Gerd Bruder

Perceptually-Inspired User Interfaces for Computer-Mediated Realities

Perceptually-Inspired User Interfaces for Computer-Mediated Realities

Habilitationsschrift

zur Erlangung des Nachweises besonderer Befähigung zu selbständiger wissenschaftlicher Forschung in der Informatik vorgelegt am Fachbereich Informatik der Fakultät für Mathematik, Informatik und Naturwissenschaften der Universität Hamburg

vorgelegt von

Gerd Bruder

geb. am 1. Dezember 1981 in Datteln

2016

Contents

List of Figures

List of Tables

Acknowledgments

This habilitation thesis presents results of my research at the Immersive Media Group (IMG) at the Julius-Maximilians-Universität Würzburg as well as at the Human-Computer Interaction (HCI) research group at the Universität Hamburg. These acknowledgments try to address all those, who have contributed along the way.

At first, I am deeply grateful to Frank Steinicke for his constant support of my academic journey and a shared enthusiasm for virtual reality (VR).

Furthermore, I would like to thank the members and alumni of my previous research groups in Würzburg and Hamburg with whom I have worked together (in alphabetical order): Antje Lünstedt, Oscar Javier Ariza Nunez, Omar Janeh, Dennis Krupke, Eike Langbehn, Paul Lubos, Andreas Pusch, and Susanne Schmidt.

Special thanks go to the members of the HCI research group at the Julius-Maximilians-Universität Würzburg who have been collaborators in many projects (in alphabetical order): Martin Fischbach, Anke Giebler-Schubert, Marc Erich Latoschik, Jean-Luc Lugrin, Ilka Steinicke, and Dennis Wiebusch.

More thanks go to the co-authors of the selected publications in this habilitation thesis and to the researchers collaborating in the scope of the underlying research projects, which were partly supported by grants from the Deutsche Forschungsgemeinschaft (DFG) or VISIONAIR (in alphabetical order): Ferran Argelaguet, Benjamin Bolte, Victoria Interrante, Markus Lappe, Anatole Lécuyer, Anne-Hélène Olivier, Lane Phillips, Wolfgang Stuerzlinger, and Phil Wieland.

Moreover, I thank all the students of the local departments for their direct or indirect contribution to my research. In particular, I thank the numerous students who have participated in my experiments or completed a research project, master thesis, or bachelor thesis under my supervision.

Last but not least, I am very thankful to my parents Christel and Heiko and to my brother Ralf without whose support this thesis would not have been possible.

Overview

Tremendous advances have been made in the research fields of virtual reality (VR) and augmented reality (AR) within the last years. In particular, following the commercial success of markerless tracking sensors such as the Microsoft Kinect in 2010 as well as the investment of US$2 billion in low-cost tracked head-mounted displays (HMDs) by Facebook in 2014, a continuous increase of public interest in these fields could be observed.

The developments that we are currently seeing in the consumer domain are the accumulation of more than 50 years of research in the fields of VR and AR that have been undertaken since Ivan E. Sutherland presented his essay *The Ultimate Display* in 1965. In this essay, Sutherland outlines his vision of a display system in which a user can perceive a computer-generated virtual environment (VE) with all senses and interact with virtual objects in the same way as with physical objects in the real world. Over the years, attempts to realize his vision have resulted in two main research directions beeing established: Nowadays, VR refers to computer-generated environments that simulate real or imaginary places in which users are immersed into by utilizing, for instance, visual, auditory or haptic display technologies for a multi-sensory experience. Similarly, AR refers to predominantly real-world environments that are augmented with virtual content. These VR and AR systems are usually characterized by light-field or stereoscopic display technologies, such as HMDs or immersive projection environments, and tracking systems that measure the user's head or body pose in real time. By mapping the tracked body movements to a VE it becomes possible for users to leverage natural forms of interaction from the real world for actions in a virtual or augmented world. In particular, users can naturally explore virtual worlds by moving their body, but also use their hands for actions such as touching or grasping of virtual objects. Due to the inherently three-dimensional (3D) nature of display and interaction, these computer-mediated realities have shown great potential for application domains such as architectural design, rehabilitation and training compared to traditional desktop-based environments. Over the last years, the price of VR and AR hardware has dropped significantly and these technologies have become affordable for consumers, the software support has been greatly expanded, the latency of tracking systems has been reduced, and many display characteristics have been improved including the field of view, resolution, weight and comfort.

However, while the technology moves towards maturity in many of the aforementioned fields, there are still multiple limitations of VR and AR environments for which no acceptable hardware solutions exist and which will continue to present major challenges and shape the research field in the foreseeable future. This is mainly due to the fact that the processes of human perception and cognition in computer-mediated realities are not sufficiently under-

stood. While VR technologies can induce a convincing illusion of agency in a virtual world and an impression of being present in a life-like environment, empirical evidence shows that our perceptions differ significantly from what we would perceive in similar situations in the real world. Even if the virtual world is modeled as a photorealistic replica of a real-world environment, users tend to misperceive ego-centric distances to objects, misperceive the size of objects or misjudge their interrelations. These perceptual differences have to be documented and analyzed, causes need to be understood and eventually fixed in order to support applications that require spatial impressions that match those in the real world, such as architectural design or training. However, at the same time, research into human perception and cognition in VEs also provides novel possibilities for the design of user interfaces that arise from the ability to exploit limitations of human perception or the ability to introduce perceptual illusions that change how humans perceive the space-time continuum.

In this thesis, I present eight selected peer-reviewed publications that I have authored within the last five years, which focus on four persistent challenges in this field. According to § 2 of the "Habilitationsordnung" for publications with multiple authors: The order of the authors indicates the contributions to the paper in descending order with the major contributions lying with the author of this habilitation thesis.

Part I focuses on *touch interaction in virtual environments*. While touching or grasping of objects are some of the most basic actions for humans in the real world, which are learned since early childhood, it is not possible to replicate the perception and action of these tasks realistically with current-state VR and AR technologies. The phenomenon that arises when users try to touch an intangible virtual object that is presented stereoscopically within arm's reach is commonly referred to as *touching the void*, which often leads to confusion due to the missing haptic feedback and a significant number of overshoot errors. While a few haptic gloves have been presented that can provide near-natural force feedback, user interfaces that require the user to wear gloves for interaction have rarely been adopted by ordinary users. However, by combining touch-sensitive surfaces with stereoscopic displays it is now possible to provide users with haptic feedback when touching virtual objects on the surface without the requirement for additional instrumentation. The caveat of this approach is that an illusion of matching perceptual and motor spaces can only be induced if the virtual objects are presented stereoscopically in depth close to the display surface. This part explores the potential, limitations, and challenges of user interfaces where the flat digital world of surface computing meets the physical 3D space of free-hand interaction.

Chapter 1 presents analyses of on-display and mid-air touch volumes for 3D target objects that are displayed stereoscopically on or above a touch-sensitive tabletop display [BSS13b]. In particular, the chapter explores a perceptually-inspired interaction technique for touch interaction, in which users unnoticeably touch through an object in mid-air until they get haptic feedback from the physical surface. The chapter presents psychophysical experiments which revealed an effect of eye dominance, distance overestimation and underestimation, and behavior groups in the kinematics of touching floating objects near touch-sensitive surfaces.

Guidelines for designing user interfaces in such novel stereoscopically-enabled tabletops are discussed and an example application is presented.

Chapter 2 extends these analyses by comparing two-dimensional (2D) touch input on a touch-sensitive surface with 3D mid-air touch input in a Fitts' Law experiment, in which selection performance is computed as throughput based on individual metrics of movement times and errors [BSS13a]. The chapter identifies which of these input techniques provides higher performance when interactive objects are presented at different distances from a touch-sensitive surface. The results provide implications for user interface metaphors that seamlessly couple 2D, 2.5D and 3D touch interaction.

Part II focuses on *perceptually-inspired locomotion in virtual environments*. In the real world, walking is considered the most basic and natural forms of locomotion for healthy humans, but leveraging this form of interaction for the exploration of large-scale VEs is difficult to achieve. Although it is possible to simply map a user's physical head movement one-to-one to the movement of a camera for the rendering of a first-person view in the VE, this has the drawback that the user's physical movements are restricted by the range of the tracking sensors and the size of the physical workspace. Since the size of the virtual world is often larger than a room-scale physical workspace, straightforward implementations of omnidirectional unlimited walking are not possible. As a solution to this challenge, *redirected walking* can be applied, which introduces unnoticeable slight rotations of the VE while the user is moving. Using this approach, the user effectively starts walking in small circles in the physical workspace while having the illusion of being able to walk straight for as long as desired in any direction in the virtual world. Previous research showed that it is possible to apply this technique without the user being able to notice the manipulations if a physical workspace of at least 45m×45m is available, which is a significant limitation for practitioners aiming to incorporate natural locomotion into their systems.

Chapter 3 extends upon the previous work by introducing a novel approach to redirected walking by combining it with driving an electric motorized vehicle in the physical workspace [BIPS12]. The chapter describes the first implementation of driving a wheelchair using redirection techniques in a VR laboratory. The concept of walking-and-driving in VR is introduced, in which redirected walking is used to cover short distances while redirected driving is used for longer distances in the VE. This seamless combination of the two forms of locomotion provides near-natural vestibular and proprioceptive feedback of walking or driving as in the real world. In this chapter, psychophysical experiments focusing on rotations, translations and curvatures are presented with a wheelchair and compared to real walking. The results show that redirected driving is less noticeable than redirected walking manipulations and thus well suited for longer distance travel in VR.

Chapter 4 focuses on the question whether redirected walking can and should be applied if the size of the physical workspace does not permit it to be used without the user noticing the manipulations [BLS15]. By using a dual tasking method, an experiment was designed in which the spatial or verbal cognitive load induced by noticeable redirected walking manipulations

was measured. The results showed that redirected walking in physical workspaces of less than 10m×10m draws largely from the finite cognitive resources of the user, which can prove unsuitable in cases in which users have to perform complex cognitive tasks.

Part III focuses on *computer-mediated perceptual illusions*. While walking in the real world, sensory information such as vestibular, proprioceptive, and efferent copy signals as well as visual and auditive information create consistent multi-sensory cues that indicate one's own movement. However, when using VR or AR displays it is possible to subtly manipulate the sensory input in the different modalities, which can effect a change in the perceived motion.

Chapter 5 explores visual illusions with a video see-through AR display which are aimed at changing the perception of self-motion speed while moving in the real world [BWB+13]. The chapter introduces the concept of computer-mediated optic flow to stimulate the motion detectors in the eyes of an observer with different optic flow patterns and velocities. Three techniques were introduced and evaluated in a psychophysical experiment, which showed that it is possible to change human speed perception in the real world with subtle AR manipulations. Based on these illusions it is possible to either use them to correct misperception or to deliberately increase or decrease perceived self-motion speed if desired by an application.

Chapter 6 takes a broader view of perception and illusions in VR [BS14]. From a simple physics perspective, computer-generated virtual worlds follow some of the laws of physics of the real world, but usually do not replicate our space-time continuum entirely. In particular, when walking in VEs these self-motions can be defined by the three components speed, distance and time, which follow the simple equation that speed is the distance covered in a given time interval. However, determining these components of self-motion in the frame of reference of a human observer imposes a significant challenge to the perceptual processes in the human brain, and the resulting speed, distance and time percepts are not always veridical. This chapter presents a psychophysical experiment which analyzes the components individually and shows that motion estimation deviates from the basic mathematical relation of motions between speed, distance and time.

Part IV focuses on *spatial perception with stereoscopic displays*. Empirical studies of spatial perception in VR and AR have revealed significant differences to the real world. When viewing stereoscopically displayed VEs, users tend to overestimate or underestimate egocentric distances to objects, misjudge spatial relations or the size of objects depending on distance and display setup. While many factors have been identified that have an effect on spatial perception in VR and AR, the major causes of these perceptual differences still remain elusive, as are approaches that might be used to reduce the effects. These perceptual differences between geometry shown on VR or AR displays and actual objects in the real world hinder the broad acceptance of such display technologies for many applications domains including architectural design and training.

Chapter 7 analyzes differences between the display hardware and geometric rendering parameters as potential causes of size and distance misperception [BPS12]. The chapter de-

scribes a model of the effects of the field of view and the interpupillary distance in an on-axis viewing condition on the size and distance of virtual objects. The model is tested in an experiment in which the geometric rendering parameters have been varied relative to the hardware parameters of an HMD. The results showed an asymmetric effect that variations in the field of view strongly affected distance estimation, whereas variations in the interpupillary distance mainly affected size estimation.

Chapter 8 presents an analysis of the effect of the screen distance and parallax in off-axis viewing conditions [BAOL16]. The chapter describes an experiment which analyzes ego-centric distance judgments when considering effects of the distance to the projection screen and the stereoscopic parallax of a displayed target object as the main factors. The results showed a significant asymmetric effect of the factors on distance estimation. A second experiment further analyzes the factors contributing to these effects and confirmed the effects with a high-resolution projection setup and improved accommodative stimuli. The results showed that the distance to a projection wall is an important characteristic for veridical distance estimation in immersive projection installations. The chapter provides guidelines for the design of projection setups that minimize misperception.

Bibliography

[BAOL16] Gerd Bruder, Ferran Argelaguet, Anne-Hélène Olivier, and Anatole Lécuyer. CAVE Size Matters: Effects of Screen Distance and Parallax on Distance Estimation in Large Immersive Display Setups. *Presence: Teleoperators & Virtual Environments*, 25(1):1–16, 2016.

[BIPS12] Gerd Bruder, Victoria Interrante, Lane Phillips, and Frank Steinicke. Redirecting Walking and Driving for Natural Navigation in Immersive Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18(4):538–545, 2012.

[BLS15] Gerd Bruder, Paul Lubos, and Frank Steinicke. Cognitive Resource Demands of Redirected Walking. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 21(4):539–544, 2015.

[BPS12] Gerd Bruder, Andreas Pusch, and Frank Steinicke. Analyzing Effects of Geometric Rendering Parameters on Size and Distance Estimation in On-Axis Stereographics. In *Proceedings of the ACM Symposium on Applied Perception (SAP)*, pages 111–118, 2012.

[BS14] Gerd Bruder and Frank Steinicke. Threefolded Motion Perception During Immersive Walkthroughs. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST)*, pages 177–185, 2014.

[BSS13a] Gerd Bruder, Frank Steinicke, and Wolfgang Stuerzlinger. To Touch or not to Touch? Comparing 2D Touch and 3D Mid-Air Interaction on Stereoscopic Table-

top Surfaces. In *Proceedings of ACM Symposium on Spatial User Interaction (SUI)*, pages 9–16, 2013.

[BSS13b] Gerd Bruder, Frank Steinicke, and Wolfgang Stuerzlinger. Touching the Void Revisited: Analyses of Touch Behavior On and Above Tabletop Surfaces. In *Proceedings of the International Conference on Human-Computer Interaction (INTERACT)*, pages 278–296, 2013.

[BWB+13] Gerd Bruder, Phil Wieland, Benjamin Bolte, Markus Lappe, and Frank Steinicke. Going With the Flow: Modifying Self-Motion Perception with Computer-Mediated Optic Flow. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 67–74, 2013.

Part I

Touch Interaction in Virtual Environments

Chapter 1

1 Touching the Void Revisited: Analyses of Touch Behavior On and Above Tabletop Surfaces

Recent developments in touch and display technologies made it possible to integrate touch-sensitive surfaces into stereoscopic three-dimensional (3D) displays. Although this combination provides a compelling user experience, interaction with stereoscopically displayed objects poses some fundamental challenges. If a user aims to select a 3D object, each eye sees a different perspective of the same scene. This results in two distinct projections on the display surface, which raises the question where users would touch in 3D or on the two-dimensional (2D) surface to indicate the selection. In this chapter, we analyze the relation between the 3D positions of stereoscopically displayed objects and the on-surface as well as off-surface touch areas. The results show that 2D touch interaction works better close to the screen but also that 3D interaction is more suitable beyond 10cm from the screen. Finally, we discuss implications for the development of future touch-sensitive interfaces with stereoscopic display.

1.1 Introduction

Recent exhibitions and the entertainment market have been dominated by two different technologies: (i) (multi-)touch-sensitive surfaces and (ii) stereoscopic three-dimensional (3D) displays. Interestingly, these two technologies are orthogonal, as (multi-)touch is about input, whereas 3D stereoscopic display is about output. Both technologies have the potential to provide more intuitive and natural interaction with a wide range of applications, including urban planning, architectural design, collaborative tabletops, or geospatial applications. First commercial hardware systems have recently been launched, e. g., [CKC+10], and interdisciplinary research projects explore interaction with stereoscopic content on 2D touch surfaces, e. g., [iMU13, InS13]. Moreover, an increasing number of hardware solutions provide the means to sense human gestures and postures not only on surfaces, but also in 3D space,

e. g., the Kinect, the Three-Gear system, or Leap Motion. The combination of these novel technologies provides enormous potential for a variety of new interaction concepts.

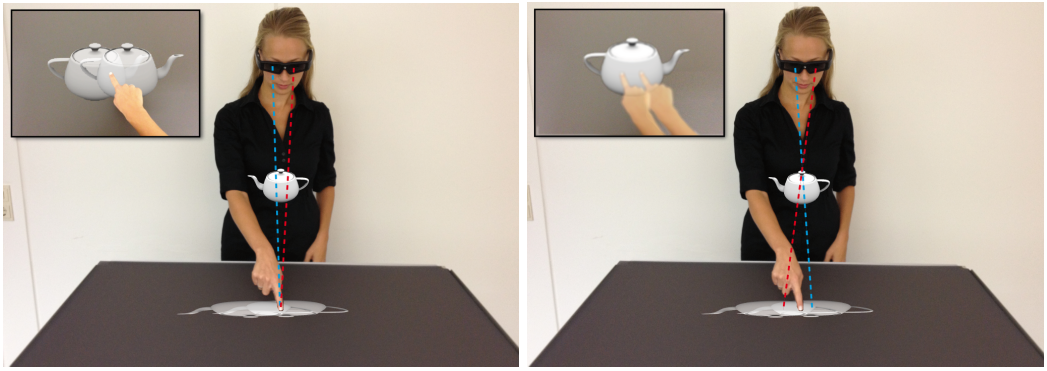
Until recently, research in the area of (multi-)touch interaction was mostly focused on monoscopically displayed data. For this, the ability to directly touch elements without additional input devices has been shown to be very appealing for novice as well as expert users. Also, passive haptics and multi-touch displays have both shown their potential to considerably improve the user experience [BWB06]. Touch surfaces build a consistent and pervasive illusion in perceptual and motor space that the two-dimensional graphical elements on the surface can be touched. Yet, three-dimensional data limits this illusion of place and plausibility [Sla09]. Such 3D data sets are either displayed monoscopically, which has been shown to impair spatial perception and performance in common 3D tasks, or stereoscopically, which can cause objects to appear detached from the touch surface [MHG12, SSV+09, BSS13a].

Stereoscopic display technology has been available for decades. Recently, it was revived due to the rise of 3D cinema, upcoming 3D televisions and 3D games. With stereoscopic displays, each eye sees a different perspective of the same scene through appropriate technology. This requires rendering of two distinct images on the display surface. When using stereoscopic technology to display each projection to only one eye, objects may be displayed with *negative*, *zero*, or *positive parallax*, corresponding to their appearance in front, at, or behind the screen. Objects with zero parallax appear attached to the projection screen and are perfectly suited for touch interaction. In contrast, it is more difficult to apply direct-touch interaction techniques to objects that appear in front of or behind the screen [HCC07, PFC+97, RDH09]. In this chapter, we focus on the major challenge of touching objects that appear in front of the projection screen. Two methodologies can be used for touching such stereoscopic objects on a tabletop display:

1. If the touch-sensitive surface captures only direct contacts, the user has to penetrate the stereoscopically displayed object to touch the 2D surface behind it [VSB+10, VSBH11].

2. Alternatively, if the system can capture finger movements in front of the screen, the user may virtually "touch" the object in *mid-air*, i. e., in 3D space.

Due to the discrepancy between perceptual and motor space and the missing passive haptic feedback, both approaches provide natural feedback only for objects rendered with zero parallax. This poses the questions where users "touch" a stereoscopically displayed object in 3D space. Here, one issue is the well-documented issue of misperception of distances in virtual 3D scenes [LK03]. Another problem arises from potential touch locations on the 2D display surface, as there are two distinct projections, one for each eye. If the user penetrates the object while focusing on her finger, the stereoscopic impression of the object is disturbed, since the user's eyes are not accommodated and converged to the projection screen's surface anymore. Thus, the left and right stereoscopic images of the object's projection appear blurred and can usually not be merged as illustrated in Figure 1.1(a). However, focusing on the virtual object causes a disturbance of the stereoscopic perception of the user's finger, since her eyes are converged on the object's 3D position, see Figure 1.1(b). If a 3D tracking system is used,



(a) (b)

Figure 1.1: Illustration of the main problem of 2D touch interaction with stereoscopically displayed 3D data: The user is either focused (a) on her finger, which makes the selection ambiguous, or (b) on the object, which disturbs the visual perception of the finger.

the user can see a stereoscopic image while converging her eyes to her finger. Yet, due to the vergence-accommodation conflict [BSS13a, CKC+10], the virtual object will appear blurred in comparison to the real finger (see Figure 1.2).

In this chapter, we address the challenge of how to interact with stereoscopic content in front of a touch-sensitive tabletop surface. Towards this, we also analyze touch behavior when touch sensing is constrained to the 2D screen surface. In order to allow the user to select arbitrary objects, a certain area of the touch surface, which we refer to as *on-surface target*, must be assigned to each object. In the monoscopic case, the mapping between an on-surface touch area and the intended object point in the virtual scene is straightforward. Yet, with stereoscopic projection this mapping is more problematical. In particular, since there are different projections for each eye, the question arises *where* users touch the surface when they try to "touch" a stereoscopic object. In principle, the user may touch anywhere on the surface to select a stereoscopically displayed object. However, according to our previous work [VSBH11], the most likely alternatives that users try to touch are the (see Figure 1.4):

- *midpoint* (M) between the projections for both eyes,
- projection for the *dominant eye* (D), or
- projection for the *non-dominant eye* (N).

