19:14. ISSN: 1544-3558. DOI: 10.1145/2010325.2010329. URL: http://doi.acm.org/10.1145/2010325.2010329 (cited on page 60).

[Wil+06] Betsy Williams, Gayathri Narasimham, Tim P. McNamara, Thomas H. Carr, John J. Rieser, and Bobby Bodenheimer. “Updating orientation in large virtual environments using scaled translational gain”. In: *Proceedings of Symposium on Applied Perception in Graphics and Visualization (APGV)*. ACM. 2006, pages 21–28 (cited on pages 61, 157).

[Wil+07] Betsy Williams, Gayathri Narasimham, Bjoern Rump, Timothy P. McNamara, Thomas H. Carr, John Rieser, and Bobby Bodenheimer. “Exploring Large Virtual Environments with an HMD when Physical Space is Limited”. In: *Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization*. APGV ’07. Tubingen, Germany: ACM, 2007, pages 41–48. ISBN: 978-1-59593-670-7. DOI: 10.1145/1272582.1272590. URL: http://doi.acm.org/10.1145/1272582.1272590 (cited on pages 61, 63).

[Wil10] Andrew D. Wilson. “Using a Depth Camera As a Touch Sensor”. In: *ACM International Conference on Interactive Tabletops and Surfaces*. ITS ’10. Saarbrücken, Germany: ACM, 2010, pages 69–72. ISBN: 978-1-4503-0399-6. DOI: 10.1145/1936652.1936665. URL: http://doi.acm.org/10.1145/1936652.1936665 (cited on page 55).

[WB05] Chadwick A. Wingrave and Doug A. Bowman. “Baseline factors for raycasting selection”. In: *Proceedings of HCI International*. 2005 (cited on page 50).

[WS98] Bob G. Witmer and Michael J. Singer. “Measuring presence in virtual environments: A presence questionnaire”. In: *Presence: Teleoperators and Virtual Environments* 7 (1998), pages 225–240. (Cited on pages 35, 202, 203).

[WMW09] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. “User-defined gestures for surface computing”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM. 2009, pages 1083–1092 (cited on page 54).

[WW12] Jonathan R. Wolpaw and Elizabeth Winter Wolpaw. “THE FUTURE OF BCIS: MEETING THE EXPECTATIONS”. In: *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press, 2012, page 387 (cited on page 169).

[WS02] Marjorie Woollacott and Anne Shumway-Cook. “Attention and the control of posture and gait: a review of an emerging area of research”. In: *Gait Posture* 16.1 (2002), pages 1–14 (cited on page 148).

[WB03] Mike Wu and Ravin Balakrishnan. “Multi-finger and Whole Hand Gestural Interaction Techniques for Multi-user Tabletop Displays”. In: *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*. UIST ’03. Vancouver, Canada: ACM, 2003, pages 193–202. ISBN: 1-58113-636-6. DOI: 10.1145/964696.964718. URL: http://doi.acm.org/10.1145/964696.964718 (cited on pages 23, 24).

[XB16] Robert Xiao and Hrvoje Benko. “Augmenting the Field-of-View of Head-Mounted Displays with Sparse Peripheral Displays”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’16. San Jose, California, USA: ACM, 2016, pages 1221–1232. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858212. URL: http://doi.acm.org/10.1145/2858036.2858212 (cited on page 166).

[Zan+05] Catherine A. Zanbaka, Benjamin C. Lok, Sabarish V. Babu, Amy Catherine Ulinski, and Larry F. Hodges. “Comparison of path visualizations and cognitive measures relative to travel techniques in a virtual environment”. In: *IEEE Transactions on Visualization and Computer Graphics* 11 (2005), pages 694–705 (cited on page 148).

[ZK16] Markus Zank and Andreas Kunz. “Eye tracking for locomotion prediction in redirected walking”. In: *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)*. Mar. 2016, pages 49–58. DOI: 10.1109/3DUI.2016.7460030 (cited on page 62).

[Zel+02] Robert C. Zeleznik, Joseph J. LaViola, Daniel Acevedo Feliz, and Daniel F. Keefe. “Pop through button devices for VE navigation and interaction”. In: *Proceedings of the IEEE Virtual Reality conference (VR)*. 2002, pages 127–134. DOI: 10.1109/VR.2002.996515 (cited on page 58).

[Zha96] Shumin Zhai. *Human performance in six degree of freedom input control*. University of Toronto, 1996 (cited on page 38).

[ZBM94] Shumin Zhai, William A. S. Buxton, and Paul Milgram. “The Silk Cursor: Investigating Transparency for 3D Target Acquisition”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’94. Boston, Massachusetts, USA: ACM, 1994, pages 459–464. ISBN: 0-89791-650-6. DOI: 10.1145/191666.191822. URL: http://doi.acm.org/10.1145/191666.191822 (cited on page 48).

[ZMB96] Shumin Zhai, Paul Milgram, and William A. S. Buxton. “The Influence of Muscle Groups on Performance of Multiple Degree-of-freedom Input”. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*. CHI ’96. Vancouver, British Columbia, Canada: ACM, 1996, pages 308–315. ISBN: 0-89791-777-4. DOI: 10.1145/238386.238534. URL: http://doi.acm.org/10.1145/238386.238534 (cited on page 39).

[Zil+13] David Zilch, Gerd Bruder, Frank Steinicke, and Frank Lamak. “Design and Evaluation of 3D GUI Widgets for Stereoscopic Touch-Displays”. In: *Proceedings of the GI-Workshop VR/AR*. 2013, pages 37–48 (cited on page 66).

[Zim+87] Thomas G. Zimmerman, Jaron Lanier, Chuck Blanchard, Steve Bryson, and Young Harvill. “A Hand Gesture Interface Device”. In: *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*. CHI ’87. Toronto, Ontario, Canada: ACM, 1987, pages 189–192. ISBN: 0-89791-213-6. DOI: 10.1145/29933.275628. URL: http://doi.acm.org/10.1145/29933.275628 (cited on page 12).

[Zon14] Ray Zone. *Stereoscopic Cinema and the Origins of 3-D Film, 1838-1952*. University Press of Kentucky, 2014 (cited on page 26).

Electronic

[3DC14] 3DConnexion. *3D Connexion Products*. http://www.3dconnexion.de/products/3a.php. 2014. (Visited on 10/10/2005) (cited on pages 25, 38).

[3Ge14] 3Gear Systems. *3Gear*. http://threegear.com/. http://threegear.com/. 2014 (cited on pages 42, 193).

[Ada17] Adafruit. *Adafruit NeoPixel Library*. 2017. URL: https://github.com/adafruit/Adafruit_NeoPixel (cited on page 162).

[Bou99] Paul Bourke. *Calculating Stereo Pairs*. http://paulbourke.net/stereographics/stereorender/. 1999. (Visited on 1999) (cited on page 46).

[Cmg13] Cmglee. *Human Photoreceptor Distribution*. Nov. 29, 2013. URL: https://commons.wikimedia.org/wiki/File:Human_photoreceptor_distribution.svg (cited on page 29).

[Com17] Computer Dictionary Online. *Gorilla Arm*. Dec. 30, 2017. URL: http://www.computer-dictionary-online.org/?q=gorilla%20arm (cited on page 14).

[Dig16] Digi Capital. *Augmented/Virtual Reality Report 2016*. 2016. URL: http://www.digi-capital.com/news/2016/01/augmentedvirtual-reality-revenue-forecast-revised-to-hit-120-billion-by-2020/#.WJCN_IDQcUE (cited on page 11).

[Dre10] Maayan Dreamer. *Flying Shark*. 2010. URL: http://www.feng-gui.com/research/IronManSim/index.htm (cited on pages 132, 133).

[Far30] Philo Taylor Farnsworth. *Television system*. US Patent 1,773,980. Aug. 1930 (cited on page 22).

[Ham28] Laurens Hammond. *Stereoscopic-picture-viewing apparatus*. US Patent 1,658,439. Feb. 1928. URL: https://www.google.de/patents/US1658439 (cited on page 26).

[HTC17] HTC. *HTC Vive Product*. 2017. URL: https://www.vive.com/us/product/ (cited on pages 26, 41).

[Int00] International Organization for Standardization. *ISO/DIS 9241-9 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9: Requirements for non-keyboard input devices*. 2000 (cited on pages 49, 50, 78, 80, 94, 105).

[Kni13] Rob Knies. *Collaboration, expertise produce enhanced sensing in Xbox One*. Oct. 2, 2013. URL: https://blogs.technet.microsoft.com/microsoft_blog/2013/10/02/collaboration-expertise-produce-enhanced-sensing-in-xbox-one/ (cited on pages 42, 135).

[Lea12] Leap Motion Inc. *Leap Motion*. http://www.leapmotion.com/. https://www.leapmotion.com. 2012 (cited on pages 12, 13, 42, 66).

[Mic12] Microsoft. *Kinect*. https://developer.microsoft.com/en-us/windows/kinect. 2012 (cited on pages 42, 66).

[Mic17] Microsoft. *Inside-out tracking*. Dec. 10, 2017. URL: https://docs.microsoft.com/en-us/windows/mixed-reality/enthusiast-guide/tracking-system (cited on page 41).

[Nas14] Alessandro Nassiri. *Brewster Stereoscope*. July 1, 2014. URL: https://commons.wikimedia.org/wiki/File:IGB_006055_Visore_stereoscopico_portatile_Museo_scienza_e_tecnologia_Milano.jpg (cited on page 44).

[Nat17] NaturalPoint Inc. *OptiTrack*. http://www.optitrack.com/. http://www.optitrack.com. 2017 (cited on page 41).

[Nin18] Nintendo. *Wii Fit Balance Board*. Feb. 16, 2018. URL: http://wiifit.com/ (cited on page 42).

[Ocu13] OculusVR. *Oculus Rift*. http://www.oculusvr.com. 2013 (cited on pages 14, 41).

[Son17] Sony Cooperation. *Playstation VR Full specifications*. 2017. URL: https://www.playstation.com/en-us/explore/playstation-vr/tech-specs/ (cited on page 26).

[Sto15] Seth Stoll. *Icaros VR prototype*. Mar. 25, 2015. URL: http://lumacode.com/articles/icaros-vr-prototype.html (cited on page 133).

[TF12] John G. Tang and Anthony M. Fadell. *Peripheral treatment for head-mounted displays, U.S. Patent US 20080088936 A1, Apple Inc.* 2012 (cited on page 161).

[Tha17] Thales Visionix, Inc. *InterSense*. http://www.intersense.com/. https://www.intersense.com. 2017 (cited on page 41).

[Uni13] Unity Technologies. *Unity3D Game Development Software*. 2013. URL: %5Curl%7Bhttp://unity3d.com/%7D (visited on 2013) (cited on pages 59, 67).

A. Appendices

A.1 Hardware

Here we will describe the hardware developed throughout this thesis. This Section is based on [ALS15; Lub+14b].