A precise approach to this mapping is important to ensure efficient interaction and correct selections, in particular in a densely populated virtual scene. First, we determine a precise *on-display target area* where users touch the screen to select a 3D object. Second, we compare this approach with systems where the user's finger can be tracked in 3D space, and where users virtually touch objects in mid-air 3D space. The results of this experiment provide guidelines for the choice of touch technologies, as well as the optimal placement and parallax of interactive elements in stereoscopic touch environments.

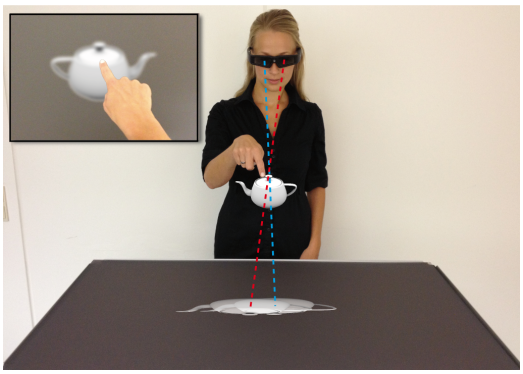


Figure 1.2: Illustration of the main problem of 3D mid-air interaction with stereoscopically displayed 3D data: The user sees a stereoscopic image while converging to her finger, but due to the vergence-accommodation conflict, the virtual object appears blurred in comparison to the finger [BSS13a].

In summary, our contributions are:

- An analysis of on-display touch areas for 3D target objects in stereoscopic touch-sensitive tabletop setups.
- A direct comparison of 2D touch and 3D mid-air selection precision.
- Guidelines for designing user interfaces for stereoscopic touch-sensitive tabletops.

The remainder of this chapter is structured as follows. Section 1.2 summarizes related work in touch interaction and stereoscopic display. Section 1.3 describes the experiments we conducted to identify 2D/3D touch behavior. Section 1.4 presents the results, which are discussed in Section 1.5. An example application using the derived guidelines is described in Section 1.6. Section 1.7 concludes the chapter.

1.2 Background

Recently, many approaches for extending multi-touch interaction techniques to 3D applications with monoscopic display have been proposed [HCC07, MCG10, PHH12, RDH09]. For instance, Hancock et al. [HCC07] presented the concept of shallow-depth 3D, i. e., 3D interaction within a limited range, to extend interaction possibilities with digital 2D surfaces. However, direct touch interaction with stereoscopically displayed scenes introduces new challenges [SSV+09], since the displayed objects can float in front of or behind the interactive display surface. Müller-Tomfelde et al. presented anaglyph- or passive polarization-based stereoscopic visualization combined with FTIR-based touch detection on a multi-touch enabled wall [MTSH+10], and discussed approaches based on mobile devices for addressing the formulated parallax problems. The parallax problem described in the introduction is known from the two-dimensional representation of the mouse cursor within a stereoscopic image [SSV+09]. While the mouse cursor can be displayed stereoscopically on top of objects [SSV+09] or monoscopically only for the dominant eye [TS13], movements of real objects

in the physical space, e. g., the user's hands, cannot be constrained such that they appear only on top of virtual objects. Grossman and Wigdor [GW07] provided an extensive review of the existing work on interactive surfaces and developed a taxonomy for this research. This framework takes the perceived and actual display space, the input space and the physical properties of an interactive surface into account. As shown in their work, 3D volumetric visualizations are rarely considered in combination with 2D direct surface input.

Even on monoscopic touch surfaces, the size of the human fingers and the lack of sensing precision can make precise touch screen interactions difficult [Fer08, VB04]. Some approaches have addressed this issue, for example, by providing an adjustable [BWB06] or fixed cursor offset [PWS88], by scaling the cursor motion [BWB06] or by extracting the orientation of the user's finger [HIW+09].

1.2.1 Kinematics of Touch

The kinematics of point and grasp gestures and the underlying cognitive functions have been studied by many research groups [GCE08, MMD+87, WWG03]. For instance, it has been shown that total arm movement during grasping consists of two distinct component phases:

1. an initial, *ballistic phase* during which the user's attention is focused on the object to be grasped (or touched) and the motion is basically controlled by proprioceptive senses, and

2. a subsequent *correction phase* that reflects refinement and error-correction of the movement, incorporating visual feedback in order to minimize the error between the hand or finger, respectively, and the target [Int00].

Furthermore, MacKenzie et al. [MMD+87] have investigated the real-time kinematics of limb movements in a Fitts' task and have shown that, while Fitts' law holds for the total limb-movement time, humans usually start sooner decelerating the overall motion, if the target seems to require more precision in the end phase. The changes of the kinematics and control of the reaching tasks within virtual environments have also been investigated [BSS13a, dlRDO+10, RDH09]. Valkov et al. [VSB+10] showed that users are, within some range, insensitive to small misalignments between visually perceived stereoscopic positions and the sensed haptic feedback when touching a virtual object. They proposed to manipulate the stereoscopically displayed scene in such a way that the objects are moved towards the screen when the user reaches for them [VGH12, VSB+10]. However, the problem is that objects have to be shifted in space, which may lead to a disturbed perception of the virtual scene for larger manipulations.

1.2.2 3D Touch for 3D Objects

To enable direct "touch" selection of stereoscopically displayed 3D objects in space, 3D tracking technologies can capture a user's hand or finger motions in front of the display surface.

Hilliges et al. [HIW+09] investigated an extension of the interaction space beyond the touch surface. They tested two depth-sensing approaches to enrich multi-touch interaction on a tabletop setup. Although 3D mid-air touch provides an intuitive interaction technique, touching an intangible object, i. e., *touching the void* [CKC+10], leads to potential confusion and a significant number of overshoot errors. This is due to a combination of three factors: depth perception is less accurate in virtual scenes than in the real world, see e. g., [SM90], the introduced double vision, and also vergence-accommodation conflicts. A few devices, such as the CyberGrasp, support haptic feedback when touching objects in space, but require extensive user instrumentation. A similar option for direct touch interaction with stereoscopically rendered 3D objects is to separate the interactive surface from the projection screen, as proposed by Schmalstieg et al. [SES99]. In their approach, the user is provided with a physical *transparent prop*, which can be moved in front of the object of interest. This object can then be manipulated via single- or multi-touch gestures since it has almost zero parallax with respect to the prop.

1.2.3 2D Touch for 3D Objects

Recently, multi-touch devices with non-planar surfaces, such as cubic [dKOD08] or spherical ones [BWB08], were proposed. Other approaches are based on controlling the 3D position of a cursor through multiple touch points [BF07, SVH11]. These can specify 3D axes or points for indirect object manipulation. Interaction with objects with negative parallax on a multi-touch tabletop setup was addressed by Benko et al.'s balloon selection [BF07], as well as Strothoff et al.'s triangle cursor [SVH11], which use 2D touch gestures to specify height above the surface. Valkov et al. [VSBH11] performed a user study, in which they displayed 3D objects stereoscopically either in front of or behind a large vertical projection screen. They recorded user behavior when instructed to touch the virtual 3D objects on the display surface. They identified that users tend to touch between the projections for the two eyes with an offset towards the projection for the dominant eye. However, the results suffered from a large variance between participants. Hence, it is unclear how far these results can be applied to different setups, such as mobile screens or tabletops, where users have an easy frame of reference due to the bezel. Also, they may engage in different touch behavior due to physical support and gravity.

So far, no comparative analysis exists for 2D and 3D touch interaction in stereoscopic tabletop setups. Thus, it remains unclear if 2D touch is a viable alternative to 3D mid-air touch.

1.3 Experiments

Here we describe our experiments in which we analyzed the touch behavior as well as the precision of 2D touch and 3D mid-air touches. We used a standard ISO 9241-9 selection task setup [MMD+87] on a tabletop surface with 3D targets displayed at different heights above

the surface, i. e., with different negative stereoscopic parallaxes.

1.3.1 Participants

Ten male and five female participants (ages 20−35, M=27.1, heights 158−193cm, M=178.3cm) completed the experiment. Participants were students or members of the departments of computer science, media communication or human-computer interaction. Three participants received class credit for participating in the experiment. All participants were right-handed. We used the Porta and Dolman tests to determine the sighting dominant eye of participants [MOB03]. This revealed eight right-eye dominant participants (7 males, 1 female) and five left-eye dominant participants (2 males, 3 females). The tests were inconclusive for two participants (1 male, 1 female), for which the 2 tests indicated conflicting eye dominance. All participants had normal or corrected to normal vision. One participant wore glasses and four participants wore contact lenses during the experiment. None of the participants reported known eye disorders, such as color weaknesses, amblyopia or known stereopsis disruptions. We measured the interpupillary distance (IPD) of each participant before the experiment, which revealed IPDs between 5.8cm and 7.0cm (M=6.4cm). We used each individual's IPD for stereoscopic display in the experiment. 14 participants reported experience with stereoscopic 3D cinema, 14 reported experience with touch screens, and 8 had previously participated in a study involving touch surfaces. Participants were naive to the experimental conditions. Participants were allowed to take a break at any time between experiment trials in order to minimize effects of exhaustion or lack of concentration. The total time per participant including pre-questionnaires, instructions, training, experiment, breaks, post-questionnaires, and debriefing was about 1 hour.

1.3.2 Materials

For the experiment we used a 62cm×112cm multi-touch enabled active stereoscopic tabletop setup as described in [BSS13a]. The system is shown in Figure 1.3 and uses rear diffuse illumination [MHG12] for multi-touch. For this, six high-power infrared (IR) LEDs illuminate the screen from behind. When an object, such as a finger or palm, comes in contact with the diffuse surface it reflects the IR light, which is then sensed by a camera. We use a 1024×768 PointGrey Dragonfly 2 with a wide-angle lens and a matching IR bandpass filter at 30 frames per second. We use a modified version of the NUI Group's CCV software to detect touch input on a Mac Mini server. Our setup uses a matte diffusing screen with a gain of 1.6 for the stereoscopic back projection. We used a 1280×800 Optoma GT720 projector with a wide-angle lens and an active DLP-based shutter at 60Hz per eye. We used an optical WorldViz PPT X4 system with sub-millimeter precision and sub-centimeter accuracy to track the participant's finger and head in 3D, both for 3D "touch" detection as well as view-dependent rendering. For this, we attached wireless markers to the shutter glasses and another diffused IR LED on the tip of the index finger of the participant's dominant hand. We tracked and logged both head and fingertip movements during the experiment.

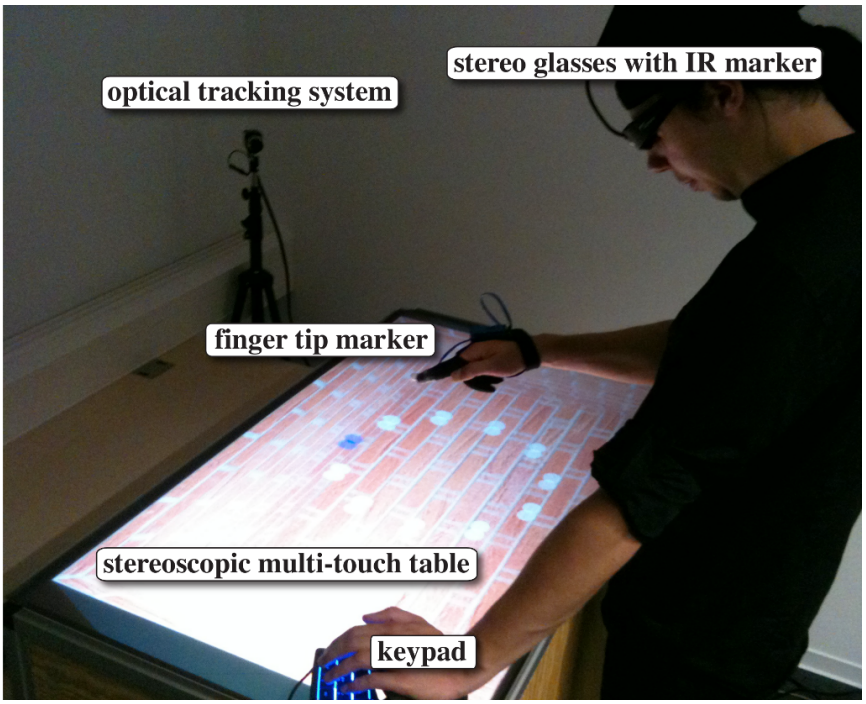


Figure 1.3: Photo of the experimental setup (with illustrations). As shown on the screen, the target objects are arranged in a circle.

The visual stimulus consisted of a 30cm deep box that matches the horizontal dimensions of the tabletop setup (see Figure 1.3). We matched the look of the scene to the visual stimuli used by Teather and Stuerzlinger [TS11, TS13]. The targets in the experiment were represented by spheres, which were arranged in a circle as illustrated in Figure 1.3. A circle consisted of 11 spheres rendered in white, with the active target sphere highlighted in blue. The targets highlighted in the order specified by ISO 9241-9 [Int00]. The center of each target sphere indicated the exact position where participants were instructed to touch with their dominant hand in order to select a sphere. For 3D touch, this was the 3D position, and for 2D touch the center of the 2D projection. The size, distance and height of target spheres were constant within circles but varied between circles. Target height was measured as positive height from the level screen surface. Participants indicated target selection using a Razer Nostromo keypad with their non-dominant hand. The virtual scene was rendered on an Intel Core i7 3.40GHz computer with 8GB of main memory, and an Nvidia Quadro 4000 graphics card.

1.3.3 Methods

The experiment used a 2×5×2×2 within-subjects design with the method of constant stimuli, in which the target positions and sizes are not related from one circle to the next, but presented randomly and uniformly distributed [Fer08]. The independent variables were selection technique (2D touch vs. 3D mid-air touch), target height (between 0cm and 20cm, in steps

of 5cm), as well as target distance (16cm and 25cm) and target size (2cm and 3cm). Each circle represented a different index of difficulty (ID), with combinations of 2 distances and 2 sizes. The ID indicates overall task difficulty [Fit54]. It implies that the smaller and farther a target, the more difficult it is to select quickly and accurately. Our design thus uses four uniformly distributed IDs ranging from approximately 2.85bits to 3.75bits, representing an ecologically valuable range of difficulties for such a touch-enabled stereoscopic tabletop setup. As dependent variables, we measured the on- as well as off-display touch areas for 3D target objects.

The experiment trials were divided into two blocks: one for the 2D and one for the 3D touch technique. We randomized their order between participants. At the beginning of each block, participants were positioned standing in an upright posture in front of the tabletop surface as illustrated in Figure 1.3. To improve comparability, we compensated for the different heights of the participants by adjusting a floor mat below the participant's feet, resulting in an (approximately) uniform eye height of 1.85cm for each participant during the experiment. The experiment started with task descriptions, which were presented via slides on the tabletop surface to reduce potential experimenter bias. Participants completed 5 to 15 training trials with both techniques to ensure that they correctly understood the task and to minimize training effects. Training trials were excluded from the analysis.

In the experiment, participants were instructed to touch the center of the target spheres as accurately as possible (either with 2D or 3D touch), for which they had as much time as needed. For this, participants had to position the tip of the index finger of their dominant hand inside the 3D sphere for the 3D touch condition, or push their finger through the 3D sphere until it reached the 2D touch surface. Participants did not receive feedback whether they "hit" their target, i. e., participants were free to place their index finger in the real world where they perceived the virtual target to be. We did this to evaluate the often-reported systematical overestimation or underestimation of distances in virtual scenes, which can be observed even for short grasping-range distances [SM90], as also tested in this experiment. Moreover, we wanted to evaluate the impact of such misperceptions on touch behavior in stereoscopic tabletop setups. We tracked the tip of the index finger in both 2D and 3D touch conditions. When participants wanted to register the selection, they had to press a button with their non-dominant hand on the keypad. We recorded a distinct 2D and 3D touch position for each target location for each configuration of independent variables, with a total of 20 circles and 220 recorded touch positions per participant.

1.4 Results

In this section, we summarize the results from the 2D and 3D touch experiment. We had to exclude two participants from the analysis who obviously misunderstood the task. We analyzed these results with a repeated measure ANOVA and Tukey multiple comparisons at the 5% significance level (with Bonferonni correction).

1.4.1 2D Touch

For the 2D touch technique, we evaluated the judged 2D touch points on the surface relative to the potential projected target points, i. e., the midpoint (M) between the projections for both eyes, as well as the projection for the dominant (D), and the non-dominant (N) eye, as illustrated in Figure 1.4. Figure 1.5 shows scatter plots of the distribution of the touch points from all trials in relation to the projected target centers for the dominant and non-dominant eye for the different heights of 0cm, 5cm, 10cm, 15cm and 20cm (bottom to top). We normalized the touch points in such a way that the dominant eye projection D is always shown on the left, and the non-dominant eye projection N is always shown on the right side of the plot. The touch points are displayed relatively to the distance between both projections.

As it is illustrated in Figures 1.4 and 1.5, we observed three different behaviors when participants used the 2D touch technique. In particular, eight participants touched towards the midpoint, i. e., the center between the dominant and non-dominant eye projections. This includes the two participants for whom eye dominance estimates were inconclusive. We arranged these participants into the group GM. Furthermore, three participants touched towards the dominant eye projection D, which we refer to as group G_D, and three participants touched towards the non-dominant eye projection N, which we refer to as group G_N. This

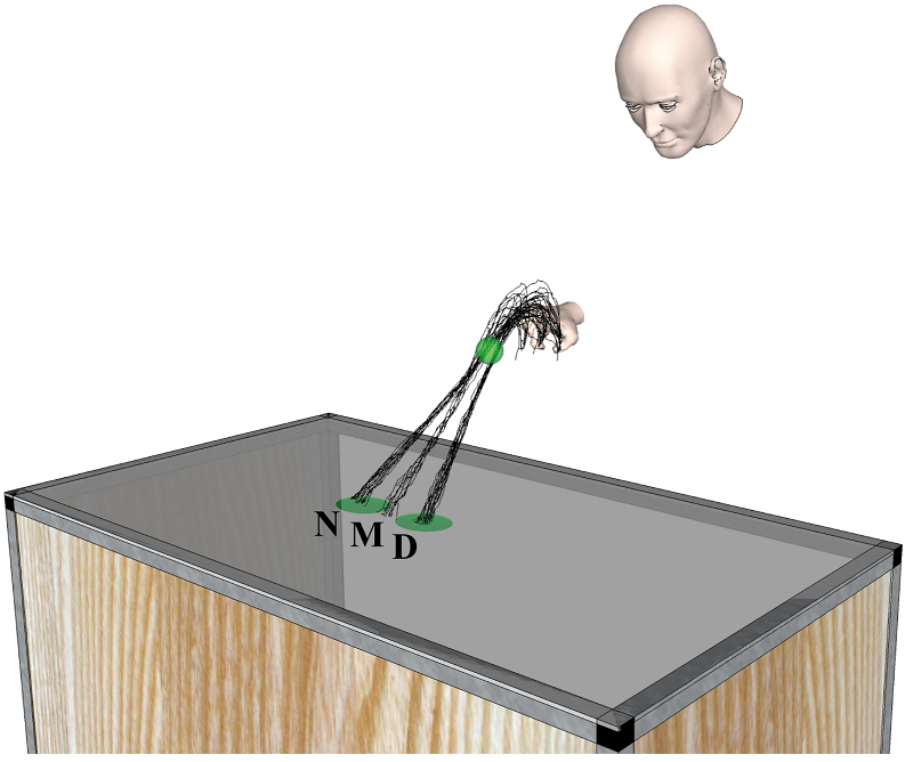


Figure 1.4: Illustration of finger movement trails for user groups touching towards the dominant eye projection (D), non-dominant eye projection (N), or towards the midpoint (M) using the 2D touch technique [BSS13b]. The trails have been normalized and are displayed here for a right-eye dominant user.

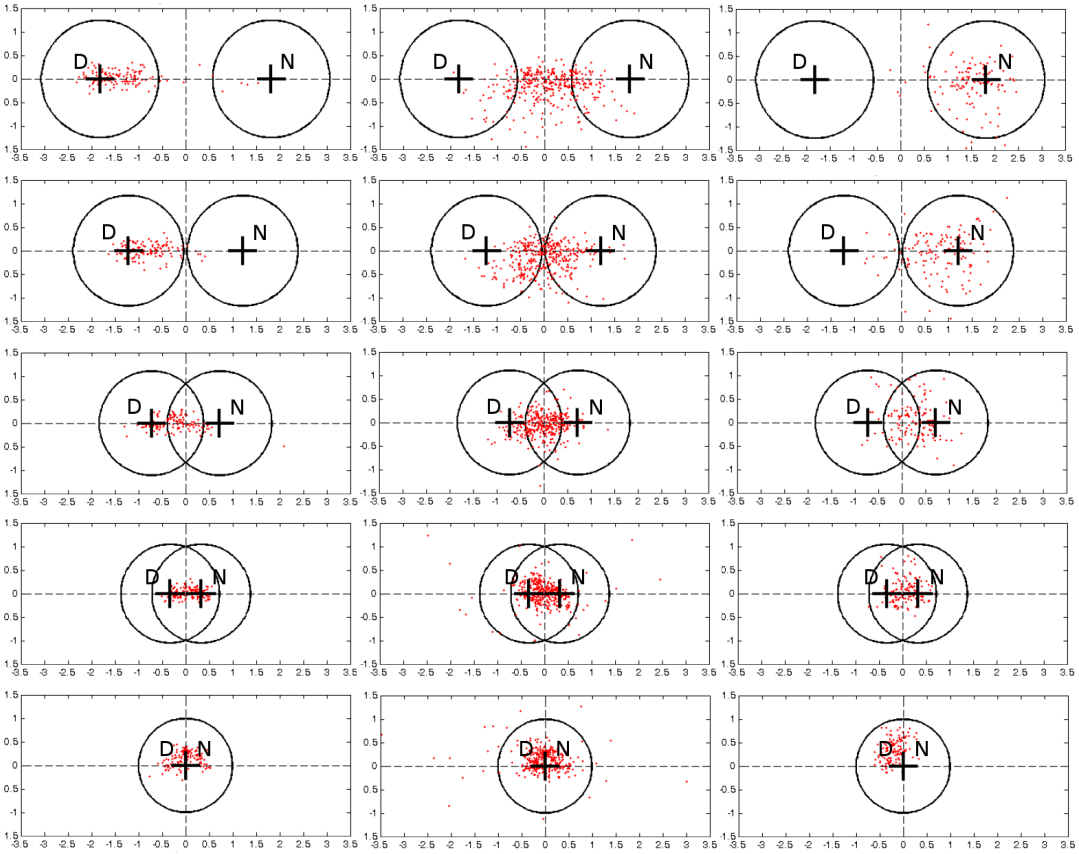


Figure 1.5: Scatter plots of relative touch points between the dominant (D) and non-dominant (N) eye projections of the projected target centers on the surface for the 2D touch technique. Black crosses indicate the two projection centers. Black circles indicate the approximate projected target areas for the dominant and non-dominant eye. Top to bottom rows show results for 20cm, 15cm, 10cm, 5cm, and 0cm target heights. The left column shows participant behavior for dominant-eye touches (3 participants), the middle for center-eye touches (8 participants), and the right for non-dominant-eye touches (3 participants). Note that the distance between the projection centers depends on the target height.

points towards an approximately 50/50% split in terms of behaviors in the population, i. e., between group GM and the composite of groups G_D and G_N.