A.1.1 iSPACE

In this section, we describe the **i**nteractive **SPA**tial Surfa**CE** (iSPACE), a system combining classical multi-touch-based interaction with direct mid-air selection and manipulation of stereoscopically projected virtual objects. By utilizing a large state-of-the-art Ultra HD 3D display and a high-performance touch frame, the system offers high-quality collaborative exploration. Recently, particularly due to the revival of 3D movies in cinemas, the market of stereoscopic 3D displays experiences massive growth, giving everyone access to the formally expensive technology. Additionally, new display technologies allow an increase of display resolution, enabling retinal resolution even on large displays (cf. Section 2.5). These large 3D screens can be enhanced by adding multi-touch capabilities through so-called touch frames. These technologies dominated recent exhibitions and the entertainment market and have the potential to provide more intuitive and natural interaction setups for a wide range of collaborative scenarios, including geospatial applications, urban planning, architectural design, or collaborative tabletops. For (multi-)touch interaction with monoscopically displayed data, the ability to directly touch elements without additional input devices has been shown to be very appealing for novice and expert users [BWB06]. By 2013, first commercial hardware systems with multi-touch technology and 3D stereoscopic display have been launched [Fis+12; Hac+11; Riv+10] and interdisciplinary research projects explore interaction with stereoscopic content on 2D touch surfaces [Val+11; Val+10b] (cf. Section 2.6.1.5). Moreover, an increasing number of hardware solutions provide the means to sense hand and finger poses and gestures in 3D space without input devices or instrumentation. We developed the iSPACE (cf. Figure A.1), which combines these different technology approaches, and provides enormous potential for a variety of new interaction concepts, in particular for collaborative exploration.

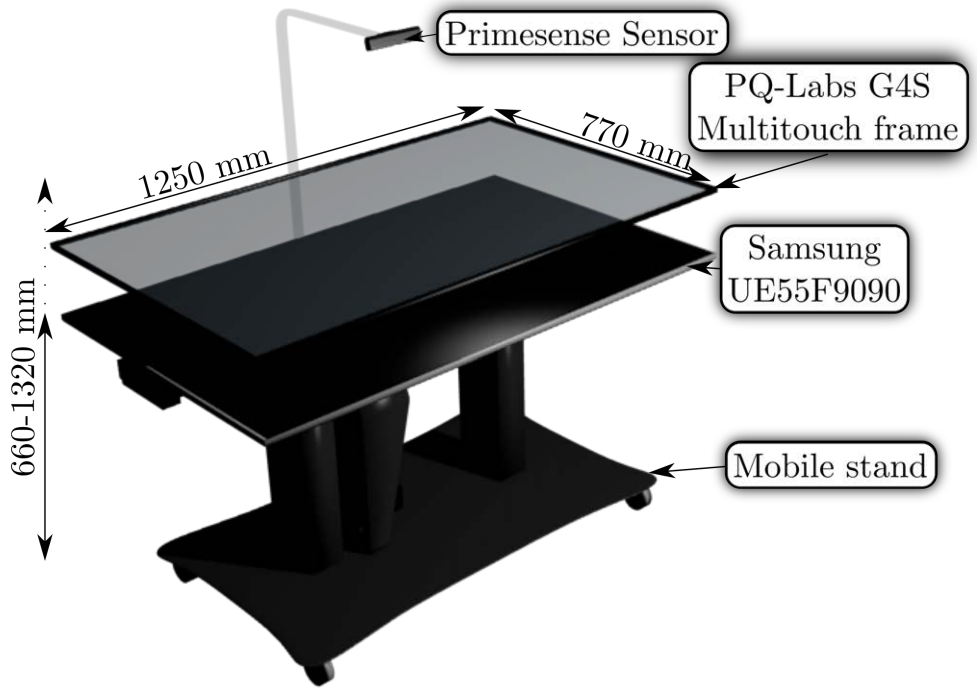


Figure A.1: Schematic illustration of the interactive Spatial Surface in the tabletop mode.

A.1.1.1 Interactive Spatial Surface Setup

As illustrated in Figure A.1, iSPACE consists of three main components: a 3D-capable Ultra high definition (UHD) television, a portable, electrically adjustable for height, stand and the input system consisting of a cell-lightning touch frame for multi-touch-detection, a depth camera for hand tracking and an IR camera for head tracking. To display the 3D images in high-fidelity detail, we used a Samsung UE55F9000 55 inches TV. It uses low weight active shuttering glasses to stereoscopically display images. The maximum resolution is 3840×2160, which is often called UHD or 4k. Actual 4k resolution would refer to a horizontal pixel resolution of 4096 pixels. Also, unlike other setups which require expensive Quadro graphics cards, any modern graphics card with HDMI capabilities can generate the 3D scenes, either Side-by-Side or Top-Bottom and the TV automatically displays the images in a frame sequential mode. Using a TV instead of a projector eliminated the need for calibration of mirrors or lenses and offered the advantage of being able to display dark scenes, as well as bright scenes nearly correctly. The high resolution and the large size of the screen allow multiple users to collaborate in the same space. For the multi-touch-recognition, we decided to use a PQ-Labs G4S frame, which offers smooth tracking of 32 or more touch points by utilizing the patented Cell Imaging Technology, allowing frame rates of 200 fps and detection of points with as low a size as 1.5 mm. The extremely high frame rate allows tracing of quick, natural movements and the technology is less sensitive to environment lighting conditions, allowing the combination with other IR devices, such as the Primesense Carmine. Hence, several users can simultaneously interact in this setup. To be able to offer proper perspective for monoscopic and stereoscopic scenes, we decided to use the Naturalpoint TrackIR dedicated 6 DOF head tracking camera with a FOV of 51.7°, a resolution of 1850 sub-pixels/degree, a sample rate of 120 fps and a response time of 9 ms. Indeed, only a single user can perceive correct perspective in the stereoscopic case, but this perspective can be passed on to other users. Improved shutter glasses similar to the technology presented by Fröhlich et al. could offer frame sequential stereo for more users [Frö+05]. Since studies have shown that direct touch is only viable for heights up to 10 cm above the screen [BSS13c] and above that height mid-air interaction is feasible, we decided to