We found a significant main effect of the three groups ($F(2,11)$=71.267, p<.001, η_p^2=.928) on the on-surface touch areas. Furthermore, we found a significant two-way interaction effect of the three groups and target heights ($F(8,44)$=45.251, p<.001, η_p^2=.892) on the on-surface touch areas. The post-hoc test revealed that the on-surface target areas, see Figure 1.5, significantly (p<.001) vary for objects that are displayed at heights of 15cm or higher. For objects displayed at 10cm height group G_D and G_N vary significantly (p<.02). For objects displayed below 10cm, we could not find any significant difference. As illustrated in Figure 1.5, for these heights the projections for the dominant and non-dominant eye are close together, and participants touched almost the same on-screen target areas.

Considering the on-surface touch areas, we found that on average the relative touch point for

group G_D was 0.97D+0.03N for projection points D$\in\mathbb{R}^2$ and N$\in\mathbb{R}^2$, meaning the participants in this group touched towards the projection for the dominant eye, but slightly inwards to the center. The relative touch point for group G_N was 0.11D+0.89N, meaning the participants in this group touched towards the projection for the non-dominant eye, again with a slight offset towards the center. Finally, for group GM we found that on average the relative touch point for this group was 0.504D+0.596N. We could not find any significant difference for the different heights, i. e., the touch behaviors were consistent throughout the tested heights.

However, we observed a trend of target height on the standard deviations of the horizontal distributions (x-axis) of touch points for all groups as shown in Figure 1.5. For 0cm target height we found a mean standard deviation (SD) of 0.29cm, for 5cm SD 0.32cm, for 10cm SD 0.42cm, for 15cm SD 0.52cm, and for 20cm SD 0.61cm. For the vertical distribution (y-axis) of touch points and at 0cm target height we found a mean SD of 0.20cm, for 5cm SD 0.20cm, for 10cm SD 0.25cm, for 15cm SD 0.29cm, and for 20cm SD 0.30cm.

In summary, the results for the 2D touch technique show a significant effect for the different user groups on the on-surface touch area over the range of tested heights. These on-surface touch areas vary significantly for objects displayed at heights of 10cm and higher.

1.4.2 3D Touch

We analyzed the tracked physical 3D "touch" points where participants judged the perceived center of the mid-air target spheres for the 3D touch technique in terms of their deviation from their actual position in the 3D virtual scene. Figure 1.6 shows scatter plots of the distribution of judged target positions in relation to the 3D target centers for the different target heights over all trials. The red dots indicate the center positions of the spheres as judged by the participants. The black wireframe spheres illustrate the actual position and size of the objects. We normalized the judged positions relative to the optical view angle towards the target center. We found no significant difference in the judged positions for the three groups identified in Section 1.4.1 and pooled the data.

We analyzed the effect of target height on the participants' judgments. We found a significant main effect of target height on the distances of judged positions from the displayed target centers. Mauchly's test indicated that the assumption of sphericity had been violated for effects of height on the distances of judged positions ($\chi^2(9)$=62.388, p<.001), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ϵ=.302). The results show that the distances of judged positions significantly differs for heights ($F(1.21, 15.725)$=12.846, p<.002, η_p^2=.497).

A Tukey post-hoc test with Bonferroni correction revealed that participants estimated the target centers significantly closer to the actually displayed target centers for the 0cm targets in comparison to targets displayed at 20cm height (p<.002). For all other heights, the results suggest that the higher the targets are displayed, the larger are the deviations. Pooling over all participants, we observed mean distances to target centers of M=0.56cm (SD=0.27cm) for 0cm target height, M=0.88cm (SD=0.53cm) for 5cm, M=0.97cm (SD=0.61cm) for 10cm,

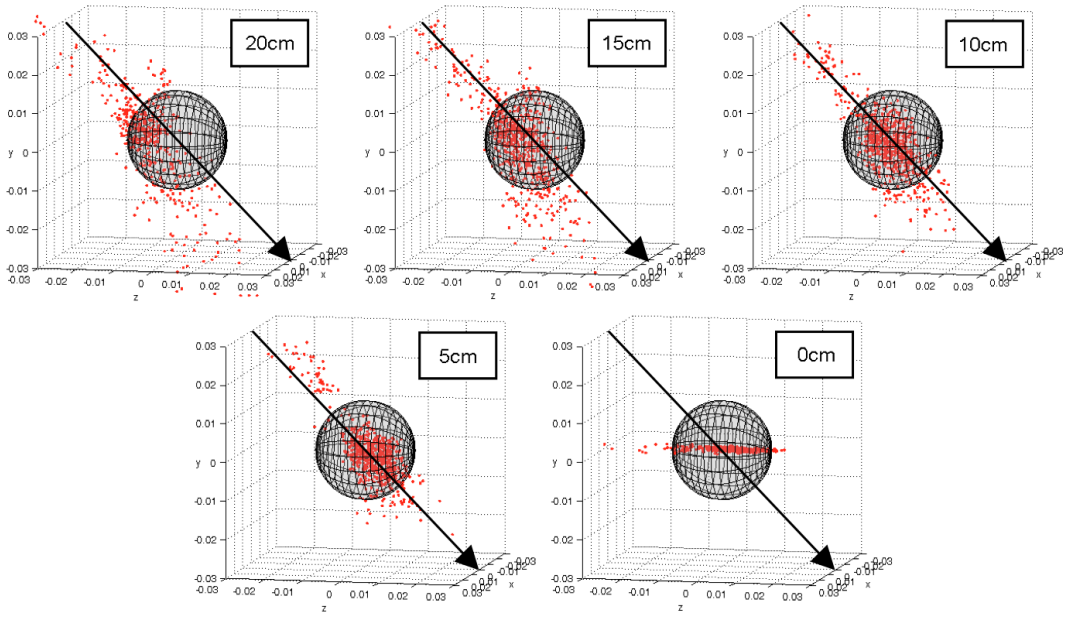


Figure 1.6: Scatter plots of judged positions of the 3D target centers for the 3D touch technique over all participants. Black wireframe spheres indicate the targets. The diagonal arrow illustrates the normalized view angle. The five diagrams show results for 20cm, 15cm, 10cm, 5cm, and 0cm target heights.

M=1.32cm (SD=0.93cm) for 15cm, and M=1.90cm (SD=1.48cm) for 20cm. The results suggest that the physical constraints provided by the touch surface at 0cm height reduced judgment errors for objects at zero parallax relative to the other heights. We found no significant difference when comparing to the results for the 2D touch technique at 0cm target height as presented in Section 1.4.1.

As it can be seen in Figure 1.6, participants made larger errors along the view axis than along the orthogonal axes. For the mid-air target positions, we found a mean standard deviation of 1.43cm along the optical line-of-sight, a mean SD of 0.36cm parallel to the touch surface, and a mean SD of 0.50cm orthogonal to the other axes. Furthermore, these deviations increased with increasing target heights. For the different target heights above the surface we observed standard deviations of judged positions along the optical line-of-sight of SD=2.20cm (for 20cm target height), SD=1.52cm (15cm), SD=1.05cm (10cm), and SD=0.94cm (5cm). On the other hand, we observed standard deviations of judged positions orthogonal to the view axis parallel to the touch surface of only SD=0.49cm (20cm), SD=0.39cm (15cm), SD=0.30cm (10cm), and SD=0.27cm (5cm). Finally, we found standard deviations of judged positions orthogonal to the other axes of only SD=0.70cm (20cm), SD=0.55cm (15cm), SD=0.41cm (10cm), and SD=0.35cm (5cm). We further analyzed the data to determine whether deviations in judged target positions result from underestimation or overestimation of distances from the observer to the mid-air targets [BSS13a, CKC+10]. We observed a mean distance underestimation of 0.25% (SD=2.93%). Surprisingly, we found a distance overestimation of M=0.4% (SD=2.00%) and M=1.0% (SD=2.25%) for heights of 5cm and 10cm, respectively.

Yet, we found an underestimation of M=−0.54% (SD=2.67%) and M=−0.98% (SD=4.18%) for heights of 15cm and 20cm, respectively.

In summary, the results for the 3D touch technique show a significant effect of stereoscopic parallax over the range of tested heights on the precision and accuracy of judging the position of a target object.

1.5 Discussion

Our results provide interesting guidelines on how touch interaction in 3D stereoscopic tabletop setups should be realized. First of all and in contrast to previous work [VSBH11], our results show evidence for a twofold diversity of 2D touch behaviors of users. As shown in Figure 1.5, roughly half of the participants in our study touched through the virtual object towards the center between the projections, and the other half touched towards projections determined by a single eye. The second group roughly splits in half again depending if they touch the projection for the dominant or non-dominant eye. Our results differ from the findings by Valkov et al. [VSBH11]. Using a setup with a large vertical projection plane they observed that participants touched towards the center projection, with a slight offset towards the dominant eye. With 3 participants touching towards the dominant eye, and 3 participants towards the non-dominant eye in our study, user behavior in tabletop environments cannot be explained by this model. As a guideline, we suggest that the center between the projection for the left and right eye can be used to detect selections of objects stereoscopically displayed with less than 10cm height since we did not observe significant differences between participants at such heights. In order to reliably detect selections for objects higher above the screen, i. e., with larger parallaxes, our results suggest that for each user a calibration would be required. Our results confirm that this approach is highly beneficial since participants touched *consistently* for all heights towards the dominant, center, or non-dominant projection.

For practical considerations and to evaluate the ecological validity of using the 2D touch technique for selections of targets at a height between 0cm and 10cm, we computed the minimal on-surface touch area that supports 95% correct detection of all 2D touch points in our experiment. Due to the similar distributions of touch points between the three behavior groups for these heights shown in Figure 1.5, we determined the average minimal 95% on-surface region over all participants. Our results show that an elliptical area with horizontal and vertical diameter of 1.64cm and 1.07cm with a center in the middle between the two projections is sufficient for 95% correct detection. This rule-of-thumb heuristic for on-surface target areas is easy to implement and ecologically valuable considering the fat finger problem [HCRC88, VB04]. Due to this problem objects require a relatively large size of between 1.05cm to 2.6cm for reliable acquisition, even in monoscopic touch-enabled tabletop environments.

The results of our second experimental condition reveal that distinct differences exist between the 3D mid-air touch technique and the 2D touch technique. These differences impact the relative performance and applicability for interaction with objects displayed stereoscop-

ically at different heights above the surface. We found no behavior groups or effects of eye dominance on the distribution of judged 3D target positions. Our results show that target height has an effect on precision and accuracy of 3D selections, with large errors mainly along the optical line-of-sight, which we believe to correlate with distance misperception. For 3D objects displayed close to the display surface up to 10cm, touching objects in 2D on the surface by touching "through" the stereoscopic impression is more accurate than 3D mid-air touching. Considering that much research has shown that 3D mid-air touches of virtual objects suffer from low accuracy and precision due to visual conflicts, including vergence-accommodation mismatch, diplopia, and distance misperception [BSS13a, CKC+10], it is a promising finding that the reduction of 3D selection tasks to 2D input with the 2D touch technique can improve performance for tabletop surface with stereoscopically displayed objects. However, the results also show that the accuracy for 2D touching of objects displayed above the screen decreases significantly for large negative parallax. The findings are encouraging for stereoscopic visualization on (multi-)touch surfaces. They suggest that virtual objects do not have to be constrained exactly at the zero-parallax level, but may deviate up to 10cm before 2D touch accuracy is significantly degraded [VSB+10, VSBH11]. For such distances, the 2D touch technique is a good choice and instrumenting users with gloves or 3D markers can be avoided. Overall, our results show that it is possible to leverage stereoscopic cues in tabletop setups for an improved spatial cognition.

As a guideline for future tabletop setups with direct 2D touch input, the results suggest that touch-enabled 3D objects should not be displayed above an interactive display surface at more than about 10cm height. Above that, the disadvantages outperform the benefits and 3D interaction techniques should be used in that region, as they will provide more accurate interaction possibilities.

1.6 Example Application: Stereoscopic 3D Widgets

Our experiments have shown that the 2D touch technique has enormous potential as a new interaction paradigm for stereoscopic multi-touch surfaces as long as the objects are displayed with less than 10cm above the surface. In this region, our 2D touch technique is a more accurate choice. While this constraint appears to limit the application scenarios in which one could use the 2D touch technique, it also ensures a simple implementation for interaction, in particular, a clear definition of on-display target areas as described in Section 1.5. Moreover, the size and scale of many virtual objects used in actual tabletop applications suit this constraint. For instance, 3D widgets can be displayed stereoscopically on any multi-touch surface and provide the user with a natural haptic feedback experience when she virtually touches them.

In order to evaluate the quality of the 3D touch technique in a real-world application, we adapted a simple visualization application for virtual caravans (see Figure 1.7). With this application, customers can evaluate various types of caravans with several different features. The 3D widgets on the menu plane allow users to change the visual appearance of the caravan,

Figure 1.7: User interacting with a virtual scene in a stereoscopic multi-touch tabletop setup using touch-enabled 3D widgets. The widgets in the graphical user interface were rendered with negative parallax of up to 10cm height.

lighting parameters, turn on signals, headlamps etc. We implemented the on-surface target areas of these 3D widgets as described in Section 1.5. The highest widgets, i. e., the 3D buttons on the menu panel, are displayed about 10cm above the surface. We used the same physical setup as described in Section 1.3.2. For this application, we used the Unity3D game engine for the generation and rendering of the virtual scene. Unity3D provides a simple development environment for virtual scenes, animations and interactions. In order to synchronize virtual camera objects with the movements of a user, we integrated the MiddleVR for Unity software framework. MiddleVR supports streaming of motion data from our tracking system to Unity3D using the Virtual Reality Peripheral Network (VRPN) protocol. With this we stream head poses to Unity3D, resulting in a correct perspective from the user's point of view at all times.

We presented this application to four users and made several interesting observations. First, all users acknowledged the stereoscopic display when viewing the 3D scene. Second, most users immediately understood that the menu panel with the 3D widgets provides a means to interact with the setup. Surprisingly, when users tried to "touch" the 3D widgets, they adapted their actions to the affordances provided by the widget. For instance, when they pressed the toggle switch, usually they touched its lifted part, although we did not distinguish between touch positions on the surface. We see this as further indication that stereoscopic display in combination with a touch-enabled surface does indeed support the notion of 3D physical interaction elements. Finally, none of the users complained about non-reactive 3D widgets, which might have occurred if they missed the on-surface target areas. This suggests that the shape and size of the on-surface touch areas, as determined by our above study, is sufficient for using stereoscopic 3D widgets in tabletop setups.

1.7 Conclusion

In this chapter, we evaluated and compared 2D touch and 3D touch interaction techniques for scenes on touch-sensitive tabletop setups with stereoscopic display. We analyzed the differences of 3D mid-air touch input and a technique based on reducing the 3D touch problem to two dimensions by having users touch "through" the stereoscopic impression of 3D objects, resulting in a 2D touch on the display surface. We identified two separate classes of user behavior, with one group that touches the center between the projections, whereas the other touches the projection for the dominant or non-dominant eye. The results of the experiment show a strong interaction effect between input technique and the stereoscopic parallax of virtual objects.

The main contributions of this work are:

- We identified two separate classes of user behavior when touching "through" stereoscopically displayed objects.
- We compared precision and accuracy of 2D/3D direct touch input, which revealed that the 2D touch technique is a viable alternative to 3D touch interaction for object selection up to about 10cm height from the display surface.
- We determined on-surface target regions that support a simple implementation of the 2D touch technique. This enables intuitive touch input for 3D objects and widgets in stereoscopic 3D tabletop applications.

The results are encouraging for stereoscopic visualization in future touch-enabled tabletop setups since no additional instrumentation and tracking technology is needed for objects with a small stereoscopic parallax. An interesting question for future work is if the results can be applied to portable setups, where the orientation of the touch-sensitive surface varies during the interaction. We plan to further pursue these topics to provide compelling user experiences and effective user interfaces for touch-sensitive stereoscopic display surfaces. Moreover, we plan to investigate also how the 2D and 3D touch methods compare in terms of the speed-accuracy tradeoff.

Chapter 2

2 To Touch or not to Touch? Comparing 2D Touch and 3D Mid-Air Interaction on Stereoscopic Tabletop Surfaces

Recent developments in touch and display technologies have laid the groundwork to combine touch-sensitive display systems with stereoscopic three-dimensional (3D) display. Although this combination provides a compelling user experience, interaction with objects stereoscopically displayed in front of the screen poses some fundamental challenges: Traditionally, touch-sensitive surfaces capture only direct contacts such that the user has to penetrate the visually perceived object to touch the 2D surface behind the object. Conversely, recent technologies support capturing finger positions in front of the display, enabling users to interact with intangible objects in mid-air 3D space. In this chapter, we perform a comparison between such 2D touch and 3D mid-air interactions in a Fitts' Law experiment for objects with varying stereoscopical parallax. The results show that the 2D touch technique is more efficient close to the screen, whereas for targets further away from the screen, 3D selection outperforms 2D touch. Based on the results, we present implications for the design and development of future touch-sensitive interfaces for stereoscopic displays.

2.1 Introduction

Two different technologies dominated recent exhibitions and the entertainment market: (multi-)touch-sensitive surfaces and 3D stereoscopic displays. These technologies have the potential to provide more intuitive and natural interaction setups for a wide range of areas, including geo-spatial applications, urban planning, architectural design, or collaborative tabletops. These two technologies are orthogonal, as (multi-)touch is about *input* and 3D stereoscopic visualization about *output*. First commercial hardware systems have recently

been launched (e. g., Nintendo 3DS), and interdisciplinary research projects explore interaction with stereoscopic content on 2D touch surfaces (e. g., [iMU13, InS13]). 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 (e. g., Leap Motion). The combination of these novel technologies provides enormous potential for a variety of new interaction concepts.

Until recently, research in the area of (multi-)touch interaction was mostly focused on monoscopically displayed data. There, the ability to directly touch elements has been shown to be very appealing for novice as well as expert users. Also, passive haptics and multi-touch capabilities have both shown their potential to improve the user experience [BWB06]. Touch surfaces build a consistent and pervasive illusion in perceptual and motor space that two-dimensional graphical elements on the surface can be touched. Yet, three-dimensional data limits this illusion of place and plausibility [Sla09]. 3D data sets are either displayed monoscopically, which has been shown to impair spatial perception in common 3D tasks, or stereoscopically, which can enrich the experience and interaction, but causes objects to appear detached from the touch surface [MHG12, SSV+09].

Stereoscopic display technology has been known for decades. It has recently been revived in the rise of 3D cinema and 3D televisions. With stereoscopic displays, each eye sees a different perspective of the same scene through appropriate technology. This requires showing two distinct images on the display. Objects may be displayed with *negative*, *zero*, or *positive* parallax, corresponding to in front, at, or behind the screen. Objects with centroid at *zero parallax* appear attached to the screen and are perfectly suited for touch interaction. In contrast, it is more difficult to apply direct-touch interaction techniques to objects that appear in front of or behind the screen [HCC07, PFC+97, RDH09]. In this chapter we focus on the major challenge in this context, namely objects that appear in front of the screen such as a virtual object floating above the surface within the user's personal interaction space [DCJH13]. Teather and Stuerzlinger [TS11] provide a review of interaction techniques for distant objects behind the screen.

Two methodologies can be used for interacting with stereoscopic objects in front of a tabletop display:

1. If the touch-sensitive surface captures only direct contacts, the user has to penetrate the visually perceived object to touch the 2D surface behind the object [VSBH11, VSB+10].

2. Alternatively, if finger poses in front of the screen can be captured, the user can directly interact with the intangible object in 3D space.

Due to the discrepancy between perceptual and motor space and missing haptic feedback, both approaches provide natural feedback only for objects rendered with zero parallax. One question posed by this issue is where users "touch" a stereoscopically displayed intangible object in 3D space, considering the misperception of distances in virtual 3D scenes [LK03]. Conversely, it also brings up the issue where users "touch" a stereoscopically displayed object on a 2D display surface, considering that there are two distinct projections for each

eye [VSBH11]. If the user penetrates the object while focusing on her finger, the stereoscopic impression of the object is disturbed, since the user's eyes are not accommodated and converged to the display surface. Thus, the left/right image pairs of the object appear blurred and can potentially not be merged (see Figure 1.1(a)). Yet, focusing on the virtual object causes a disturbance of the stereoscopic perception of the user's finger, since her eyes are converged on the object's 3D position (see Figure 1.1(b)). When the user selects an object in 3D space, by holding her finger in front of the screen, she can see a stereoscopic image while converging to her finger. However, due the vergence-accommodation conflict, the virtual object will appear blurred in comparison to the real finger (see Figure 1.2).

In this chapter we address the challenge of how to interact with stereoscopic content in front of a touch-sensitive tabletop surface. We evaluate interaction with touch-sensitive screens to select a 3D object, and compare this approach to systems where the user's finger is tracked in 3D space. We use a Fitts' Law experimental design to determine differences in 3D object selection performance for varying object parallax in front of the screen. The results of this experiment provide guidelines for the choice of touch technologies, as well as the optimal placement and parallax of interactive elements in stereoscopic touch environments.

Our contributions are:

- A direct comparison of the performance of 2D touch and 3D mid-air selection for different spatial configurations of interactive 3D objects.
- Guidelines for designing user interfaces for stereoscopic touch-sensitive tabletop setups.

The remainder of this chapter is structured as follows. Section 2.2 summarizes background information on touch interaction and stereoscopic display. Section 2.3 describes the experiment we conducted to evaluate and compare 2D/3D interaction performance. Section 2.4 presents the results, which are discussed in Section 2.5. Section 2.6 concludes the chapter.

2.2 Background

Recently, many approaches for extending multi-touch interaction techniques to 3D applications with *monoscopic* display have been proposed [HCC07, MCG10, PHH12, RDH09, WIH+08]. In order to extend interaction possibilities with monoscopic 2D surfaces, Hancock et al. [HCC07] presented approaches for 3D interaction within a limited range above the surface. Yet, interaction with stereoscopically displayed scenes introduces new challenges [SSV+09], since the displayed objects can float in front of or behind the interactive display surface.

2.2.1 Interaction with Stereoscopic Objects

In this section we describe work related to interaction with stereoscopically displayed objects. In particular, we discuss 2D touch and 3D mid-air selection techniques.

3D Mid-Air Interaction Techniques

To enable selection of stereoscopically displayed 3D objects in space, 3D tracking technologies capture a user's hand or finger motions in front of the display surface. The kinematics of point and grasp gestures in 3D space and the underlying cognitive functions have been studied [GCE08, MMD+87, WWG03]. For instance, it has been shown that the arm movement during grasping consists of two distinct phases: (1) an initial, *ballistic phase* during which the user's attention is focused on the object to be grasped (or touched). The motion is essentially controlled by proprioception, and (2) a *correction phase* that reflects refinement and error-correction of the movement, incorporating visual feedback in order to minimize the error between the hand or finger and the target [LCE08]. MacKenzie et al. [MMD+87] investigated real time kinematics of limb movements in a Fitts' task and showed that, while Fitts' Law holds for the total limb-movement time, humans decelerate the motion sooner, if the target seems to require more precision in the end phase. The changes of the kinematics and control for reaching tasks within virtual environments have been investigated [DKK07, VFML04].

Hilliges et al. [HIW+09] investigated extending the interaction space beyond the touch surface. They tested two depth-sensing approaches to enrich multi-touch interaction on a tabletop with monoscopic display. Although 3D "mid-air" interaction provides an intuitive technique, it has been shown that touching an intangible object, i. e., *touching the void* [CKC+10], leads to confusion and a significant number of overshoot errors. This is due to the fact that depth perception is less accurate in virtual scenes compared to the real world, as well as the introduced double vision and vergence-accommodation conflicts. Bruder et al. [BSS13a] investigated the effects of visual conflicts on 3D selection performance with stereoscopic tabletop displays. Some devices, such as the CyberGrasp, support haptic feedback when touching objects in space, but require extensive user instrumentation. Other approaches are based on the user moving tangible surfaces in 3D space to align with floating objects, e. g., through transparent props [CKC+10], or on controlling the 3D position of a cursor through multiple touch points [BF07, SVH11]. Toucheo uses 2D projections to define widget for interaction with objects presented stereoscopically above a multi-touch display [HBCd11]. Yet, the projection direction for Toucheo is straight down towards the display surface. This paradigm does not work well for objects that are stacked one above the other, as their projections then conflict.

2D Touch Techniques

Recently, multi-touch devices with non-planar surfaces, such as cubic [dKOD08] or spherical [BWB08], were proposed. These can specify 3D axes or points for indirect object manipulation. Interaction with objects with negative parallax on a multi-touch tabletop setup was addressed by Benko et al.'s balloon selection [BF07], as well as Strothoff et al.'s triangle cursor [SVH11], which use 2D touch gestures to specify height above the surface.

Valkov et al. [VSBH11] performed a user study, in which they displayed 3D objects stereoscopically in front or behind a large vertical projection screen. They instructed users to touch

the virtual 3D objects by touching *through* the objects until their finger hit the display surface and recorded user behavior. This study found that users tended to touch between the projections for the two eyes with an offset towards the projection for the dominant eye. Bruder et al. [BSS13b] further analyzed stereoscopic 2D touch interaction and identified three distinct user behaviors (see Figure 1.4): users consistently touched either towards the dominant eye projection, the non-dominant one, or the midpoint between the projections. While these three behaviours varied between participants, they found little within-participants variation.

In a different study, Valkov et al. [VSB+10] showed that users are, within some range, insensitive to small misalignments between visually perceived stereoscopic positions and the sensed haptic feedback when touching a virtual object. Moreover, users are less sensitive to discrepancies between visual and tactile feedback for objects with negative parallax. They proposed to manipulate the stereoscopically displayed scene so that objects are moved towards the screen when the user reaches for them [VGH12, VSB+10]. This only works for objects displayed close (approximately 5cm) to the surface. Yet, the problem is that objects have to be shifted in space, which leads to a disturbed perception of the virtual scene for larger manipulations.

So far, no comparative analysis exists for 2D touch and 3D mid-air interaction in stereoscopic tabletop setups. Thus, it remains unclear if 2D touch is a viable alternative to 3D mid-air selection.

2.2.2 Fitts' Law and Selection

Fitts' Law [Fit54] is a well-known empirical model for user performance in selection tasks. The model predicts the movement time MT for a given target distance D and size W by $MT = a + b \times \log_2(D/W+1)$; where a and b are empirically derived. The log term is the *index of difficulty (ID)* and indicates overall task difficulty. This implies that the smaller and farther a target, the more difficult it is to select accurately. A valuable extension supported by an international standard [Int00] is the use of "effective" measures. This post-experiment correction adjusts the error rate to 4% by re-sizing targets to their effective width (W_e). This enables the computation of effective throughput, a measure that incorporates both speed and accuracy, by "normalizing" the accuracy as effective scores. This *throughput* is computed as $TP = \log_2(D_e/W_e + 1)/MT$, where D_e is the effective distance (average of measured movement distances), and W_e the effective width (standard deviation of error distances multiplied by 4.1333 [MI08]). Previous 3D research [TS11] suggests that one should use the point closest to the target along the ray to compute an accurate representation of the effective width W_e, as using the actual 3D cursor position would artificially inflate the effective measure. In essence, this suggestion projects the 3D task into 2D before computing throughput for touch-based interaction techniques. Even more recent work [TS13] reveals that the distortion due to perspective also has an effect. This work recommends the use of the 2D projections of sizes and distances to compute a screen-projected throughput for all *remote-pointing* techniques, such as ray-pointing.

2.3 Experiments

Here we describe our experiments to compare the performance of 2D touch and 3D mid-air interaction. We used a Fitts' Law selection task on a tabletop setup with 3D targets displayed on the surface or at different heights above the surface, i. e., with different negative stereoscopic parallax.

2.3.1 Experimental Setup

For the experiment we used a 62cm×112cm active stereoscopic multi-touch tabletop setup. The system is shown in Figure 1.3. The setup uses a matte diffusing screen with a gain of 1.6. For stereoscopic back projection screen we use a 1280×800 Optoma GT720 projector at 120Hz. The active DLP-based shutter glasses are driven by the projector at 60Hz per eye. We use an optical WorldViz Precision Position Tracking X4 system with sub-millimeter precision and accuracy to track the participant's finger and head for view-dependent rendering. For this, we attached wireless markers to the shutter glasses and another diffused IR LED on the tip of the index finger of the participant's dominant hand. We tracked and logged both head and fingertip movements during the experiment. The view of the 3D scene was rendered stereoscopically using off-axis projections. We measured an end-to-end latency of approximately 55ms between physical movements and a visual response.

The visual stimulus used in the experiment is a 3D scene in a 30cm deep box, fit to the horizontal dimensions of the physical tabletop setup (see Figure 1.3). We matched the look of the scene to the visual stimuli used in [BSS13b, BSS13a, TS13, TS11] for improved comparability. The targets in the experiment were represented by spheres, arranged in a circle (see Figure 1.3). A circle consisted of 11 spheres rendered in white, with the active target sphere highlighted in blue. The targets highlighted in the order specified by ISO 9241-9 [Int00]. The center of each target sphere indicated the exact position where participants were instructed to touch with their dominant hand in order to select a sphere. Participants indicated target selection using a Razer Nostromo keypad with their non-dominant hand. The target spheres highlighted green when the finger of the user was within the target to provide participants with feedback about successful selection, to minimize systematic errors in Fitts' Law experiments [MMD+87]. Head-tracked off-axis stereoscopic display was active in all conditions. The size, distance, and height of target spheres were constant within circles, but varied between circles. In other words, targets were at a constant height for each circle of targets. Target height was measured upwards from the level screen surface. All target spheres were presented with positive height, i. e., in front of the screen. The virtual environment was rendered on an Intel Core i7 computer with 3.40GHz processors, 8GB of main memory, and an Nvidia Quadro 4000 graphics card.

2.3.2 Methods

The experiment used a 2 × 5 × 2 × 2 within-subjects design with the method of constant stimuli. The independent variables were selection technique (2D touch vs. 3D mid-air interaction), target height (0cm to 20cm, in steps of 5cm), as well as distances between targets (16cm and 25cm) and size (2cm and 3cm). Each circle represented a different index of difficulty with combinations of 2 distances and 2 sizes. This yielded four uniformly distributed IDs ranging from approximately 2.85bits to 3.75bits, representing an ecologically valuable range of Fitts' Law task difficulties for a touch screen setup. Each circle used one of 5 different target height, between 0cm and 20cm in steps of 5cm. Distances between targets, sizes and heights were not related from one circle to the next, but presented randomly and uniformly distributed. The dependent variables were movement time, error distance, error rate (percentage of targets missed), and effective throughput.

The experiment trials were divided into two blocks: one for 2D touch selections and one for 3D mid-air selections. We randomized their order between participants. At the beginning of each block, participants were positioned standing in an upright posture in front of the tabletop surface (see Figure 1.3). To remove a potential confound in terms of target visibility and view angle, we compensated for the different heights of participants by adjusting the height of a floor mat below the participant's feet, resulting in an eye height of about 185cm for all participants during the experiment. The experiment started with task descriptions, which were presented via slides on the projection surface in order to reduce potential experimenter biases. Participants had to complete 5 to 15 training trials for both techniques to minimize later training effects. These training trials were excluded from the analysis. In order to compensate for misperceptions of the targets, we performed a calibration phase based on Bruder et al. [BSS13b]. During this calibration, participants were instructed to touch the center of the target spheres as accurately as possible with 2D touch as well as 3D mid-air selection. Participants had as much time as needed and they were free to place their index finger in the real world where they perceived the virtual target to be. We used the resulting calibrated positions to define the target centers in the Fitts' Law trials for each participant as described in [BSS13b, BSS13a].

After the calibration, participants were instructed to select the targets as quickly and accurately as possible, a common instruction in Fitts' Law experiments [TS13, TS11]. Participants received visual feedback when their finger was inside a target, by targets turning green. Then, participants indicated selection by pressing a key with their non-dominant hand. If participants pressed the key while the target sphere was not green, we recorded this as a selection error and advanced the trial state. We computed the distance of the position of the tip of the index finger to the calibrated sphere center. A valid 3D selection occurred if this distance was less than the sphere radius for 3D mid-air interactions. For 2D touch interactions, we computed the projected 3D target position and size on the 2D touch surface (see Figure 1.4). Then we judged a 2D touch selection to be valid if the finger position was within the projected circle (cf. [VSBH11]). There were 11 recorded target selections per circle. Circles were shown

twice to each participant in randomized order for each configuration of independent variables. Each participant completed a total of 80 circles, with 880 recorded target selections.

In addition to the performance data collected in the Fitts' Law trials, we also asked participants to judge various characteristics of the techniques through subjective questionnaires. Before and after the 2D/3D interaction conditions, participants were asked to complete a Simulator Sickness Questionnaire (SSQ). Moreover, asthenopia, visual discomfort symptoms, were measured with a questionnaire about blurred vision, ocular soreness, itching of the eyes, increased blinking, heaviness of the eyes, and double vision on 4-point scales (0=none, 1=slight, 2=moderate, 3=severe), i. e., analogous to the SSQ sickness symptoms. After each technique, participants were asked to complete a Slater-Usoh-Steed (SUS) presence questionnaire, a NASA TLX mental workload questionnaire, as well as a general usability questionnaire, in which we asked participants to judge the technique according to the criteria learnability, efficiency, memorability, errors, and satisfaction on 5-point Likert scales.

2.3.3 Participants

10 male and 5 female participants (ages 20-35, M=27.1) completed the experiment. Participants were students or members of the local university. 3 participants received class credit for participating in the experiment. All participants were right-handed. All participants had normal or corrected to normal vision. 1 participant wore glasses and 4 participants wore contact lenses during the experiment. None of the participants reported known eye disorders, such as color weaknesses, amblyopia or known stereopsis disruptions. We verified the ability for stereoscopic vision of all participants. We measured the interpupillary distance (IPD) of each participant before the experiment [WGTCR08], which revealed IPDs between 5.8cm and 7.0cm (M=6.4cm). We used each individual's IPD for stereoscopic display in the experiment. 14 participants reported experience with stereoscopic 3D cinema, 14 with touch screens, and 8 had previously participated in a study involving touch surfaces. Participants were naïve to the experimental conditions. Participants were allowed to take a break at any time between trials to minimize effects of exhaustion or lack of concentration. The total time per participant was about 1.5 hours.

2.4 Results

Here we summarize the results of the experiment. We had to exclude two participants from the analysis who misunderstood the task (i. e., showed 100% incorrect selections). All other trials have been included in the analysis. As stated above, we used for each participant the calibrated target positions as valid target centers. Results 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 (with Bonferonni 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.

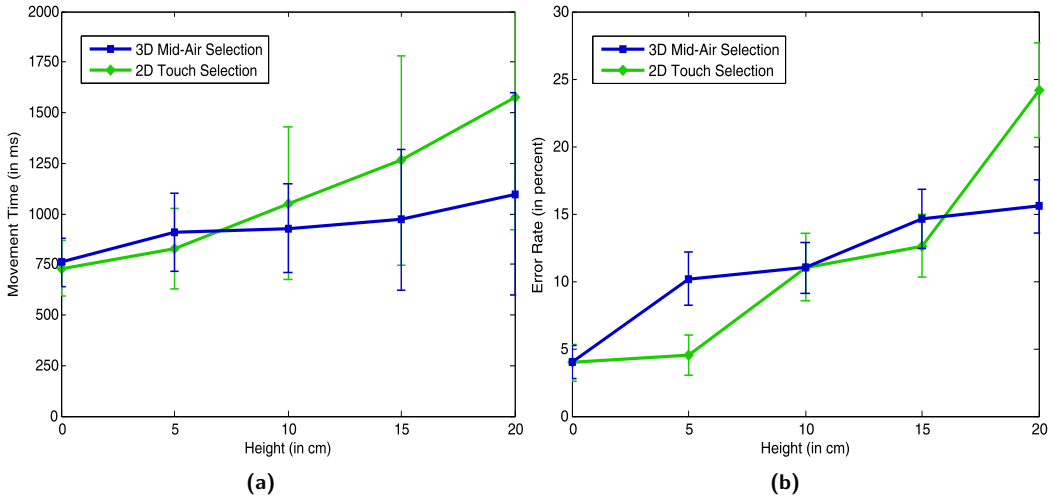


Figure 2.1: Results for Fitts' Law trials with target object height on the horizontal axis and pooled for (a) movement time and (b) error rate on the vertical axis. The error bars show the standard error.

2.4.1 Movement Time

The results for the movement time are illustrated in Figure 2.1(a). We found no significant main effect of technique ($F(1,12)$=3.870, p>.05, η_p^2=.244) on movement time. The average movement time during the experiment was M=1090ms (SD=521ms) for 2D touch, while 3D selection had M=934ms (SD=324ms).

The results show that the movement time for heights differs significantly ($F(1.272,15.265)$=27.127, p<.001, η_p^2=.693). Post-hoc tests revealed that the movement time was significantly increased when objects were displayed with heights of 15cm (p<.05) or 20cm (p<.001) in comparison to 0cm. As expected, we found a significant main effect of the ID on movement time ($F(1.220,14.635)$=23.061, p<.001, η_p^2=.658).

We found a significant two-way interaction effect between technique and height ($F(1.360,16.319)$=9.453, p<.01, η_p^2=.441). Post-hoc tests revealed that participants took significantly longer with 2D touch than 3D selection when objects were displayed with a height of 20cm (p<.05). We found no significant difference between the techniques for lower heights.

2.4.2 Error Rate

The results for error rate are illustrated in Figure 2.1(b). We found no significant main effect of technique ($F(1,12)$=0.009, p>.05, η_p^2=.001) on error rate. The average error rate during the experiment was M=11.6% (SD=18.5%) for 2D touch, while 3D selection had M=11.3% (SD=14.1%).

The results show that the error rate for heights differs significantly ($F(1.848,22.172)$=17.186, p<.001, η_p^2=.589). Post-hoc tests revealed that the error rate was significantly increased when objects were displayed with a height of 20cm (p<.05)

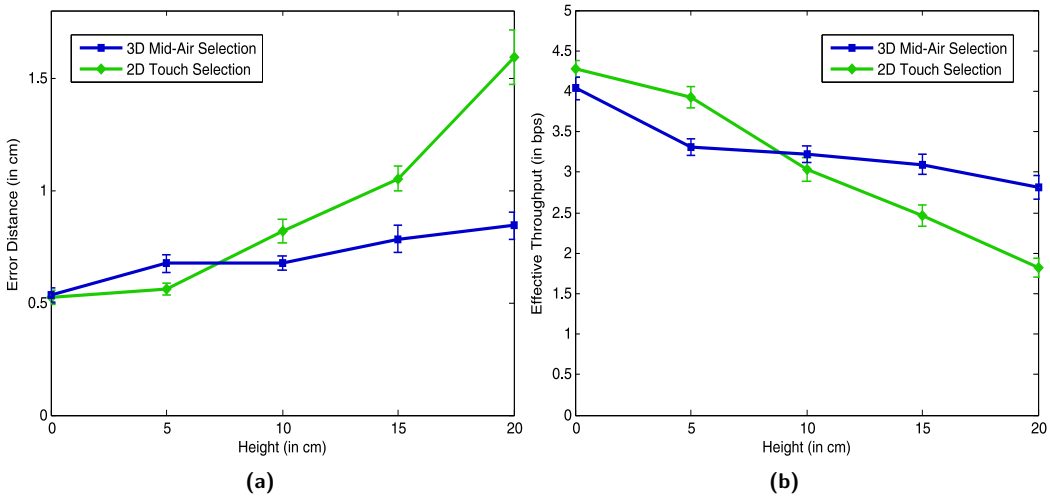


Figure 2.2: Results for Fitts' Law trials with target object height on the horizontal axis and pooled for (a) error distance and (b) effective throughput (combining errors and movement time) on the vertical axis. The error bars show the standard error.

in comparison to 0cm. As expected, we found a significant main effect of the ID on error rate ($F(3, 36)$=15.359, p<.001, η_p^2=.561).

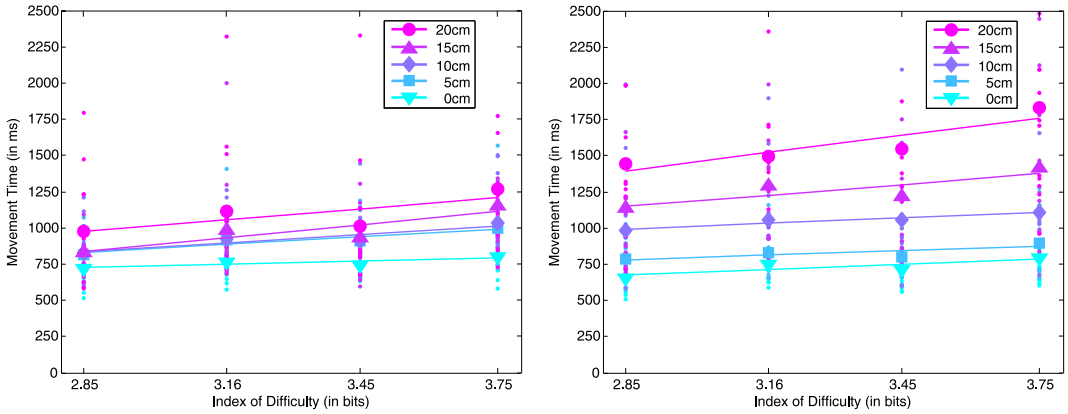
We found no significant two-way interaction effect between technique and height ($F(1.798, 21.570)$=2.685, p>.05, η_p^2=.183).

2.4.3 Error Distance

The results for the error distances, between the center of each sphere and the finger position during selection, are illustrated in Figure 2.2(a). We found a significant main effect of technique ($F(1, 12)$=5.115, p<.05, η_p^2=.299) on the error distance. Participants made significantly larger errors when using 2D touch (M=0.91cm, SD=0.62cm) in comparison to 3D selection (M=0.70cm, SD=0.35cm).

The results show that the error distance for the height differs significantly ($F(1.419, 17.032)$=34.99, p<.001, η_p^2=.745). Post-hoc tests revealed that participants made significantly larger errors when objects were displayed with heights of 15cm (p<.05) or 20cm (p<.001) in comparison to 0cm. As expected, we found a significant main effect of the ID on error distance ($F(1.28, 15.361)$=5.669, p<.03, η_p^2=.321).

We found a significant two-way interaction effect between technique and height ($F(1.427, 17.120)$=11.293, p<.002, η_p^2=.485). Post-hoc tests revealed that participants made significantly larger errors with 2D touch than 3D selection when objects were displayed with a height of 20cm (p<.01). We found no significant difference between the techniques for lower heights.



(a) 3D Mid-Air Selection **(b)** 2D Touch Selection

Figure 2.3: Models for (a) 3D mid-air selection and (b) 2D touch selection: solid lines are regressions of the measured movement time for the five target heights.

2.4.4 Effective Throughput

The results for the effective throughput are shown in Figure 2.2(b). We found no significant main effect of technique ($F(1,12)$=1.658, p>.05, η_p^2=.121) on throughput. The average throughput during the experiment was M=3.11bps (SD=1.29bps) for 2D touch, while 3D selection had M=3.30bps (SD=0.98bps).