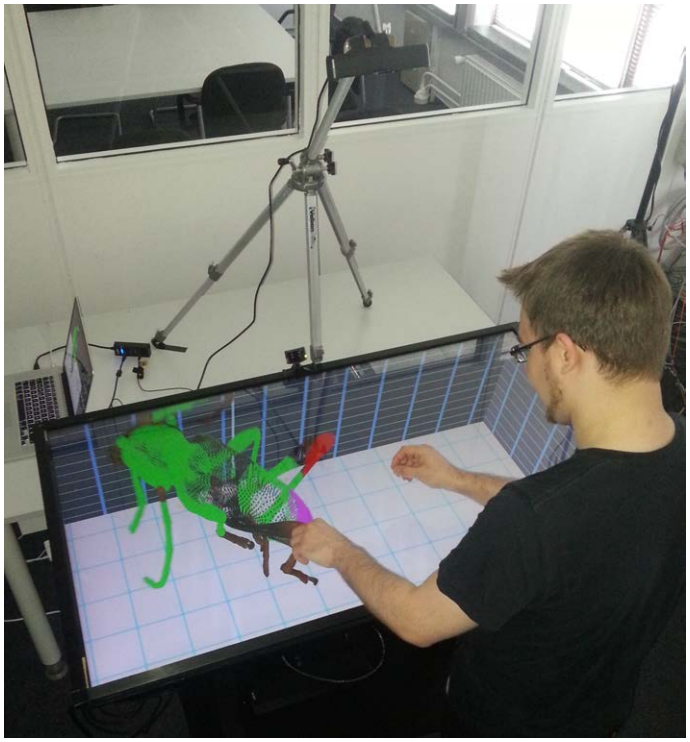


Figure A.2: Illustration of work at the interactive Spatial Surface for annotating a point-cloud data set with the headtracking marker at camera position.

expand the touch interaction through a Primesense Carmine sensor and the 3Gear NimbleSDK. The NimbleSDK allows tracking and virtual reconstruction of a user's hands [3Ge14]. This allows us to create new interaction concepts or adapt existing ones for interactive touch tables. For example, the Handle Bar Metaphor, which users know from smartphones in a two-dimensional context, can be now used in all three dimensions [Son+12]. Furthermore, the hand reconstruction may be used for physics simulations to enhance the user's sense of presence. While the sensor has to be above the user's hands, it works without a surface, allowing the usage in the vertical setup, as well. For instance, the pointing of a user at objects within the VE can be tracked and utilized. As illustrated in Figure A.1, the stand features wheels, enabling new interaction techniques by moving the table around in the real world. Also, the stand is tiltable as well as adjustable in height in an electromechanical way, adding further degrees of freedom. For instance, the surface can be either used as a table or as a vertical screen depending on the application domain. In both scenarios, multiple users can stand around a table surface or vertical projection screen and collaborate. The setup even allows a smooth transition between both paradigms and therefore provides novel ways of collaboration, such as the exploration of architecture or city planning. Where conventional tools usually offer a simple top-down view, iSPACE allows the smooth transition of top-down views to a side-view, providing more precise distance views.

A.1.1.2 Discussion

In this section, we introduced the iSPACE, a setup offering high-fidelity, stereoscopic 3D visualization with multi-touch for collaborative workflows. Due to blending touch interaction with mid-air interaction the setup offers a potential for new interaction metaphors that enhance the stereoscopically presented content. As illustrated in Figure A.2, the setup provides an interesting space for collaborative interactions, for example, in the context of collaborative point-cloud exploration and annotations, based on [Lub+14a]. High resolution combined with stereoscopic

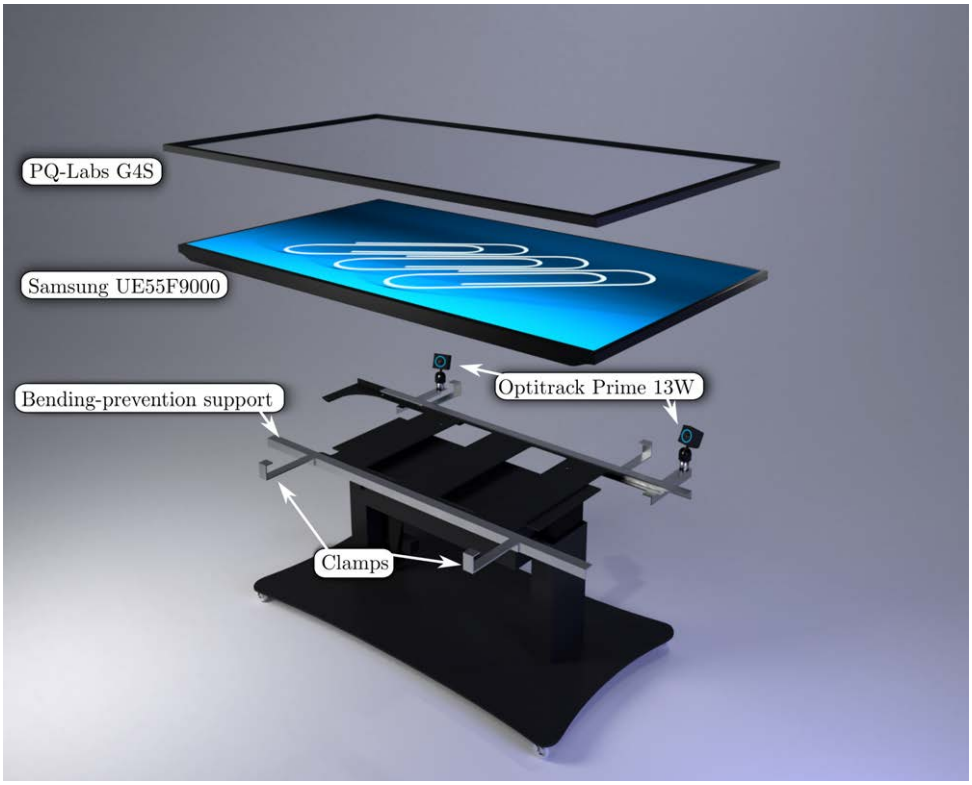


Figure A.3: Improved version of the iSPACE.

display and head-tracking supports better depth perception. Touch interaction allows simple 3D manipulations tasks such as scaling, translation or rotations. Mid-air interaction is required to select specific volumes in the point cloud. Annotations can be performed via speech input. All concepts could be used by multiple users simultaneously. The combination of mid-air interaction with touch input is a novel concept, offering a potential for further research. To summarize, we believe that the iSPACE provides enormous potential for a variety of new interaction concepts, in particular for collaborative exploration of 3D datasets.