The results show that the throughput for heights differs significantly ($F(1.696, 20.358)$=71.995, p<.001, η_p^2=.857). Post-hoc tests revealed that throughput was significantly reduced when objects were displayed with heights of 10cm (p<.05), 15cm (p<.001) or 20cm (p<.001) in comparison to 0cm. As expected, we found a significant main effect of the ID on throughput ($F(3,36)$=8.083, p<.001, η_p^2=.402).

We found a significant two-way interaction effect between technique and height ($F(2.408, 28.898)$=23.979, p<.001, η_p^2=.666). Post-hoc tests revealed that throughput was significantly higher with 3D selection than 2D touch when objects were displayed with a height of 20cm (p<.05). In addition, we found a trend that the throughput was also higher with 3D selection for objects displayed with a height of 15cm (p<.08). In contrast, we found the inverse trend for objects displayed with a height of 5cm (p<.07). Here, throughput for 2D selection was higher. We found no significant difference between the techniques for lower heights.

2.4.5 Modeling

Fitts' Law can also be used as a predictive model, by regressing movement time on the index of difficulty. We performed this analysis for both techniques at the five different heights. The regression lines for movement time are presented in Figure 2.3. The predictive quality of the model (as expressed by χ^2 values) is very high for 2D touch (for heights 0cm to 20cm χ^2=0.18, 0.06, 0.006, 0.04, and 0.037) and for 3D selection (for height 0cm to 20cm χ^2=0.10,

0.06, 0.08, 0.24, and 0.01).

2.4.6 Questionnaires

Also, the results were normally distributed according to a Shapiro-Wilk test at the 5% level. Before and after each of the 2D touch and 3D selection conditions, we asked participants to judge their level of simulator sickness and visual discomfort. Results were analyzed using paired-samples t-tests. For simulator sickness, we found a significant difference between the two conditions ($t(13)$=2.86, p<.02), with an average increase of mean SSQ-scores of 5.61 (SD=16.15) for the 2D touch technique, and 12.16 (SD=12.77) for 3D selections, which may be explained by missing physical support during 3D selections (cf. [BIB+11]). We found no significant difference ($t(13)$=0.16, p>.05) for the asthenopia questionnaire between the two techniques, but we observed a general before-after increase in visual discomfort for both 2D touch (M=0.18, SD=0.37) and 3D selection (M=0.19, SD=0.33). Again, the results do not exceed typical effects in stereoscopic display environments. For the reported sense of feeling present in the virtual scene, we did not observe a significant difference ($t(13)$=0.60, p>.05) for mean SUS-scores for 2D touch (M=3.92, SD=1.15) and 3D selection (M=4.08, SD=1.14). Both scores indicate a high sense of presence. We did not find a significant difference ($t(13)$=0.15, p=.88) between 2D touch (M=2.85, SD=0.43) and 3D selection (M=2.92, SD=0.56) on the mean five general usability criteria scores learnability, efficiency, memorability, errors, and satisfaction. Individual usability scores for 2D touch respectively 3D selection were (M=3.15 & M=3.00) for learnability, (M=3.54 & M=3.29) efficiency, (M=3.08 & M=3.43) memorability, (M=2.31 & M=2.71) errors, and (M=2.46 & M=2.00) for satisfaction. We could not find any significant differences between 2D touch and 3D mid-air selection for these metrics. We found no significant difference ($t(13)$=0.46, p>0.05) between 2D touch (M=10.44, SD=3.27) and 3D selection (M=9.91, SD=3.07) for the NASA TLX mental workload questionnaire scores.

At the end of the experiment, we collected additional subjective preferences in an informal debriefing session. One participant remarked here notably:

"Selecting low objects was much easier on the surface
though it seemed counterintuitive at first!"

This comment was representative for many responses regarding the 2D touch technique. All but one participant preferred touching through 3D objects for objects close to the display surface.

2.5 Discussion

The results from the Fitts' Law experiment reveal distinct characteristics of the 2D touch and 3D mid-air selection techniques, which impact their performance and applicability for interaction with objects displayed stereoscopically at different parallaxes. For 3D objects displayed up to 10cm above the display surface, touching objects in 2D on the surface by touch-

ing "through" the stereoscopic projection outperforms 3D mid-air selection in all considered metrics. Since much research has shown that 3D mid-air selection of virtual objects suffers from low accuracy and precision [BIB+11], e. g., due to visual conflicts, including vergence-accommodation mismatch, diplopia, and distance misperception [CKC+10], it is a promising finding that the reduction of 3D selection tasks to 2D input with the 2D touch technique can improve performance for tabletops with stereoscopically displayed objects. However, while interactions with both techniques are equal for objects at 0cm height, the results also show that the performance of the 2D touch technique decreases drastically for large negative parallax in comparison to 3D mid-air selection. At 20cm height, 2D touch performance is less than half in terms of throughput compared to performance on the screen. 3D mid-air selection performance drops much more slowly, decreasing only by about 30% at 20cm height.

For scenarios with stereoscopic visualization on (multi-)touch surfaces, the findings are still encouraging. They suggest that interactive 3D objects do not have to be constrained at the zero-parallax level, but may deviate up to 10cm before performance with the 2D touch technique is significantly degraded. For such distances, touch input is a good choice. Overall, our results show that it is indeed possible to leverage stereoscopic distance and interposition cues over a considerable range in touch-sensitive tabletop setups for improved spatial understanding of virtual data sets.

In our experiment, we compensated for different viewer heights by raising all participants to a consistent head level. We did this to compensate for the potential confound that a lower viewpoint has a smaller 3D view volume due to (relatively) earlier clipping by the far and near sides of the display. In future commercial systems, we expect that stereoscopic touch tables could be height adjusted to accommodate for the height of each user.

In summary we suggest the following guidelines for the realization of touch interaction in 3D stereoscopic tabletop setups: For tabletop setups using the 2D touch technique, interactive virtual objects (e. g., buttons or other elements of graphical user interfaces) should not be displayed more than 10cm above the interactive display surface. Above that, the disadvantages outperform the benefits and 3D interaction techniques should be used.

2.6 Conclusion

In this chapter, we compared interaction techniques for tabletop setups with stereoscopic display. We analyzed the differences between 3D mid-air selection and a technique based on reducing the 3D selection problem to two dimensions by touching "through" the stereoscopic impression of 3D objects, i. e., a 2D touch on the display. The experimental results show a strong interaction effect between input technique and the stereoscopic parallax of virtual objects for all performance metrics, including movement time, errors, and effective throughput. Our main findings are:

- The 2D touch technique outperforms 3D mid-air selection for objects up to ca. 10cm height above the display surface.

- 3D mid-air selection is a better alternative for higher targets.
- Performance decreases faster for the 2D touch technique than for 3D selection with increasing height of virtual objects.

The results are encouraging for stereoscopic visualization in future touch-sensitive tabletop setups, since no additional tracking technology is needed for objects with small negative parallax. Recent sensing technologies for finger poses above display surfaces (e. g., Leap Motion) will thus realize their benefits mostly only for objects at least about 10cm above the surface.

As a direction for future work, we cannot yet tell if these results hold for portable setups, where the orientation of the touch-sensitive surface can change during interaction. We will pursue this topic to design more compelling user experiences as well as effective user interfaces for touch-sensitive stereoscopic display surfaces.

Part II

Perceptually-Inspired Locomotion in Virtual Environments

Chapter 3

3 Redirecting Walking and Driving for Natural Navigation in Immersive Virtual Environments

Walking is the most natural form of locomotion for humans, and real walking interfaces have demonstrated their benefits for several navigation tasks. With recently proposed redirection techniques it becomes possible to overcome space limitations as imposed by tracking sensors or laboratory setups, and, theoretically, it is now possible to walk through arbitrarily large virtual environments. However, walking as sole locomotion technique has drawbacks, in particular, for long distances, such that even in the real world we tend to support walking with passive or active transportation for longer-distance travel. In this chapter we show that concepts from the field of redirected walking can be applied to movements with transportation devices. We conducted psychophysical experiments to determine perceptual detection thresholds for redirected driving, and set these in relation to results from redirected walking. We show that redirected walking-and-driving approaches can easily be realized in immersive virtual reality laboratories, e. g., with electric wheelchairs, and show that such systems can combine advantages of real walking in confined spaces with benefits of using vehicle-based self-motion for longer-distance travel.

3.1 Introduction

Immersive virtual environments (VEs) are often characterized by head-mounted displays (HMDs) or immersive projection technologies, as well as a tracking system for measuring head position and orientation data. Navigation in such immersive VEs is often performed with interaction devices, such as joysticks or wands, which allow users to initiate self-motion in virtual scenes, but often provide unnatural inputs and feedback from the body about virtual self-motion. Although such setups can provide users with a sense of moving through three-dimensional virtual scenes, these *magical* forms of virtual self-motion [BKH97] have often

revealed degraded performance in wayfinding tasks and mental map buildup when compared to *natural* forms of self-motion from the real world [RL09, SFR+10].

In the real world, we navigate with ease by walking, running, driving etc., but in immersive VEs realistic simulation of these forms of self-motion is difficult to achieve. While moving in the real world, sensory information such as vestibular, proprioceptive, and efferent copy signals as well as visual information create consistent multi-sensory cues that indicate one's own motion, i. e., acceleration, speed and direction of travel. Traveling through immersive virtual environments by means of *real walking* is considered the most basic and intuitive way of moving within the real world, and is an important activity to increase the naturalness of virtual reality (VR)-based interaction [UAW+99]. Keeping such a dynamic ability to navigate through large-scale immersive VEs is of great interest for many 3D applications, such as in urban planning, tourism, or 3D entertainment. However, natural self-motion in immersive VEs imposes significant practical challenges [WCF+05].

An obvious approach for leveraging natural self-motion for immersive VEs is to transfer the user's tracked head movements to changes of the camera in the virtual world by means of isometric mappings. Then, a one meter movement in the real world is mapped to a one meter movement of the virtual camera in the corresponding direction in the VE. This technique has the drawback that a user's movements are restricted by a limited range of the tracking sensors and a rather small workspace in the real world. The size of the virtual world often differs from the size of the tracked laboratory space so that a straightforward implementation of omni-directional and unlimited walking is not possible. Thus, virtual locomotion methods are required that enable locomotion over large distances in the virtual world while remaining within a relatively small workspace in the real world.

As a solution to this challenge, researchers transferred findings from the field of perceptual psychology to address the space limitations of immersive VR setups. Based on perceptual studies showing that vision often dominates proprioception and vestibular sensation when the senses disagree [Ber00, DB78], researchers found that users tended to unwittingly compensate with their body to small inconsistencies in visual stimulation while walking in immersive VEs, which even allows guiding users along paths in the real world that differ from the perceived path in the virtual world [Raz05]. In principle, using this *redirected walking* it becomes possible to explore arbitrarily large virtual scenes using redirection techniques, while the user is guided along circular paths in a considerably smaller tracked interaction space in the laboratory. Recent studies on navigation and spatial disorientation in confined virtual spaces suggest that redirected walking can provide users with similar benefits for navigation as real walking, and a significantly improved performance over virtual flying and other travel techniques [PFW11, SKFB11].

However, although (redirected) walking is a simple navigation technique, it has practical drawbacks, in particular, when traveling over long distances. Even in the real world we support long-distance travel by various forms of traveling devices. Thus, in this chapter we propose supporting natural movements in immersive VEs by moving with traveling devices in the real world (e. g., electric wheelchairs or scooters). Although such devices can make it

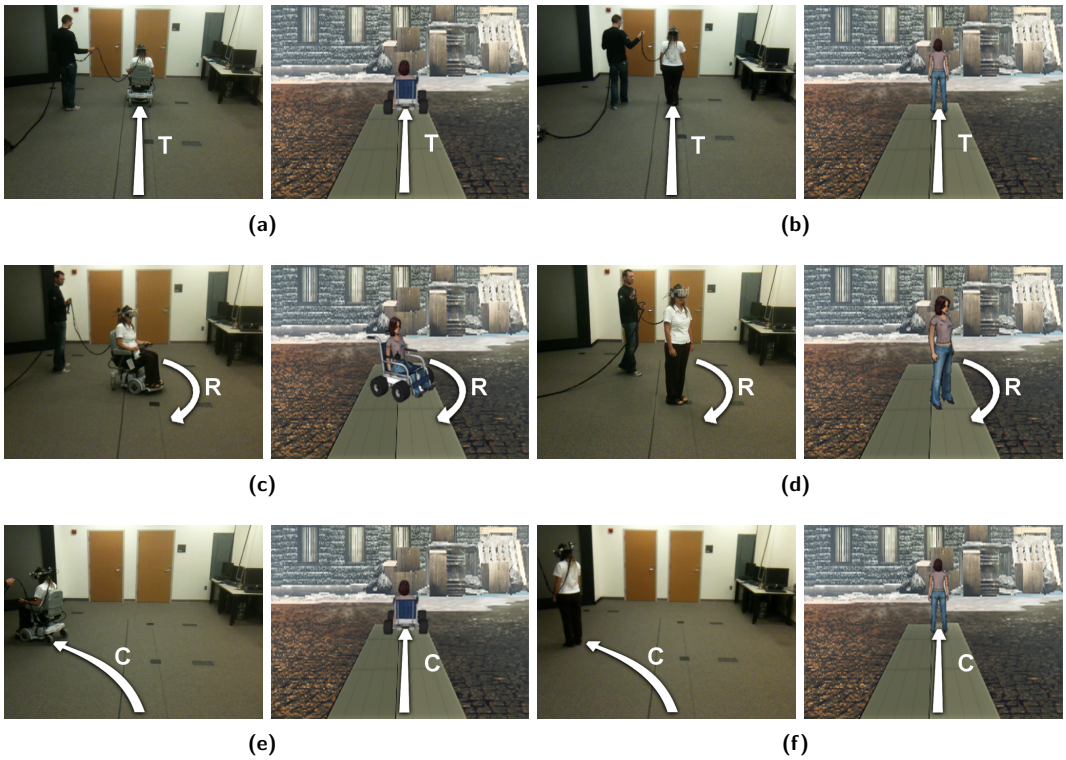


Figure 3.1: Redirected walking-and-driving in immersive virtual environments: (a), (c), (e) user steering an electric wheelchair with a head-mounted display in the virtual reality laboratory, and (b), (d), (f) real walking counterparts. The renderings illustrate virtual representations of translations (T), rotations (R) and physical curvatures (C) [SBJ+10].

more comfortable to travel long distances in VEs, while supporting natural vestibular and proprioceptive feedback, using traveling devices that move in the real world imposes the same problems in terms of space restrictions as real walking. Therefore, concepts similar to redirected walking may be applied to redirect a user's path of travel using such devices. However, since users receive different self-motion cues from the real and virtual world during walking and driving it has to be carefully analyzed whether or not and to what extent redirection techniques can be applied when users steer such traveling devices.

In this chapter we propose, evaluate and discuss *redirected walking-and-driving*, which allows users of immersive VEs to cover long distances in realistic virtual scenes with near-natural vestibular and proprioceptive feedback by steering a traveling device, while retaining the ability to switch to walking depending on the navigation requirements, similar to the real world. In particular, we show that redirected driving can easily be incorporated in head-tracked immersive virtual reality laboratories by adapting an electric wheelchair for virtual traveling.

The remainder of this chapter is structured as follows. Section 3.2 provides an overview of virtual self-motion. In Section 3.3 we present redirected driving in a head-tracked VR laboratory. In Section 3.4 we describe psychophysical experiments that we conducted to

determine perceptual differences in the detectability of manipulations of translations and rotations when walking or driving in a virtual scene. Section 3.5 concludes the chapter.

3.2 Background

Moving through a virtual scene is one of the most essential interaction tasks in virtual reality environments, for which various different technologies and techniques have been introduced. Virtual self-motion approaches can be divided in *locomotion* and *traveling* user interfaces.

Locomotion and Traveling

Defined as active self-propulsion, locomotion encompasses repetitive motions of legs and body during walking, but also propulsion of human-powered vehicles like bicycles, scooters, skaters, or manual wheelchairs [Hol02]. In particular, the key characteristic of locomotion that distinguishes it from passive motion is that proprioceptive and kinesthetic information from the body while moving can be integrated with visual self-motion cues by the perceptual system. A significant body of work has shown the benefits of proprioceptive cues of physical motion in spatial tasks [CGBL98, RL09], with some disagreement about whether the motion needs to be walking [RL09] or whether simple physical rotation would suffice [RSA+06]. Results imply that perception of virtual geometry, motions and distances may be enhanced by the ability to locomote [Hol02]. Moreover, the features of energy expenditure and sensorimotor integration are hypothesized to yield an increased sense of presence in immersive VEs [SUS95]. Typical problems of locomotion interfaces are user exertion when moving over long distances, and the limited physical space when transferring actual movements of a user from a real-world laboratory to a potentially infinite VE.

Traveling user interfaces encompass approaches that are not based on repetitive limb or body motions for initiating or controlling movements. Examples are virtual steering techniques which combine head orientation tracking with hand-based input, e. g., with wands or joysticks, to initiate translations of the user's virtual viewpoint. Since users receive conflicting sensory information caused by visually indicated motions that are not matched by proprioceptive and vestibular cues from their body, such approaches may limit the user's sense of feeling present in a VE [SUS95]. To provide a cognitive grounding for virtual traveling techniques, and to provide physical self-motion cues when traveling in a virtual scene, motion simulators can be used. Motion simulators consist of a mockup of a real-world vehicle, such as a car or aircraft, which may be steered by the user, while receiving visual feedback about the motions, as well as vestibular and proprioceptive feedback from a motion platform [YJN+96]. Motion platforms used in simulators represent a mature technology area that is not addressed in this chapter. In contrast to simulating movements in the real world with motion simulators, we propose using vehicles that actually move in the physical world, and are steered by the user. Examples of such motion devices include electric wheelchairs, scooters, roller skates, and bicycles. With such devices, users receive consistent multisensory cues about self-motions in the virtual and real world, including visual, vestibular and inertial feedback, while limiting

user exertion when traveling over long distances. However, the same limitations apply to users moving with a vehicle through the laboratory space as for users walking in the limited workspace provided by tracking sensors.

Redirection Techniques

Different approaches to redirect a user in immersive VEs have been proposed. An obvious approach is to scale translational movements, for example, to cover a virtual distance that is larger than the distance traveled in the physical space. With most redirection techniques, however, the virtual world is slowly rotated around the center of a standing or walking user until the user is oriented in such a way that no physical obstacles block the path of travel [GNRH05, KBMF05, PFW11, Raz05]. For instance, if the user wants to walk straight ahead for a long distance in the virtual world, small rotations of the camera redirect the user to walk unconsciously on an arc in the opposite direction in the real world. When redirecting a user, the visual sensation is consistent with motion in the VE, but vestibular and proprioceptive sensations reflect motion in the physical world. If the induced manipulations are small enough, the user has the impression of being able to walk in the virtual world in any direction without restrictions. A vast body of research has been undertaken in order to identify thresholds that indicate the tolerable amount of deviation between sensations from the virtual world and physical world while the user is walking. In this context, Steinicke et al. [SBJ+10] conducted a series of psychophysical experiments to identify detection thresholds for redirected walking gains. Therefore, they compared manipulations with a range of gains, which have been applied to rotations, translations, and curved paths, while participants had to discriminate between virtual and real motions (see Figure 3.1).

In this chapter, we show that redirection techniques can be applied not only for locomotion but also for traveling, with a user steering a physical vehicle that actually moves through the laboratory space.

3.3 Redirected Driving

Redirected driving for moving vehicles in a limited VR laboratory space can be implemented with the same approaches as used to enable redirected walking. In particular, since redirection is a software-based process that makes use of perceptual limitations of humans with the goal to subconsciously affect a user's movements in the real world compared to virtual movements, many of the controllers developed for redirected walking can be directly applied for manipulating a user's movements when steering a vehicle [Raz05, SBJ+10].

Redirection of walking and driving differs in terms of different cues provided to users about movements in the real and virtual world. For instance, walking users may adapt to manipulations of the visual stimuli, e. g., optic flow movement velocity and direction cues [BSW11, GL99, LBv99], in the VE by adaptation of muscles used for walking straight, or turning [Raz05]. Adaptation of traveling direction and velocity of users driving with a vehicle may require different muscle groups, which are integrated with different couplings

and levels of conscious access to motor control information in human perception and action processes [EB04].

3.3.1 Combining Walking and Driving

Redirected walking-and-driving can be implemented with the same software-based techniques, and even in the same VR setup. In particular, provided the user's head position and orientation can be tracked in the VR laboratory, basic mappings from real head movements to virtual camera motions are independent of whether the user is using a vehicle in the real world to travel, or whether the user is walking (see Figure 3.1). As a result, for basic setups, no additional hardware is required to enable combined walking-and-driving. However, if users are immersed in a VE using an HMD, the virtual scene is displayed exclusively to the user, while blocking visual information about the vehicle from the real world, i. e., it may be required to track the position and orientation of the vehicle in the laboratory to display a registered virtual counterpart to the user when required (see Figure 3.1). Combining walking-and-driving in VR environments provides users with advantages of walking in focus regions, as well as an intuitive means of traveling over longer distances.

3.3.2 Redirecting Self-Motion

In head-tracked immersive VR environments, user movements are typically mapped isometrically to virtual camera motions. For each frame the change in position and orientation measured by the tracking system is used to update the virtual camera state for rendering the new image that is presented to the user. The new camera state can be computed from the previous state defined by tuples consisting of the position $pos_n \in \mathbb{R}^3$ and orientation $(yaw_n, pitch_n, roll_n) \in \mathbb{R}^3$ at frame $n \in \mathbb{N}$ in the scene with the tracked change in position $\Delta pos \in \mathbb{R}^3$ and orientation $(\Delta yaw, \Delta pitch, \Delta roll) \in \mathbb{R}^3$. In the general case, we can describe a scaled mapping from real to virtual motions as follows:

$$pos_{n+1} \;=\; pos_n \;+\; g_T \cdot \Delta pos,$$

$$\begin{cases} yaw_{n+1} &= yaw_n &+ &g_{R[yaw]} \cdot \Delta yaw, \\ pitch_{n+1} &= pitch_n &+ &g_{R[pitch]} \cdot \Delta pitch, \\ roll_{n+1} &= roll_n &+ &g_{R[roll]} \cdot \Delta roll, \end{cases}$$

with translation gains $g_T \in \mathbb{R}$ and rotation gains $(g_{R[yaw]}, g_{R[pitch]}, g_{R[roll]}) \in \mathbb{R}^3$ [SBJ+10]. As discussed by Interrante et al. [IRA07], translation gains may be selectively applied to translations in the main walk direction.

Camera rotations can also be introduced relative to head translations. In particular, if the virtual scene is slowly rotated around the user's viewpoint while the user is walking straight, the user adapts to the virtual rotation by rotating in the real world. Such physical path bending manipulations are specified as rotation angles per walking distance [Raz05], or circular path radii in the real world [SBJ+10], with curvature gains defined as $g_C = \frac{1}{r}$, for

radius $r \in \mathbb{R}^+$, and $g_C = 0$ for $r = \infty$.

High-level redirected walking controllers usually incorporate one or more of these techniques to manipulate a user's walking direction or travel distance in the real world relative to the VE [BSH09, NHS04, PFW11, Raz05]. To support this process, researchers determined the amount of manipulation that users are unaware of for each of these techniques in the field of redirected walking [SBJ+10], such that controllers could try to determine the least noticeable combination of the manipulations in the context of the user's current state in the real laboratory and virtual scene.

3.3.3 Hypothesis

Since previous research on detectability of redirection manipulations has focused mainly on users walking with a HMD in a laboratory environment, it is still largely unknown how the human perceptual system integrates differences in self-motion information from the real and virtual world when steering a traveling device, such as when seated in an electric wheelchair. However, diverging findings in the fields of redirected walking and motion platforms suggest differences in discrimination performance and detectability of manipulations [HAH02, Raz05, vR01, YB85]. In particular, it is not well-understood how the sophisticated perceptual processes involved in posture stability during natural walking contribute to self-motion perception, e. g., when coordinating over 50 muscles or muscle groups to maintain the body in a repetitive forward progression [BR06, LAPD06], in comparison to seated traveling, which limits the number of available self-motion cues. We hypothesize that

H1) with an electric wheelchair participants will be less accurate at detecting discrepancies of real and virtual self-motions,

which is suggested by a reduced number of real-world self-motion cues when seated compared to when walking, and suggests advantages of redirected driving over redirected walking for longer-distance traveling in a large virtual scene.

3.4 Psychophysical Evaluation of Redirected Driving

In this section, we evaluate redirected driving in three experiments, which we conducted to analyze detectability of manipulations of translations and rotations when driving an electric wheelchair, and compare the results to redirected walking based on an implementation of the same redirection techniques. Therefore, we analyzed participants' estimation of physical movements compared to simulated virtual motions while varying the parameters of the redirection techniques, which provides information on how the traveling technique affects the just noticeable difference between physical and virtual motions, as well as practical thresholds that can be applied in redirection controllers.

Figure 3.2: Visual stimulus generated with Crytek's CryEngine 3 in the walking and driving trials.

3.4.1 Experiment Design

We performed the experiments in a 11m×9.5m darkened laboratory room. The participants wore an nVisor SX60 HMD (1280×1024@60Hz, 60° diagonal field of view) for the stimulus presentation. We used a 3rdTech Hiball 3100 Wide Area Tracker to track the position and orientation of an optical sensor that we fixed on the HMD. The Hiball tracker provided sub-millimeter precision and accuracy of position data, as well as <0.01° angular precision and <0.02° angular accuracy of orientation data at an update rate between 1000−2000Hz during the experiments. For visual display, system control and logging, we used an Intel computer with Core i7 processors, 6GB of main memory and Nvidia Quadro FX 1500 graphics card.

For the trials with electric wheelchair we used a Hoveround MPV 5 Power Wheelchair, which provides variable speed settings of up to 8 km/h, a 22.7" turning radius (adjustable by participants to zero around the head position), and joystick control (see Figure 3.1). We used settings of approximately 2.34 km/h top speed, 0.13 m/s^2 acceleration and -0.83 m/s^2 deceleration for linear movements, as well as 44 deg/s top speed, 45 deg/s^2 acceleration and -90 deg/s^2 deceleration for angular movements. During the experiment, ambient city noise was presented to the participants over the headphones in the nVisor SX60 HMD to reduce auditive orientation cues from the laboratory.

In order to focus participants on the tasks, no communication between experimenter and participant was performed during the experiments. All instructions were displayed on slides in the VE, and participants judged their perceived motions via button presses on a Nintendo Wii Remote controller. The visual stimulus consisted of a virtual city environment rendered with Crytek's CryEngine 3 (see Figure 3.2).

Participants

8 male and 4 female (ages 19−51, $M = 26.9$) participants completed the experiment. All participants were undergraduate or graduate students or members of the department of computer science. All had normal or corrected to normal vision. No participant had a disorder of balance. 1 participant had no experience with 3D games, 5 had some, and 6 had much experience. 5 of the participants had experience with walking in an HMD environment. All participants were naïve to the experimental conditions. All participants had experience with steering the electric wheelchair using its joystick controller due to a 3-minute familiarization phase before the experiment. The total time per participant including pre-questionnaire, instructions, training, experiments, breaks, and debriefing was 1.5 hours, of which participants spent approximately 1 hour wearing the HMD. Participants were allowed to take breaks at any time.

Methods

We used a within-subjects design, and the method of constant stimuli in a *two-alternative forced-choice* (2AFC) task. In the method of constant stimuli, the applied gains are not related from one trial to the next but presented randomly and uniformly distributed. The participant chooses between one of two possible responses, e. g., "Was the virtual movement *smaller* or *larger* than the physical movement?"; responses like "I can't tell." are not allowed. When the participant cannot detect the signal, the participant is forced to guess, and will be correct on average in 50% of the trials [SBJ+10].

The gains at which the participant responds "smaller" in half of the trials is taken as the *point of subjective equality* (PSE), at which the participant judges the virtual motion to match the physical movement. As the gain decreases or increases from this point the ability of the participant to detect the difference between physical and virtual motion increases, resulting in a psychometric curve for the discrimination performance. The discrimination performance pooled over all participants is represented with a fitted psychometric function, for which we used the common Weibull function for 2AFCs [FHW11, Kle01]. The PSEs give indications about how to parameterize the redirection technique such that virtual motions appear natural to users, while manipulations with values close to the PSEs will often go unnoticed by users. Typically, the points are taken as thresholds, at which the psychometric curve reaches the middle between the chance level and 100% correct detections (cf. Steinicke et al. [SBJ+10]). We define the detection threshold (DT) for gains smaller than the PSE to be the point at which the participant has 75% probability of choosing the "smaller" response and the detection threshold for gains larger than the PSE to be the point at which the participant chooses the "smaller" response in only 25% of the trials (since the correct response was then chosen in 75% of the trials). The detection thresholds indicate which practical range of manipulations can be applied in redirection controllers.

We measured the participants' sense of presence with the SUS questionnaire [UCAS99], and simulator sickness with the Kennedy-Lane SSQ before and after each experiment. The

wheelchair and walking trials were conducted in separate blocks, of which the order was randomized between participants. The order of the experiments in each condition was randomized for each participant.

3.4.2 Experiment E1: Rotation Discrimination

We analyzed the impact of the physical locomotion methods walking and driving with independent variable $g_{R[yaw]}$ (cf. Section 3.3) on discrimination of real and virtual rotations.

Materials

We instructed the participants to turn their head and body around in the VE until the scene changed. The rotation angle in the real world was randomized between 67.5° and 112.5°, with an average rotation angle of 90°. The virtual rotation angle was scaled with rotation gains $g_{R[yaw]}$ between 0.4 and 1.6 in steps of 0.2. We randomized the independent variables over all trials and tested each 4 times. In total, each participant performed 28 rotations in-place when standing, as well as when seated in the wheelchair. We instructed participants to alternate clockwise and counterclockwise rotations, which were counterbalanced for all gains.

For each trial, after a participant performed the rotation in the VE, the participant had to decide whether the simulated virtual rotation was smaller (down button) or larger (up button) than the physical rotation with the Wii Remote controller. The next trial started immediately after the participant judged the previous motion. The procedure was identical for rotations when standing and with the wheelchair. To control rotations with the wheelchair, participants used the joystick to initiate a rotation either to the left or right, corresponding to counterclockwise and clockwise rotations, respectively. The physical rotation speed with the wheelchair of 44 deg/s approximated the mean turning speed of 41 deg/s while standing.

Results

Figure 3.3 shows the pooled results for the gains $g_{R[yaw]} \in \{0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6\}$ on the x-axis with the standard error over all participants. The y-axis shows the probability for estimating the virtual rotation as smaller than the real rotation. The black psychometric function shows the results for standing participants, and the gray function for participants rotating with the wheelchair. We observed a chi-square goodness of fit of the psychometric function of $\chi^2 = 0.6990$ for standing and $\chi^2 = 0.3822$ for the wheelchair. We did not observe a difference in responses for clockwise and counterclockwise rotations, as well as for the different physical rotation angles, and pooled the data. From the psychometric functions we determined PSEs at $g_{R[yaw]} = 0.9544$ for standing, and $g_{R[yaw]} = 1.0111$ for the wheelchair condition. A practically applicable range of manipulations with rotation gains is given as the interval between the lower and upper detection thresholds, which we determined from the psychometric functions as $g_{R[yaw]} \in [0.6810, 1.2594]$ for standing participants, and $g_{R[yaw]} \in [0.7719, 1.2620]$ for the electric wheelchair.

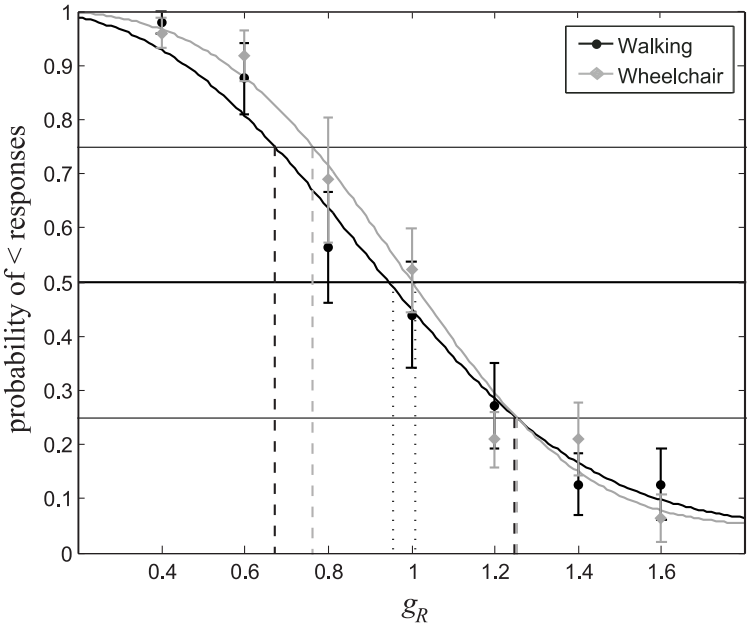


Figure 3.3: Pooled results of in-place rotations while standing (black function), and seated in the wheelchair (gray function). The x-axis shows the applied parameter $g_{R[yaw]}$. The y-axis shows the probability of estimating the virtual rotation as smaller than the real rotation.

Discussion

The results show a significant impact of parameter $g_{R[yaw]}$ on responses. For participants standing and rotating in-place, the results approximate results found by Steinicke et al. [SBJ+10]. In particular, the participants' responses indicate a slight underestimation of rotations in the VE of approximately 4.56%, while Steinicke et al. found an underestimation of approximately 4%. For participants rotating while seated in the electric wheelchair, the results indicate no bias towards over- or underestimation of virtual rotations. The detection thresholds in the standing condition define a possible manipulation range of rotations that can cause a real rotation to deviate from a fixed virtual rotation between −20.60% and +46.84% (see Section 3.3). In the wheelchair condition, real rotations can deviate between −20.76% and +29.55%.

The results are interesting, in particular, considering the duality of movement cues provided from the real and virtual world during rotations. From the VE, participants primarily received visual cues, e. g., optic flow [LBv99], as well as limited cues from ambient auditive sources of city noise. From the real world, participants in both conditions received vestibular feedback about their angular head motion. Differences between the two conditions mainly show for proprioceptive feedback. While standing participants received proprioceptive cues about the motion of their body, such cues were limited in the wheelchair condition. Moreover, participants had to initiate rotations by pushing the joystick of the wheelchair all the way to the left or right to initiate counterclockwise or clockwise rotations, respectively. This suggests that participants got the same proprioceptive cues about the state of their hand

in all trials, independent of the virtual motion. It remains unclear if the differences in the responses were caused by cue integration processes [EB04], or cognitive effects of the traveling technique [AKLG04].

3.4.3 Experiment E2: Translation Discrimination

We analyzed the impact of the physical locomotion methods walking and driving with independent variable g_T (cf. Section 3.3) on discrimination of real and virtual travel distances.

Materials

We instructed the participants to walk or drive forward along a displayed straight path in the virtual scene until the scene changed (see Figure 3.2). The travel distance in the real world was randomized between 2.5−3.5m. The virtual travel distance was scaled with translation gains g_T between 0.4 and 1.6 in steps of 0.2. As proposed by Interrante et al. [IRA07], we applied translation gains only to translations in the main walk direction, i. e., we did not scale lateral translations and head bobbing. We randomized the independent variables over all trials and tested each 4 times. In total, each participant performed 28 translation trials when walking, as well as when driving with the wheelchair.

For each trial, after a participant performed the translation in the VE, the participant had to decide whether the simulated virtual translation was smaller (down button) or larger (up button) than the physical translation with the Wii Remote controller. After the participant judged the previous motion, participants were guided to the start position in the real world for the next trial via two 2D markers on a uniform background. The next trial started immediately once the participant assumed the start position and orientation for the next trial. The procedure was identical for translations when walking, and translations in the wheelchair. To control translations with the wheelchair during the trials, participants used the joystick to initiate a straight translation in forward traveling direction. The physical traveling speed with the wheelchair of 2.34 km/h approximated the mean walking speed of 2.7 km/h.

Results

Figure 3.4 shows the pooled results for the gains $g_T \in \{0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6\}$ on the x-axis with the standard error over all participants. The y-axis shows the probability for estimating the virtual translation as smaller than the real translation. The black psychometric function shows the results for walking participants, and the gray function for participants traveling with the wheelchair. We observed a chi-square goodness of fit of the psychometric function of $\chi^2 = 0.5372$ for walking, and $\chi^2 = 0.0258$ for the wheelchair. We did not observe a difference in responses for the different physical traveling distances and pooled the data. From the psychometric functions, we determined PSEs at $g_T = 1.0824$ for walking, and $g_T = 1.1508$ for driving with the wheelchair. A practically applicable range of manipulations with translation gains is given as the interval between the lower and upper detection thresholds,

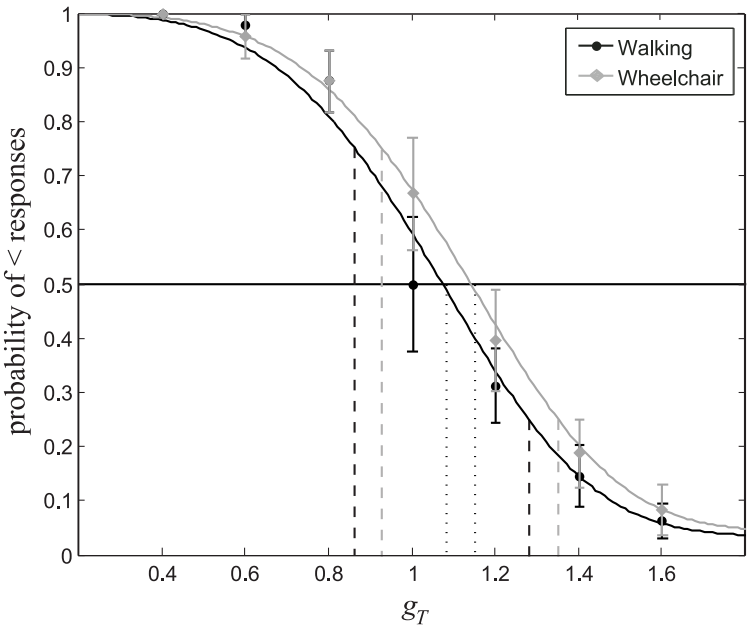


Figure 3.4: Pooled results of translations while walking (black function), and driving in the wheelchair (gray function). The x-axis shows the applied parameter g_T. The y-axis shows the probability of estimating the virtual translation as smaller than the real translation.

which we determined from the psychometric functions as $g_T \in [0.8724, 1.2896]$ for walking, and $g_T \in [0.9378, 1.3607]$ for driving with the electric wheelchair.

Discussion

The results show a significant impact of parameter g_T on responses. For walking participants, the results approximate results found by Steinicke et al. [SBJ+10]. In particular, the participants' responses indicate a slight overestimation of translations in the VE of approximately 8.24%, while Steinicke et al. found an overestimation of approximately 7%. For participants driving with the electric wheelchair, the results indicate a stronger bias towards overestimation of virtual rotations of approximately 15.08%. The detection thresholds in the walking condition define a possible manipulation range of translations that can cause a real translation to deviate from a fixed virtual translation between -22.46% and $+14.62\%$ (see Section 3.3). In the wheelchair condition, real translations can deviate between -26.51% and $+6.63\%$.

Different cues provided from the real and virtual world during walking and driving may have caused the differences. Participants received visual cues about translations in the VE, as well as limited cues from ambient city noise. Participants in both conditions received vestibular feedback about their linear head motion in the real world. Similar to experiment E1 (see Section 3.4.2), differences between the two conditions mainly show for proprioceptive feedback during translations. While walking participants received proprioceptive cues about the motion of their body, such cues were limited in the wheelchair condition. Participants driving the wheelchair had to initiate translations by pushing the joystick all the way forward

to initiate linear movements. This suggests that participants got the same proprioceptive cues about the state of their hand in all trials, independent of virtual translations.

3.4.4 Experiment E3: Curvature Discrimination

We analyzed the impact of the physical locomotion methods walking and driving on discrimination of real and virtual motion directions.

Materials

The procedure was similar to Experiment E2 (see Section 3.4.3). We instructed the participants to walk or drive forward along a displayed straight path in the virtual scene until the scene changed. While a participant was moving forward along the virtual path, we slowly rotated the virtual camera around the participant's virtual position (cf. Section 3.3), which resulted in the participant adapting to the rotational motion in the VE by moving forward on a circular path in the real world. The travel distance in the virtual scene, i. e., the arc length of the circular path in the real world, was 3m in all trials. The virtual camera rotation was adapted to different circle radii in the real world. We mapped participants' virtual translations on circular paths in the real world with radii of 5m, 10m, 20m and 30m. The movement direction in the real world was randomized and counterbalanced for clockwise and counterclockwise progression along the circular paths. We randomized the circle radii over all trials and tested each 4 times to the left and right. In total, each participant performed 32 curvature trials when walking, as well as when driving with the wheelchair.

For each trial, after a participant performed the movement in the VE, the participant had to decide whether the participant moved on a circular path to the left (left button) or right (right button) in the real world with the Wii Remote controller. After the participant judged the previous motion, participants were guided to the start position in the real world for the next trial via two 2D markers on a uniform background. The next trial started immediately once the participant assumed the start position and orientation. The procedure was identical for walking, and driving with the wheelchair. To control movements with the wheelchair during the trials, participants used the joystick to move along the manipulated direction of travel. The physical traveling speed with the wheelchair approximated the mean walking speed of 2.7 km/h.

For the experiment, we slightly modified the joystick control of the wheelchair. An evaluation of the joystick controller showed that the 360° motion range of the joystick assumed a slightly elliptical shape, which provided a haptic indication of when the joystick was pushed straight forward, or slightly to the left or right. To reduce the haptic cues that participants received from the joystick about straightforward motions of the wheelchair in the real world, we placed a circular frame around the joystick handle.

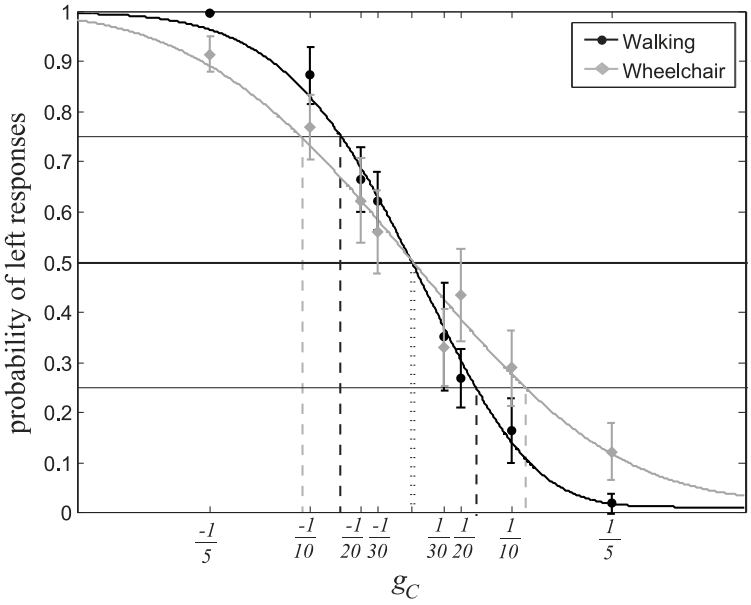


Figure 3.5: Pooled results of curvatures while walking (black function), and driving in the wheelchair (gray function). The x-axis shows the g_C gains defined as inverse circular path radius in the real world, with negative gains referring to paths bent to the left, and positive gains to rightward paths. The y-axis shows the probability of estimating the physical movement path as bent to the left.

Results

Figure 3.5 shows the pooled results for the curvature radii 5m, 10m, 20m, and 30m as curvature angle per travel distance $g_C \in \{-\frac{1}{5}, -\frac{1}{10}, -\frac{1}{20}, -\frac{1}{30}, \frac{1}{30}, \frac{1}{20}, \frac{1}{10}, \frac{1}{5}\}$ on the x-axis, with negative values referring to physical paths bent to the left, positive values referring to paths bent to the right, and the standard error over all participants. The y-axis shows the probability for estimating the real movement as bent to the left while walking straight in the VE. The black psychometric function shows the results for walking participants, and the gray function for participants traveling with the wheelchair. We observed a chi-square goodness of fit of the psychometric function of $\chi^2 = 0.2227$ for walking, and $\chi^2 = 0.2191$ for the wheelchair. From the psychometric functions we determined PSEs at a radius of 461.7m for walking, and a radius of 246.6m for driving with the wheelchair, i. e., the responses indicate that participants on average judged straight movements in the real world as straight. We did not observe a significant difference between curvatures to the left and right. A practically applicable range of manipulations is given by the detection thresholds, which we determined from the psychometric functions as radii larger or equal to 14.92m for walking, and 8.97m for driving with the electric wheelchair.