A.1.1.3 Improvements

Since the publication of this section [Lub+14b], various improvements were made to the setup, as illustrated in Figure A.3. One of the challenges we noticed was that the screen with the additional weight of the touch frame was prone to bending. Unfortunately, the touch frame is susceptible to bending and would malfunction if the setup was used for lengthy timespans in a horizontal position. We added additional support structure beneath the screen to keep it perfectly planar, as well as metal clamps to prevent accidental shifts of the touch panel to the display instead of the suggested double-sided tape. Furthermore, we replaced the TrackIR head-tracking system with a two camera Optitrack Prime13W (cf. Section 2.5) system, which provides more reliable tracking and can also be used to track other rigid bodies or single markers.

A.1.2 Hapring

Haptic devices can offer reasonable solutions regarding usability and accuracy related to touch feedback on IVEs. However, there are very few affordable devices to perform natural interaction in a 3D space, and some do not represent a suitable solution for the common ergonomic and stimuli-meaningfulness issues. In this section, we present a wireless haptic ring (HapRing) for spatial interaction, providing vibrotactile signals as well as vibration cues on a finger-basis using

a haptic actuator. Other features include inertial measurement, digital input, and support for IR camera-based tracking. This section was originally published in [ALS15].

A.1.2.1 Introduction

Providing touch feedback in spatial interactive applications through haptic technology, using devices composed by actuators, makes more meaningful experiences possible by adding sensations as responses to a user's interaction within a VE. Haptic feedback can be dynamically adjusted according to interaction conditions to provide more significant information such as cues and warnings. Accordingly, the creation of versatile and affordable devices to provide haptic feedback for simple spatial interaction tasks has not been addressed thoroughly, leaving room to propose a device for spatial applications in mid-air while keeping the fingertips free to use multi-touch tabletops, enabling active touch feedback in both cases. Most of the related solutions are focused on multi-modal experiences, centering all the efforts on the senses of sight and hearing and, as a consequence, usually no haptics technology is taken into account. The use of haptic feedback on IVEs would enable the design of new tactile stimuli to adequately convey information in the shape of timely responses, which could be combined with auditory and visual signals, producing better user performance on interactive tasks. Here we present the latest development iteration of HapRing as our proposed device to provide active haptic feedback for spatial interactive applications, eliciting dynamic feedback signals (i.e., vibrotactile patterns) according to the position between the finger and 3D virtual objects.

A.1.2.2 Related Work

Preliminary research on finger-worn devices explored the creation of input devices as well as some tactile feedback devices [Aok+09; MPT10]. Recent work provided real-time auditory and vibrotactile feedback [RPZ15; Shi+14]. Novel techniques in the field, use ultrasonic transducers and air vortex rings to create focused haptic feedback in mid-air [Gup+13; Mon+14]. However, interaction in the 3D mid-air is physically demanding and, therefore, often hinders user satisfaction and performance [Cha+10]. The increase in the DOF that have to be controlled simultaneously as well as the absence of passive haptic feedback and resulting interpenetration and occlusion issues when "touching the void" [BSS13c] are often responsible for reduced performance.

A.1.2.3 Device Prototyping

HapRing comprises the design and creation of a finger-worn device, avoiding cables by using low-energy wireless technology. A vibrational element is managed and controlled by embedded software to provide haptic feedback at different frequencies according to signals received from the computer.

Hardware and Software

The embedded software running on the HapRing is managed and controlled by an RF51822 SoC (System on a Chip) based on a 32-bit ARM Cortex, supporting Bluetooth Smart (2.4 GHz Band) for ultra-low-power wireless applications. All the components are placed on a small development board (18.5 mm $\times$ 21.0 mm) including power connectors (supporting a working voltage from 1.8 V to 3.3 V) and a set of I/O ports used to connect the micro-joystick with command button and the Linear Resonant Actuator (LRA) which is basically a shaft-less vibration motor attached to an active haptic driver (TI DRV2605). As a power source, we used a Lithium Polymer Ion (LyPo) rechargeable battery which outputs a nominal 3.7 V at 40 µA h. A voltage converter and micro USB plug were included. An optional IR LED (with 880 nm wavelength for compatibility with a PPT tracking system) is connected on the ring's top for external reference support. In addition, an IMU is integrated, providing 9 DOF data from a 3-axis combo chip (L3GD20H and LSM303D). For the communication with HapRing, the commands can be sent directly through Bluetooth. We developed an alternative interface application which receives and sends commands through a standard UDP

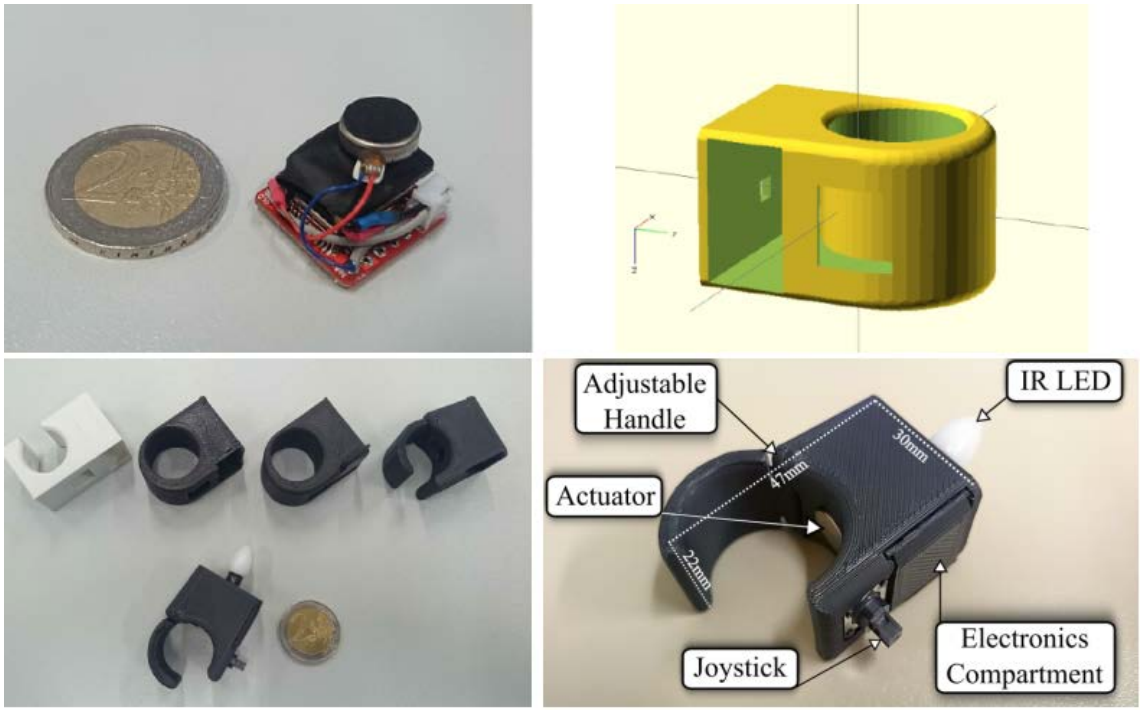


Figure A.4: Illustrations of the HapRing design process including the final prototype.

port and transmits them to HapRing. This setup is modular and would allow different computers to communicate with HapRing, as often necessary in laboratories. The HapRing project is intended to be Open Hardware, enabling the interested audience to build their own device, based on the 3D printing design files, the embedded software source code, the electronics sketches and the C# API.

Latency and Haptic Feedback

We measured the latency from the collision detection in the 3D environment to the vibration signal triggered on the actuator, which was 23 ms, by sending a byte-length message to the device using the UDP-like mechanism to send messages with a "no response flag" provided by the Bluetooth LE Specification. According to perception studies [Bat+04], the measured latency is sufficiently small based on average hand velocity for everyday interactive tasks and vibro-tactile related experiments. To provide haptic feedback in a finger basis, we use a small shaft-less vibration motor (10 mm diameter, 2.0 mm height) with an operating voltage of 2.5 V to 3.5 V. We used a PWM driver to control precisely how the motor runs, taking advantage of overdriving and active braking to guarantee responsive and clean signals. According to the vibrator characteristics, the frequencies range is 20 Hz to 220 Hz. The provided range is sufficient to elicit sensations on the epidermis' tactile corpuscles for haptic feedback related to small displacements and pressure sensations as well as simulation of small texture patterns over movement areas [CC91].

Manufacturing

Aiming to create a comfortable, useful and reliable ring device, we iterated over 3D printers, techniques and materials to get a final 3D printed chassis for our device, mount and test the electronics. The ring was sketched, modeled, sliced and 3D printed according to the constraints imposed by finger ergonomics and electronics. As such, the final design has beveled edges and a joint that allows users to adjust the size. The software tools used were OpenSCAD for 3D modeling, Slic3r for G-Code generation and an Industrial RepRap for the printing labor. As it is symmetrical, the device works just as well for left-handed users as for right-handed users. The finished prototype

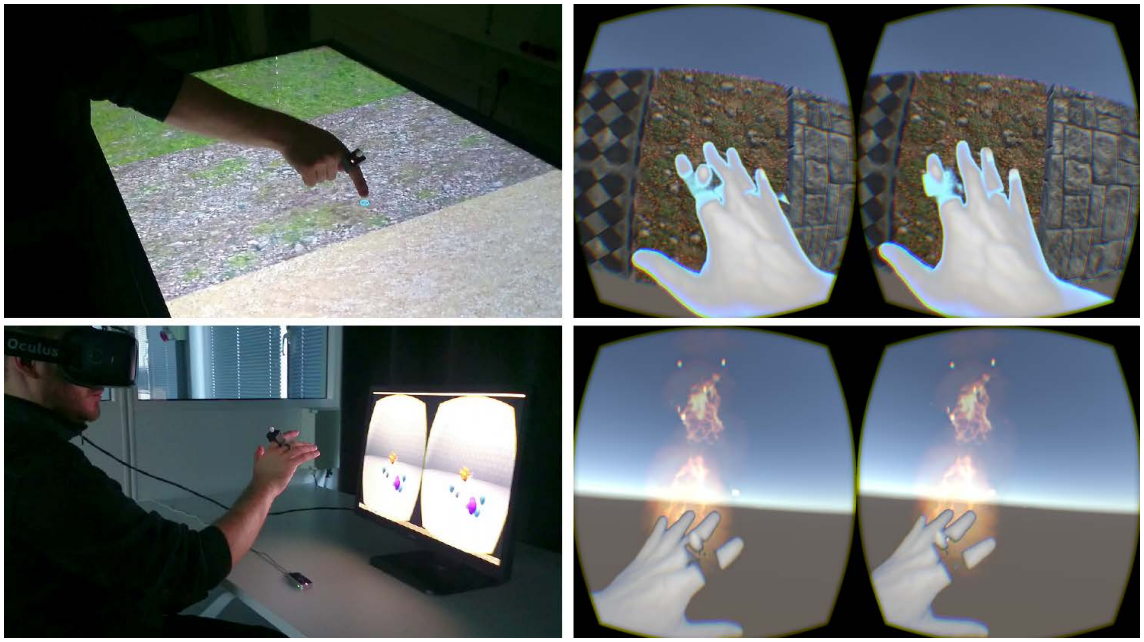


Figure A.5: Illustrations of HapRing use cases.

is illustrated in Figure A.4, which shows the various components and the size (47 mm height × 30 mm width × 22 mm depth).