Discussion

The results show a significant impact of the circular path radius on responses. The results show that the walking participants were less accurate at detecting manipulations of physi-

cal walking directions than found in a similar experiment by Steinicke et al. [SBJ+10]. In particular, our data suggests that the 75% detection threshold may be reached at a circular path radius of 14.92m, whereas the previous results suggested a radius of 22.03m. The differences may be due to the different VR setups, or participant groups, which have been suggested as potential factors [Raz05]. For driving participants the results show that the detection threshold is reached at a radius of 8.97m, which is surprisingly small compared to the walking condition, suggesting that participants can be reoriented more when driving with the wheelchair than when walking.

The difference between walking and driving may be caused by different cues provided while moving, and may be influenced by the active locomotor control. In particular, participants received audiovisual feedback about a straightforward motion in the VE in all trials, as well as angular motion cues about the path curvature, when applied scene rotations became consciously detectable for the participant. From the real world, participants in both conditions received vestibular and proprioceptive feedback about the curvature radius of the movement path in the real world, which has been found in previous studies to be linked to human locomotor control when walking, i. e., the locomotor state of the body may be adapted according to self-motion percepts [MTCR+07, Raz05]. Conversely, the movement direction in the wheelchair condition was controlled by participants using the joystick. While driving, participants pushed the joystick all the way forward, and adjusted the joystick to the left or right for virtual straight driving, i. e., participants received different feedback from the state of their hand depending on the curvature in the real world. As a result, in contrast to experiments E1 and E2, in this experiment the proprioceptive cues from the hand were not independent of the manipulation. However, visual information of the hand and joystick were blocked due to the HMD, such that due to the modified controller there were no direct cues indicating which direction of the joystick corresponds to straightforward motion (see Section 3.4.4).

3.4.5 General Discussion

The results of the three experiments suggest that detectability of virtual motion manipulations depends on the physical locomotion method. In particular, participants driving the electric wheelchair could be redirected more in experiment E3 than participants walking in the laboratory, which suggests that hypothesis H1 (see Section 3.3) holds for such manipulations. However, we did not observe comparatively larger detection thresholds for manipulations in experiments E1 and E2. The results indicate that discrimination performance of real and virtual rotations and translations is similar for participants receiving different cues while walking and driving in the wheelchair.

Moreover, the results indicate differences in the PSEs for rotations and translations between the two conditions. While rotations with the wheelchair showed no significant bias for over- or underestimation of virtual motions, in-place rotations of standing participants showed a slight overestimation of virtual rotations, which is in line with results of previous

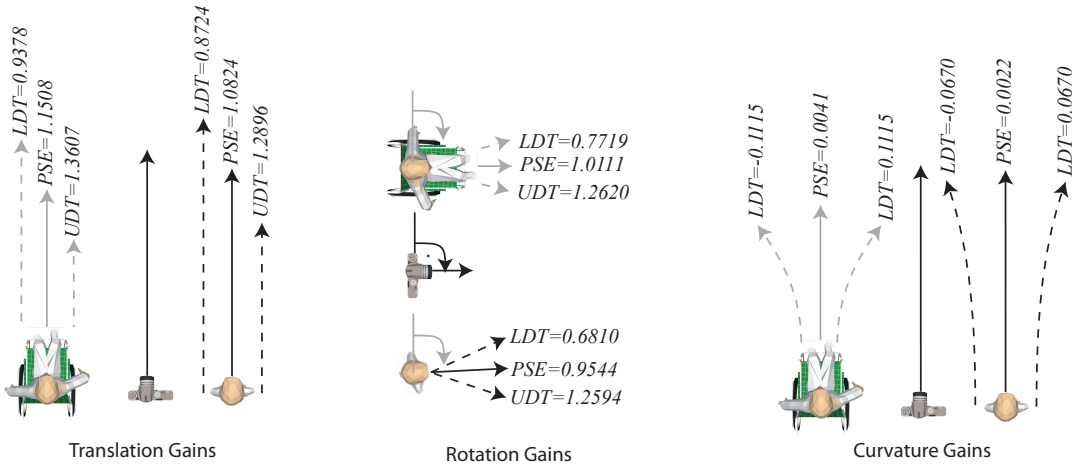


Figure 3.6: Illustration of lower and upper detection thresholds (i. e., LDT and UDT) and PSEs for the two physical locomotion means for (left) translations, (center) rotations and (right) curvatures. The wheelchair and walking illustrations indicate relative differences in physical motions to a virtual camera translation or rotation.

studies that evaluated real walking interfaces [SBJ+10]. Comparing real and virtual travel distances, the results showed an overestimation of virtual translations in both walking and driving conditions, while participants driving the wheelchair judged virtual traveling to be comparatively smaller than in the real walking condition. The results indicate that virtual translations may have to be upscaled in the wheelchair condition to provide participants with a visual stimulus of self-motion that they estimate as equal to their physical movements in the real world.

From the debriefing sessions, we gathered informal comments on the experiments. Multiple participants reported that they had difficulties estimating their actual motions in the real world when driving the wheelchair, which indicates that fewer reliable cues from physical movements could be used for the discrimination task. Compared to that, some participants commented that the wheelchair condition induced a different cognitive context when traveling in the VE, with the impression of having to go faster with the vehicle.

From the results of the Kennedy-Lane pre- and post-questionnaires we determined an average increase of simulator sickness of 6.46 (SD = 2.72) in the walking condition, and 5.78 (SD = 2.07) in the wheelchair condition. We performed a one-way repeated measures analysis of variance (ANOVA), testing the within-subjects effects of the locomotion technique, i. e., walking and driving, on the SSQ scores. We could not find any significant main effects for the SSQ scores ($F(1,22)=1.299$, $p>0.05$), i. e., we did not find any evidence that driving with the wheelchair contributes to or reduces simulator sickness symptoms. The SSQ scores approximate results of previously conducted studies involving walking in HMD environments over the time of the experiment.

The results of the presence questionnaire showed SUS mean scores of 4.82 (SD = 0.91) for walking, and 4.71 (SD = 1.01) for driving with the wheelchair. Again, we could not find a significant difference between walking and driving ($F(1,22)=0.080$, $p>0.05$), which supports

the notion that the wheelchair traveling interface can induce a similar sense of presence in participants as walking.

Furthermore, after the walking and wheelchair conditions, we asked participants to judge their fear of colliding with a wall or physical obstacle in the laboratory during the experiment. The participants judged their level of fear on a 5-point Likert-scale, with 0 corresponding to no fear, and 4 corresponding to a high level of fear. The results show an average level of fear of 1.17 (SD = 1.53) for walking, and 1.33 (SD = 1.56) for the wheelchair interface, which shows that participants felt quite safe in both conditions of the experiment. We could not find a significant difference of the reported level of fear between the conditions ($F(1,22)=0.070$, $p > 0.05$). On similar 5-point Likert-scales all participants judged that they received negligible audiovisual position or orientation cues from the real world during the trials in both conditions.

3.5 Conclusion

In this chapter we have proposed, discussed and evaluated *redirected walking-and-driving*, which denotes the locomotion user interface approach to combine redirected walking in focus regions with redirected driving to cover longer distances in virtual scenes. Both approaches provide users with near-natural vestibular and proprioceptive feedback from actually moving in the real world. The user interface can easily be implemented in head-tracked VR laboratories without extensive hard- and software requirements.

We have evaluated and compared redirection techniques for walking and driving of an electric wheelchair in psychophysical experiments. The results are promising for developers of VR user interfaces (see Figure 3.6). In particular, the results suggest that participants can be redirected on smaller circles in the laboratory when driving with the wheelchair compared to when walking (see Section 3.4.4), and participants have a tendency to regard upscaled virtual travel distances as matching smaller physical distances when driving the wheelchair (see Section 3.4.3). Both results suggest that driving may be better suited for longer-distance travel in immersive VEs than real walking.

It remains an open question how different steering interfaces may affect detectability of manipulations. While joystick control of the electric wheelchair provided no direct cues for estimation of physical rotations and translations as discussed in Sections 3.4.2 and 3.4.3, steering with the joystick interface may have provided additional cues when judging physical path curvatures (cf. Section 3.4.4). In the future, we plan to remove those cues entirely, e. g., by adapting the joystick controller for remote input in the laboratory. Compared to traditional redirected walking, which suffers from the problem that changes of a user's walking path can only be induced indirectly with potential for failure cases, we believe that redirected driving can be implemented without such failure cases, and with less detectable manipulations than for walking. Evaluating joystick control compared to other steering controllers may provide more insight into the reliability of physical cues when using such steering interfaces. Moreover, we will further evaluate perceptual and cognitive effects of combining natural locomotion

techniques for navigation in VEs, with particular focus on disorientation and mental map buildup in unknown virtual scenes, which may benefit from multisensory self-motion cues derived from actually moving in the real world, but may also be affected by integration of manipulated cues in redirected walking or driving environments.

Chapter 4

4 Cognitive Resource Demands of Redirected Walking

Redirected walking allows users to walk through a large-scale immersive virtual environment (IVE) while physically remaining in a reasonably small workspace. Therefore, manipulations are applied to virtual camera motions so that the user's self-motion in the virtual world differs from movements in the real world. Previous work found that the human perceptual system tolerates a certain amount of inconsistency between proprioceptive, vestibular and visual sensation in IVEs, and even compensates for slight discrepancies with recalibrated motor commands. Experiments showed that users are not able to detect an inconsistency if their physical path is bent with a radius of at least 22 meters during virtual straightforward movements. If redirected walking is applied in a smaller workspace, manipulations become noticeable, but users are still able to move through a potentially infinitely large virtual world by walking. For this semi-natural form of locomotion, the question arises if such manipulations impose cognitive demands on the user, which may compete with other tasks in IVEs for finite cognitive resources. In this chapter, we present an experiment in which we analyze the mutual influence between redirected walking and verbal as well as spatial working memory tasks using a dual-tasking method. The results show an influence of redirected walking on verbal as well as spatial working memory tasks, and we also found an effect of cognitive tasks on walking behavior. We discuss the implications and provide guidelines for using redirected walking in virtual reality laboratories.

4.1 Introduction

Virtual reality (VR) technologies are often used in application domains which involve simultaneous spatial tasks and goals competing for the limited cognitive resources of users. In particular, simultaneous navigation, locomotion and interaction with objects in virtual environments (VEs) are essential tasks for many three-dimensional (3D) applications, such as urban planning, training, or entertainment [SVCL13].

Locomotion in Virtual Environments

Traditionally, immersive virtual environments (IVEs) were restricted to visual displays combined with interaction devices, e. g., joystick or wand, for virtual position control. Recently, traveling through IVEs by means of intuitive, multimodal methods of generating self-motion is becoming increasingly important to improve the naturalness of VR-based interaction. In particular, traveling by means of *real walking* has received much attention recently due to results of perceptual and cognitive experiments showing that walking has significant advantages over other forms of traveling in terms of the user's sense of feeling present in the VE [UAW+99], navigational search tasks [RL09], cognitive map buildup [RVB10], and required cognitive resources [MKDO13].

However, while humans navigate with ease by walking in the real world, realistic simulation of natural locomotion is difficult to achieve in IVEs [SVCL13]. During walking in the real world, vestibular, proprioceptive, and efferent copy signals, as well as visual information create consistent multi-sensory cues which indicate one's own motion, i. e., acceleration, speed and direction of travel. Real walking can be implemented in IVEs by mapping a user's tracked head movements to changes of the camera in the virtual world, e. g., by means of a one-to-one mapping. Then, a one-meter movement in the real world is mapped to a one-meter movement of the virtual camera in the corresponding direction in the VE. While this implementation provides near-natural sensory feedback similar to the real world, it has the drawback that the user's movements are restricted by the limited range of the tracking sensors and limitations of the workspace in the real world. The size of the virtual world often differs from the size of the tracked workspace so that a straightforward implementation of omni-directional, unlimited walking is not possible. Various prototypes of locomotion devices have been developed to prevent a displacement during walking in the real world. These devices include omni-directional treadmills [BS02, SRS+11], motion foot pads, robot tiles [IHT06, IYFN05], and motion carpets [STU07]. Although these locomotion devices represent enormous technological achievements, they are still cost-prohibitive and will not be generally accessible in the foreseeable future.

Redirected Walking

Cognition and perception research suggest that cost-efficient as well as natural alternatives exist. It is known from perceptive psychology that vision often dominates proprioception and vestibular sensation when the senses disagree [Ber00, DB78]. In perceptual experiments, where human participants can use only vision to judge their motion through a virtual scene, they can successfully estimate their momentary direction of locomotion, but are worse in perceiving their paths of travel [BIL00, LBv99]. Since humans tend to unwittingly compensate for small inconsistencies during walking, it becomes possible to guide users in IVEs along paths in the real world which differ from the paths perceived in the virtual world. This *redirected walking* enables users to explore a virtual world that is considerably larger than the tracked workspace [RKW01]. Some techniques guide users on bent physical paths for

which lengths as well as active turning angles of the visually perceived paths are maintained. However, if the physical paths are bent with a radius of less than 22 meters a user may observe the discrepancy between both worlds [SBJ+10] and has to consciously follow the visual stimuli [GNRH05, NHS04], which may introduce severe cognitive demands.

Theoretical Models of Cognitive Resources

Human working memory draws from finite cognitive resources, for which several theoretical models have been proposed [GC93], which usually distinguish at least between verbal and spatial resources [BH74]. A 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, access to verbal and spatial working memory and general attention is handled by a central executive. General attention is characterized by similar demands on verbal and spatial working memory.

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. These aspects may cause users to employ strategies requiring additional cognitive resources, which compete for resources from the same pools that are utilized for successful completion of a user's primary tasks.

In this chapter, we analyze interactions between redirected walking and cognitive spatial and verbal working memory tasks. In particular, in this chapter 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 in IVEs.

The chapter is structured as follows. Section 4.2 discusses related work in the scope of the chapter. Section 4.3 details the experiment in which we evaluate cognitive demands during redirected walking. The results are presented in Section 4.4 and discussed in Section 4.5. Section 4.6 concludes the chapter.

4.2 Background

Redirected walking in IVEs is currently the focus of many research groups, analyzing locomotion and perception in both the real world and virtual worlds [SBS+12]. Basic redirection techniques make use of a *stop-and-go* approach, i. e., the virtual world is rotated around the center of stationary users until they are oriented in such a way that no physical obstacles are in front of them [KBMF05, PWF08, RKW01]. Then, users can continue to walk in the desired virtual direction. The alternative is to *continuously* apply redirection while the user

is walking through the VE [GNRH05, RKW01, SBRH08]. For instance, if users are walking straight ahead in the virtual world, small iterative rotations of the camera redirect them to walk along a circular path in the opposite direction in the real world. When redirecting a user, the visual feedback is consistent with movement in the VE, whereas proprioceptive feedback reflects movement in the physical world. When the applied manipulations are small enough, the user has the impression of being able to walk in any direction in the VE without restrictions imposed by the limited physical workspace.

Steinicke et al. [SBJ+10] have performed different experiments to identify thresholds which indicate just-noticeable differences between vision and proprioception while the user is moving. In particular, they determined thresholds for detecting translational, rotational, and curvature manipulations, which were formalized using locomotion gains [SBJ+10]. Such gains describe ratios between a user's path in the real world which is decoupled from their path in the VE. In order to determine detection thresholds for path curvature, users walked a straight path in the VE, which was bent by a curvature gain either to the left or to the right in the real world. Users had to judge if the physical path was bent left or right in a *two-alternative forced-choice (2AFC)* task. According to Steinicke et al. [SBJ+10], a straight path in the VE can be turned into a circular arc in the real world with a radius of at least 22m, for which users are not able to consciously detect manipulations. Thus, if the physical workspace in the VR laboratory supports a walking area of at least 45m×45m users can walk an infinite straight distance in the virtual world, while, in fact, they are walking a circular path in the real world. Although these psychophysical experiments revealed detection thresholds which provide insight into the ability of users to perceive redirected walking manipulations, no previous work has been performed to analyze which effect it has on cognitive resources when users are redirected. In particular, to our knowledge, no previous work provided insight into the amount of cognitive demands that are induced by redirected walking manipulations of different magnitudes on the walking user.

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 (see [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 [BDA+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 [MISH00] or even electrical stimulation [RDD+05]. For instance, using a counting-backwards task, effects on locomotion were observed in older adults, but not in young adults [BDA+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 [NZL+10].

This hypothesis is supported by experiments by Zanbaka et al. [ZLB+05], who have shown that using certain near-natural locomotion interfaces can be cognitively demanding. Nadkarni et al. [NZL+10] have shown that cognitive tasks which activate working memory and spatial attention can have an effect on human locomotion. In particular, they found that changes in gait, including speed, stride length, and double support time, were affected by cognitive tasks. In experiments by Marsh et al. [MKDO13], the dual-task selective-interference paradigm was implemented to analyze the impact of spatial and verbal cognitive tasks on locomotion. They compared the cognitive resource demands of locomotion user interfaces that varied in their naturalness as well as the impact of a restricted field of view (FOV) on cognitive working memory demands while moving in an IVE. Their results suggest that locomotion with a less natural interface can increase spatial working memory demands, and locomotion with a smaller FOV can increase general attentional demands.

4.3 Experiment

As discussed above, the application of 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 between redirected walking and two different (spatial and verbal) cognitive tasks.

4.3.1 Participants

16 participants (11 female and 5 male, ages $19 - 45$, M=27.6) completed 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. No other vision disorders have been reported by our participants. Ten participants had participated in an experiment involving head-mounted displays (HMDs) before. We measured the interpupillary distances (IPDs) of our participants before the experiment [WGTCR08]. The IPDs of our participants ranged between $5.6-6.7$cm (M=6.3cm, SD=0.3cm). We used the IPD of each participant to provide a correct perspective on the HMD. Participants were naïve 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.

4.3.2 Material

We performed the experiment in a $12\text{m}\times5\text{m}$ darkened laboratory room. As illustrated in Figure 4.1, participants wore a wireless Oculus Rift DK1 HMD for the stimulus presentation,

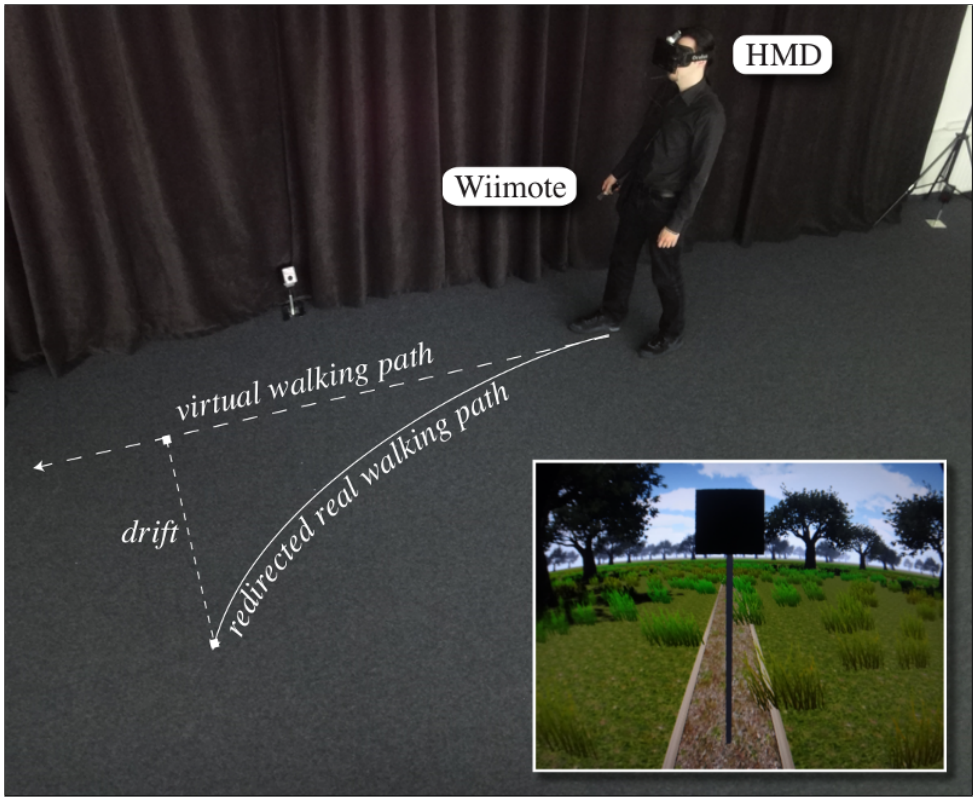


Figure 4.1: Redirected walking scenario: A user walks in the real workspace with an HMD on a curved path in comparison to the visually perceived straight path in the virtual world. The inset shows the user's view on the HMD as used for the experiment.

which provides a resolution of 640×800 pixels per eye with a refresh rate of 60Hz and an approximate 110° diagonal field of view. We attached an active infrared marker to the HMD and tracked its position within the laboratory with a WorldViz Precision Position Tracking PPT-X4 active optical tracking system at an update rate of 60Hz with sub-millimeter precision for position data in the laboratory. The head orientation was tracked with an InterSense InertiaCubeBT sensor at 180Hz update rate, which we attached to the HMD. We compensated for inertial orientation drift by incorporating the PPT-X4 optical heading plugin. The visual stimulus consisted of a simple VE with grass, trees and pavement (see Figure 4.1). We used an Asus WAVI wireless kit to transmit the rendered images at 60Hz from a rendering computer to the HMD. As claimed by the manufacturers, not more than 2ms latency are introduced due to the wireless connection. The HMD and wireless transmitter box were powered by an Anker Astro Pro2 portable battery. The boxes were carried in a small belt bag. For rendering, system control and logging we used an Intel computer with 3.4GHz Core i7 processor, 16GB of main memory and two Nvidia GeForce 780Ti SLI graphics cards. The stimuli were rendered with the Unity 3D Pro engine. In order to focus participants on the task, no communication between experimenter and participant 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 render forest and nature sounds, which minimized 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.