A.1.2.4 Testing environment

To show potential use-cases for HapRing, we extended several applications by providing haptic feedback for interactions. We developed an Unity3D plugin to simplify the integration of HapRing. Stereoscopic multi-touch tabletops, as well as environments composed by an HMD and a Leap Motion, can take advantage of the haptic feedback . In both scenarios, the user moves their finger while touching a surface and the actuator in the ring is enabled at a certain frequency, depending on the texture touched. Another case integrates HapRing into an existing HMD application to pop balloons which appear in the user's area of interaction by touching them (cf. Figure A.5). Finally, we developed an interactive application to provide warning cues according to a distance-based vibrotactile pattern when the virtual hand of the user approaches a fire source (cf. Figure A.5). More recently, we used HapRing for our experiments described in Section 3.1 and Section 4.2.

A.1.2.5 Conclusion

The presented prototype fulfills the given requirements of a wireless, haptic device, although we still need to evaluate how comfortable it is during long time spans. There are few drawbacks, such as the actuator's vibration being audible in the chassis and minor power leaks requiring more optimizations.

A.2 Questionnaires

A.2.1 SSQ

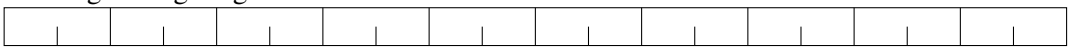
The Kennedy-Lane Simulator Sickness Questionnaire (SSQ), which we use to evaluate cybersickness (cf. Section 2.4.2.4).

1. General discomfort
   ☐ None ☐ slight ☐ Moderate ☐ Severe
2. Fatigue
   ☐ None ☐ slight ☐ Moderate ☐ Severe
3. Headache
   ☐ None ☐ slight ☐ Moderate ☐ Severe
4. Eyestrain
   ☐ None ☐ slight ☐ Moderate ☐ Severe
5. Difficulty focusing
   ☐ None ☐ slight ☐ Moderate ☐ Severe
6. Increased salivation
   ☐ None ☐ slight ☐ Moderate ☐ Severe
7. Sweating
   ☐ None ☐ slight ☐ Moderate ☐ Severe
8. Nausea
   ☐ None ☐ slight ☐ Moderate ☐ Severe
9. Difficulty concentrating
   ☐ None ☐ slight ☐ Moderate ☐ Severe
10. Fullness of head
    ☐ None ☐ slight ☐ Moderate ☐ Severe
11. Blurred vision
    ☐ None ☐ slight ☐ Moderate ☐ Severe
12. Dizzy (eyes open)
    ☐ None ☐ slight ☐ Moderate ☐ Severe
13. Dizzy (eyes closed)
    ☐ None ☐ slight ☐ Moderate ☐ Severe
14. Vertigo
    ☐ None ☐ slight ☐ Moderate ☐ Severe
15. Stomach awareness
    ☐ None ☐ slight ☐ Moderate ☐ Severe
16. Burping
    ☐ None ☐ slight ☐ Moderate ☐ Severe

A.2.2 TLX

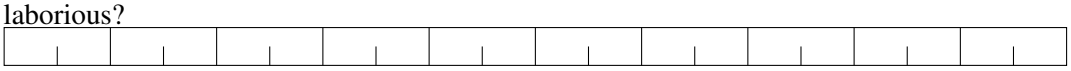
The NASA Task Load Index (TLX) [Har06] (cf. Section 2.4.3). Each question is evaluated on a 100-point scale with 5-point steps. Within this thesis we mostly used the raw scale, which means that we took the mean value from all questions for each participant. In the original version, the overall TLX score is calculated by using 15 two-alternative-forced-choice tasks, which are used for weighting each category, and then divided by 15 to get the overall score.

1. **Mental Demand**
   How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

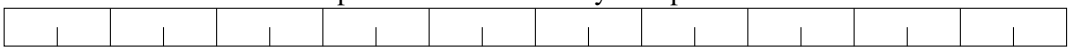
   Low High

2. **Physical Demand**
   How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

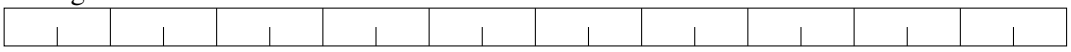
   Low High

3. **Temporal Demand**
   How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

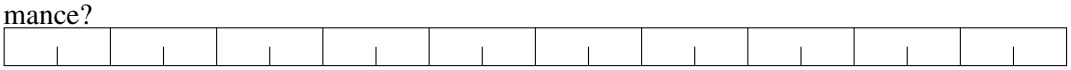
   Low High

4. **Performance**
   How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

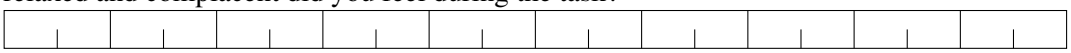
   Good Poor

5. **Effort**
   How hard did you have to work (mentally and physically) to accomplish your level of performance?

   Low High

6. **Frustration**
   How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

   Low High

1. Performance or Mental demand
2. Performance or Frustration
3. Physical demand or Temporal demand
4. Physical demand or Performance
5. Temporal demand or Mental demand
6. Temporal demand or Frustration
7. Temporal demand or Effort
8. Physical demand or Frustration
9. Frustration or Effort
10. Mental demand or Effort
11. Mental demand or Physical demand
12. Effort or Performance
13. Effort or Physical demand
14. Frustration or Mental demand
15. Performance or Temporal demand

A.2.3 Borg

The Borg-15-grade scale for ratings of perceived exertion, adapted from [Bor82]. According to the paper, the rating correlates with the heart rate.

6	
7	Very, very light
8	
9	Very light
10	
11	Fairly light
12	
13	Somewhat hard
14	
15	Hard
16	
17	Very hard
18	
19	Very, very hard
20	

The Borg-10 rating scale with ratio properties, adapted from [Bor82].

0	Nothing at all	
0.5	Very, very weak	(just noticable)
1	Very weak	
2	Weak	(light)
3	Moderate	
4	Somewhat strong	
5	Strong	(heavy)
6		
7	Very strong	
8		
9		
10	Very, very strong	almost max
•	Maximal	

A.2.4 SUS-PQ

The Slater-Usoh-Steed presence questionnaire (SUS-PQ) [Uso+99b] (cf. Section 2.4.2.4).

1. Please rate your sense of being in the virtual environment, on a scale of 1 to 7, where 7 represents your normal experience of being in a place.
   I had a sense of "being there"...
   Not at all ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 Very much
2. To what extent were there times during the experience when the virtual environment was the reality for you?
   There were times when the virtual environment was the reality for me...
   Not at all ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 Almost all the time
3. When you think back to the experience, do you think of the virtual environment more as images that you saw or more as somewhere that you visited?
   The virtual environment seems to me to be more like...
   Images that I saw ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 Somewhere that I visited
4. During the time of the experience, which was the strongest on the whole, your sense of being in the virtual environment or of being elsewhere?
   I had a stronger sense of...
   Being elsewhere ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 Being in the virtual environment
5. During the time of your experience, did you often think to yourself that you were actually in the virtual environment?
   During the experiment I often thought that I was really standing in the virtual environment...
   Not very often ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 Very much so
6. Consider your memory of being in the virtual environment. How similar in terms of the structure of the memory is this to the structure of the memory of other places you have been today? By 'structure of the memory' consider things like the extent to which you have a visual memory of the virtual environment, whether that memory is in colour, the extent to which the memory seems vivid or realistic, its size, location in your imagination, the extent to which it is panoramic in your imagination, and other such structural elements.
   I think of the virtual environment as a place in a way similar to other places that I have been today...
   Not at all ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 Very much so

A.2.5 Immersive Tendencies Questionnaire

The Immersive Tendencies Questionnaire (ITQ), which precedes the WS-PQ [WS98] (cf. Section 2.4.2.4).

1. Do you easily become deeply involved in movies or TV dramas?
   Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
2. Do you ever become so involved in a television program or book that people have problems getting your attention?
   Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
3. How mentally alert do you feel at the present time?
   Not alert ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Fully alert
4. Do you ever become so involved in a movie that you are not aware of things happening around you?
   Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
5. How frequently do you find yourself closely identifying with the character in a story line?
   Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
6. Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?
   Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
7. How physically fit do you feel today?
   Not fit ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Extremely fit
8. How good are you at blocking out external distractions when you are involved in something?
   Not very good ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Very good
9. When watching sports, do you ever become so involved in the game that you react as if you were one of the players?
   Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
10. Do you ever become so involved in a daydream that you are not aware of things happening around you?
    Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
11. Do you ever have dreams that are so real that you feel disoriented when you awake?
    Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
12. When playing sports, do you become so involved in the game that you lose track of time?
    Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
13. How well do you concentrate on enjoyable activities?
    Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Very well
14. How often do you play arcade of video games? (OFTEN should be taken to mean every day or every two day, on average.)
    Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
15. Have you ever gotten excited during a chase of fight scene on TV or in the movies?
    Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
16. Have you ever gotten scared by something happening on a TV show or in a movie?
    Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
17. Have you ever remained apprehensive or fearful long after watching a scary movie?
    Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often
18. Do you ever become so involved in doing something that you lose all track of time?
    Never ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Often

A.2.6 WS-PQ

The Witmer-Singer Presence Questionnaire (WS-PQ) [WS98] (cf. Section 2.4.2.4).

1. How much were you able to control events?
   Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Completely
2. Do you ever become so involved in a television program or book that people have problems getting your attention?
   Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Completely
3. How mentally alert do you feel at the present time?
   Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Completely
4. How much did the visual aspects of the environment involve you?
   Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Completely
5. How natural was the mechanism which controlled movement through the environment?
   Artificial ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Natural
6. How compelling was your sense of objects moving through space?
   Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Very compelling
7. How much did your experiences in the virtual environment seem consistent with your real world experiences?
   Not consistent ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Very consistent
8. Were you able to anticipate what would happen next in response to the actions that you performed?
   Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Completely
9. How completely were you able to actively survey or search the environment using vision?
   Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Completely
10. How compelling was your sense of moving around inside the virtual environment?
    Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Very compelling
11. How closely were you able to examine objects?
    Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Very closely
12. How well could you examine objects from multiple viewpoints?
    Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Extensively
13. How involved were you in the virtual environment experience?
    Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Completely engrossed
14. How much delay did you experience between your actions and expected outcomes?
    No delays ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Long delays
15. How quickly did you adjust to the virtual environment experience??
    Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Less than one minute
16. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?
    Not proficient ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Very proficient
17. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?
    Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Prevented task performance
18. How much did the control devices interfere with the performance of assigned tasks or with other activities?
    Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Interfered greatly
19. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?
    Not at all ☐0 ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 Completely

Eidesstattliche Versicherung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Ort, Datum Unterschrift