4.3.3 Methods

We used a 3×9×2 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 [SBJ+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 a fixed order of the cognitive conditions but randomized the locomotion conditions. This ensured that none of the cognitive tasks would be favored due to potential training effects (see also [MHZS13]).

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 Kennedy-Lane simulator sickness questionnaire (SSQ) [KLBL93] immediately before and after the experiment, further the Slater-Usoh-Steed (SUS) presence questionnaire [UCAS99], 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. In total, participants completed 12 training trials.

Locomotion Tasks

For each trial, participants were instructed to direct their gaze to a target sign displayed in front of them on the virtual pavement (see Figure 4.1). The target moved at a speed of 0.55m/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 2m, similar to the task used by Neth et al. [NSE+11]. The total distance walked was 7m over a duration of 12.6s, 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 next 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 [SBJ+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 in order 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 4.1. Curvature gains are expressed by $g_C := \frac{1}{r}$, with $g_C = 0$ for real walking with $r = \infty$. Gains g_C less than zero correspond to counterclockwise circular movements, whereas gains greater than zero correspond to clockwise circular movements in the physical workspace. If the injected manipulations are reasonably small, users will accurately compensate for the virtual rotations and walk along a curved path. Curvature gains $|g_C| \in [0, 0.045]$ are considered undetectable for users, which corresponds to radii of at least 22 meters (cf. [SBJ+10]). We tested gains $g_C \in \{\pm\frac{1}{2.5}, \pm\frac{1}{5}, \pm\frac{1}{10}, \pm\frac{1}{15}, 0\}$, i. e., each curvature was tested both in clockwise and in counterclockwise direction. The tested gains correspond to circular radii that fit within laboratories with a walking area of 5m×5m, 10m×10m, 20m×20m, or 30m×30m, respectively.

In the experiment we evaluated *lateral head movements* when walking straight ahead along the path in the VE, 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 in the verbal as well as spatial cognitive tasks was set to 500ms with a pseudo-randomized interstimulus interval of 1100−1500ms 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 by walking behind the 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 4.2(a), the verbal working memory task was a letter *two-back task* [GC93]. In every trial, participants were shown a continuous stream of letters that were flashed on the virtual target surface in the VE. The close distance of the target to the user ensured always 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 4.2(a)). 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 continuously appear in the center of the target surface.

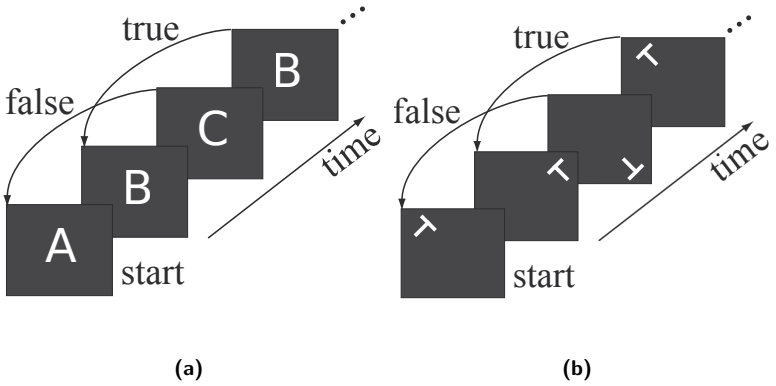


Figure 4.2: Illustration of the cognitive two-back tasks: (a) verbal working memory task and (b) spatial working memory task.

Spatial two-back working memory task

As illustrated in Figure 4.2(b), 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 target surface in the VE. The stimulus appeared in one of the four corners of the target surface rotated by $\Theta \in \{\frac{1}{4}\pi, \frac{3}{4}\pi, \frac{5}{4}\pi, \frac{7}{4}\pi\}$ radians. Participants responded 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 4.2(b)). This task did not require large verbal working memory. The displayed symbols are considered hard to verbalize (cf. [BRK07]).

4.4 Results

We found no effect of walking clockwise or counterclockwise along the circular paths and therefore pooled the data. Figure 4.3 shows the pooled results for the tested curvature gains plotted against the performance of the locomotion and cognitive tasks. The vertical bars show 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 4.3(a), and the percentage of correct responses in the cognitive tasks in Figure 4.3(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.

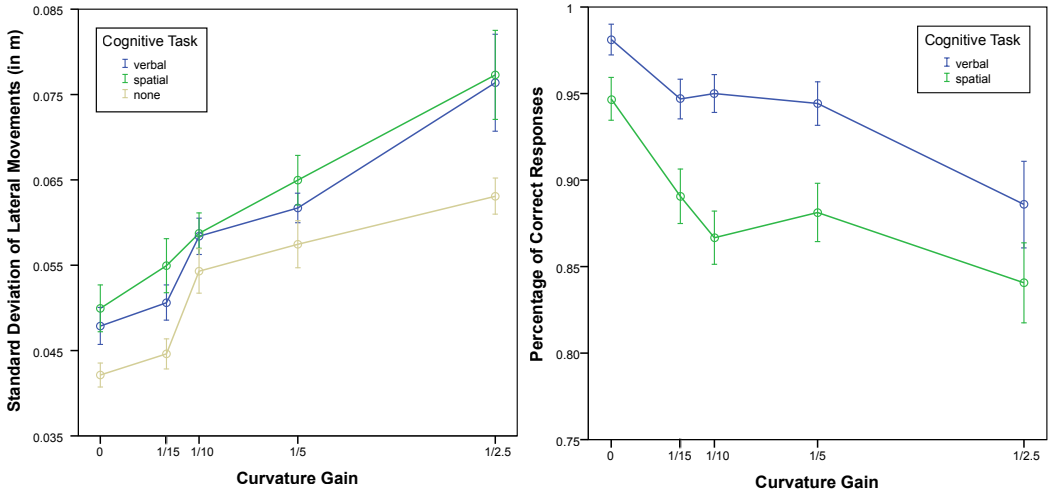


Figure 4.3: Pooled results of the experiment with (a) standard deviation of lateral movements, which indicates corrective movements during forward locomotion, for the verbal and spatial cognitive tasks and no-task condition, and (b) percentage of correct responses for the verbal and spatial cognitive tasks.

4.4.1 Locomotion Performance

We observed low lateral sway, i. e., lateral movements of the participants during forward walking, without redirected walking manipulations and without cognitive tasks, which was indicated by a standard deviation in lateral head movements of M=.042m, SD=.0058m. Increased lateral head or body movements are a common occurrence in old adults, for whom increased lateral instability during walking is hypothesized to cause more compensatory lateral movements. Since strong redirected walking manipulations require users to consciously compensate for lateral and rotational changes during virtual straightforward walking [DAK07], we hypothesized that we would find similar effects in our experiment. Indeed, lateral sway increased for both cognitive tasks as well as for larger curvature gains.

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, η_p^2=.694). Post-hoc tests showed that the lateral sway between each two curvature gains was significantly different with larger gains corresponding to larger sway (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, η_p^2=.444). Post-hoc tests showed that the lateral sway was significantly higher with the spatial task compared to no task (p=.001), as well as with the verbal task compared to no task (p=.013), 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, η_p^2=.094).

4.4.2 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, η_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<.02, η_p^2=.186). For the verbal task, the results show a significantly lower percentage of correct responses for the largest tested curvature gain $|g_C|=\frac{1}{2.5}$ compared to all other gains except for $|g_C|=\frac{1}{5}$. In particular, post-hoc tests showed for the verbal task a significant lower percentage of correct responses for curvature gain $|g_C|=\frac{1}{2.5}$ compared to curvature gain $g_C=0$ (p<.001), for curvature gain $|g_C|=\frac{1}{2.5}$ compared to curvature gain $|g_C|=\frac{1}{15}$ (p=.04), as well as for curvature gain $|g_C|=\frac{1}{2.5}$ compared to curvature gain $|g_C|=\frac{1}{10}$ (p=.027). For the spatial task, the results show a significant lower percentage of correct responses for curvature gain $|g_C|=\frac{1}{5}$ compared to curvature gain $g_C=0$ (p=.013), as well as for curvature gain $|g_C|=\frac{1}{15}$ compared to curvature gain $g_C=0$ (p<.001).

4.4.3 Questionnaires

We measured a mean SSQ-score of M=13.2 (SD=15.2) before the experiment and a mean SSQ-score of M=47.4 (SD=60.8) after the experiment. The results indicate a typical increase in simulator sickness for extensive walking with an HMD over the time of the experiment. The mean SUS-score for the sense of feeling present in the VE was M=4.71 (SD=.87), which indicates a high sense of presence [UCAS99]. 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), which is often not the case in redirected walking implementations.

4.5 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 [SBJ+10] a straight path in the VE can be turned into a circular arc in the real world with a radius of approximately 22 meters, 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 a significant increase of lateral sway for both spatial task

as well as verbal task for gains larger than $|g_C|=\frac{1}{10}$. 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 are clearly able to detect the manipulation, cognitive task performance for spatial as well as verbal tasks decreases, and in addition 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 large redirected walking manipulations require additional cognitive resources by the user which are competing for finite cognitive resources. With increasing amounts of manipulations, the required cognitive resources also increase. If manipulation with curvature gains larger than $|g_C|=\frac{1}{10}$ are applied, users are clearly able to detect the manipulation. However, 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. Based on our results, we cannot recommend applying redirected walking with curvature gain manipulations in VR laboratories with a size below 10m×10m in case users have to perform complex cognitive tasks. If possible, VR developers should apply curvature gain manipulations below the detection thresholds of $|g_C|=\frac{1}{22}$. If redirected walking is to be applied in smaller VR laboratories, we recommend combining curvature gains with rotation and translation gains to minimize the magnitude of manipulations (e. g., see [KBMF05, PFW11, SBJ+10, SBS+12]).

4.6 Conclusion

In this chapter, we presented an experiment in which we evaluated the mutual influence between 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.

For future work, we suggest comparative analyses of cognitive demands of redirected walking with other locomotion techniques such as in-place walking or 3D traveling techniques based on joysticks or gamepads.

Part III

Computer-Mediated Perceptual Illusions

Chapter 5

5 Going With the Flow: Modifying Self-Motion Perception with Computer-Mediated Optic Flow

One major benefit of wearable computers is that users can naturally move and explore computer-mediated realities. However, researchers often observe that users' space and motion perception severely differ in such environments compared to the real world, an effect that is often attributed to slight discrepancies in sensory cues, for instance, caused by tracking inaccuracy or system latency. This is particularly true for virtual reality (VR), but such conflicts are also inherent to augmented reality (AR) technologies. Although head-worn displays will become more and more available soon, the effects on motion perception have rarely been studied, and techniques to modify self-motion in AR environments have not been leveraged so far.

In this chapter, we introduce the concept of *computer-mediated optic flow*, and analyze its effects on self-motion perception in AR environments. First, we introduce different techniques to modify optic flow patterns and velocity. We present a psychophysical experiment which reveals differences in self-motion perception with a video see-through head-worn display compared to the real-world viewing condition. We show that computer-mediated optic flow has the potential to make a user perceive self-motion as faster or slower than it actually is, and we discuss its potential for future AR setups.

5.1 Introduction

Computer-mediated reality refers to the ability to manipulate one's perception of the real world through the use of wearable computers such as head-worn displays or hand-held devices. Applications include the visual augmentation of the real world with virtual objects, but can also encompass the visual "subtraction" of certain objects or information from the real world, for instance, to provide an alternative perspective [AST09]. As a result, the view of the real

world can be modified in two ways, i. e., *diminished* as well as *augmented*. Although the technology has reached a level where effective computer-mediated realities can be designed, researchers and practitioners are still attempting to solve many fundamental problems. For example, it is an often observed phenomenon that users perceive a computer-mediated reality with scene and depth distortions, which can potentially lead to poor task performance [KSF10, LDST13]. While most of the previous work in augmented reality (AR) research in this context was focussed on registration, illumination and depth problems, the perception of motion was ignored for a long time [KSF10]. With the rapid advent of applications based on wearable computers such as cell phones or head-worn displays, it can be foreseen that users will soon be able to explore computer-mediated realities, for example, by walking or driving around. Hence, it is crucial that we gain a better understanding of the perception of motion in computer-mediated realities.

It is still an open research question whether and how the perception of self-motion is affected in AR environments, i. e., if self-motion perception is changed by optical or video see-through equipment. In contrast, significant effort has been undertaken in virtual reality (VR) to determine perceptual limitations and misperception during self-motions in immersive display environments [SVCL13, VSB+10]. In particular, Razzaque et al. [Raz05] conceptually proposed and Steinicke et al. [SBJ+10] experimentally determined points of subjective equality as well as detection thresholds for a variety of walking motions in immersive head-mounted display (HMD) environments. The results revealed significant misperception of walked distances and head turn angles, which greatly limit the applicability of immersive virtual environments (IVEs) for exploration tasks, e. g., in the field of construction or architecture, in which an accurate spatial impression of virtual models is essential [IAR06].

However, the observation that self-motion perception in IVEs is limited and biased has also introduced a range of novel interface characteristics that become possible by exploiting perceptual limitations. In particular, the finding that users are often not able to consciously detect slight discrepancies between self-motion in the real world and in virtual scenes has stimulated multiple research directions that imperceptibly guide users on paths in the real world that differ from the visually perceived paths [Raz05]. A growing number of applications are based on exploiting undetectable subliminal sensory manipulations, enabling new research directions, such as the natural interaction with impossible spaces [SLF+12].

For augmented reality (AR) environments, such approaches based on exploiting limitations of self-motion perception have not been studied so far. One reason could be that it is not trivial to introduce a discrepancy between actual and visually perceived position or orientation in AR. However, research on visual illusions in IVEs based on optic flow fields recently suggested that it may be possible to change a user's self-motion perception even with AR displays [BSWL12]. The illusions were based on the direct stimulation of retinal motion detectors using the transparent overlay of rendered scenes with three-dimensional virtual flow fields [Gie97], contrast inversion or change blindness illusions [SBHW10], or time-varying displacements of object contours [FAH91]. Until now, the potential of changing self-motion perception in AR via integration of actual and apparent optic flow motion sensations has not

been considered.

The chapter is structured as follows. Section 5.2 provides background information on self-motion and optic flow. Section 5.3 presents three techniques to manipulate optic flow fields in AR. Section 5.4 describes the psychophysical experiment that we conducted to evaluate the techniques. Section 5.5 concludes the chapter.

5.2 Background

In this section, we give an overview of self-motion perception, optic flow, and illusory transformations in optic flow patterns.

5.2.1 Self-Motion Perception

The perception of self-motion speed, direction and path of travel is based on the combination and integration of various cues provided by the sensory systems [Ber00]. When moving in the real world, humans receive vestibular information about linear and angular accelerations of the head [Raz05], as well as proprioceptive and kinesthetic information about the locomotor state of the body [SVCL13]. The information from the body senses is supplemented by auditive and visual sensory information about the movement of the observer relative to the environment. Visual information consisting of landmarks and optic flow [LBv99] has often been found to *dominate* self-motion perception [BPY75, BSWL12], suggesting that visual stimulation often provides the most reliable cues about travel distance and direction. This observation is often exploited in IVEs, in which users experience virtual self-motion via visual displays, while cues from the body senses in the real world often indicate limited or no self-motion at all [SBH+09].

However, even if a user's head movements are mapped veridically to a virtual camera in IVEs, the different sensory cues often do not agree as in the real world, which results in large differences in spatial and self-motion perception [SBJ+10, SVCL13]. Various hard- and software factors have been elucidated that can cause misperception in IVEs. However, the major contributing factors remain unknown [IAR06]. Distance and size misperception effects in AR environments are often found to occur at a much lower degree than in IVEs [JSSE11]. So far, limited information exists on how technical or cognitive issues in AR environments affect self-motion perception.

5.2.2 Optic Flow

The pattern of light rays that reach an observer's eye after being structured by surfaces and objects in the world can be seen as an *optic array*. When an observer moves in the world the motion that is produced in this array is called optic flow [Gib50]. For example, forward movement produces optic flow radiating out from the point the observer is heading towards, which is visually observable as the focus of expansion (see Figure 5.1), whereas backward movements create a focus of contraction in the optic array. Especially in peripheral regions of the visual field, the sensitivity to self-motion information from optic flow is very

high [BSWL12, Pal99]. When incoming signals from the various senses are inconsistent, optic flow can dominate self-motion perception [Rie10]. For example, when a neighboring train starts to move while sitting in a stationary train, a sensation of self-motion is caused although mechanical forces on the body do not change. This phenomenon is called *vection* [BPY75]. The influence of optic flow on self-motion perception has been shown by many studies. Lee and Lishman [LL75] used a "swinging room" where the walls could be moved backward or forward unnoticed by the participant. When the movement of the walls caused optic flow that would be produced when swaying towards the wall, participants swayed or fell backward. Pailhous et al. [PFFB90] varied optic flow velocity while participants were instructed to maintain a constant walking speed. They observed modulations in stride length and cadence. Rieser et al. [RPAG95] investigated aftereffects of discrepancies between optic flow velocity and walking speed. After walking for a while, participants had to walk blindfolded to a distant target. The group that had previously walked faster than specified by the optic flow overshot the target while the other group undershot it.

5.2.3 Motion Illusions

The visual system uses all motion information available in the visual field to build a percept of self-motion. When specific motion patterns are added to an optic flow field, illusory percepts can arise even if the basic motion pattern remains unchanged. Duffy and Wurtz [DW93] presented an outward radial optic flow field that was overlapped by unidirectional horizontal motion. Although participants perceived two different motion patterns, they reported a change in their self-motion percept. The focus of expansion appeared to be shifted in the direction of the horizontal motion. A simple vector averaging of the flow fields could not account for the illusion because it predicts a shift of the focus of expansion in the opposite direction. The illusion has rather been explained by compensatory heading detection mechanisms that shift the perceived heading against the rotation of the observer [LR95]. Bruder et al. [BSWL12] manipulated peripheral optic flow in an IVE by overlaid three-dimensional motion, masked phases of camera offsets and time-varying displacements of object contours. After walking a few meters in the virtual scene, participants had to decide whether their virtual movement was smaller or larger than the physical movement. In all conditions, manipulations affected participants' percepts of self-motion. However, in the overlaid condition, only one of three manipulations had a significant impact on participants' judgments, namely motion of textures fitted to the scene. Simple particle flow and moving sinus gratings had no significant effect, probably because the visual system interpreted them as external motion in the scene, rather than self-motion. Changing global image properties can also lead to illusory percepts of self-motion derived from optic flow. In simulated driving environments it has been found that a reduction of luminance increases perceived speed while a reduction of contrast decreases it [PH12, SSR98].

Figure 5.1: Photo of a user wearing a tracked NVIS nVisor MH-60V video see-through HMD. The inset shows an illustration of the resulting expansional optic flow field during forward movements.

5.3 Computer-Mediated Optic Flow

In this section, we describe how visual self-motion feedback can be changed with video see-through displays using techniques based on optic flow manipulations. In the following, we assume that a user is wearing a video see-through HMD which is tracked in an environment that supports natural locomotion. For each frame, the change in position $\vec{p} \in \mathbb{R}^3$ and orientation with Euler angles $\vec{o} \in \mathbb{R}^3$ from the previously tracked state to the current state is measured. The differences in position and orientation divided by the elapsed time $\Delta t \in \mathbb{R}$ between rendering frames can be used to compute and manipulate the velocity of optic flow fields.

Three types of optic flow manipulations can be distinguished: those that are based on temporal transformations, those that make use of screen space transformations, and those that introduce pixel motion to increase, decrease or redirect optic flow patterns. In the following sections, we focus on translational self-motions and present different approaches to introducing optic flow patterns that differ from the visual fields naturally experienced when moving in AR environments. We discuss how these can be parametrized and steered with gains $g_v \in \mathbb{R}$ denoting scale factors of the visual self-motion velocity. Using this parametrization, gains $g_v > 1$ result in an increased visual self-motion velocity, whereas gains $g_v < 1$ decrease the visual velocity.

5.3.1 Temporal Transformations

Video see-through HMDs are based on the principle that images are captured by built-in cameras at the approximate positions of the user's eyes, processed, and presented on visual displays to the eyes of the user. Usually, for each rendering frame, the most current camera

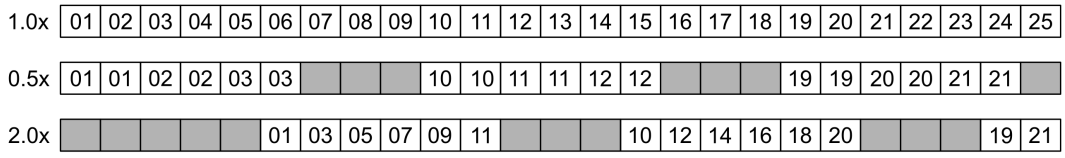


Figure 5.2: Illustration of different temporal transformations: The top row shows camera images captured at each frame and displayed to the user on the video see-through display, which results in a natural visual velocity. The center row illustrates temporal expansion by displaying each camera image for twice its natural time span resulting in half the visual velocity. The bottom row illustrates temporal compression by skipping each second camera image resulting in twice the visual velocity. The blank screens denote inter-stimulus intervals.

images are displayed to minimize latency between movements in the real world and their feedback via the AR display (illustrated in Figure 5.2(top)). However, motion feedback with video see-through displays can be changed by introducing offsets between subsequently captured camera images. Constant offsets, e. g., presenting the nth previously captured image to the user, correlate to changing the latency of the AR system. In contrast, using dynamic offsets, such as skipping every second frame or displaying each captured camera image for two frames on the HMD, results in accumulating temporal offsets and can disrupt the natural perceived temporal continuum (see Figure 5.2). While such temporal compressions or expansions provide the means to change the perception self-motion velocity, the inherent accumulation of latency reduces their applicability over time frames of more than a few hundred milliseconds.

A recently proposed solution to keeping accumulating offsets in check has been observed by Bruder et al. [BSWL12] and originates in the field of *change blindness* illusions [ROC97, SBHW10]. Change blindness denotes the phenomenon that an observer presented with a visual scene may fail to detect a significant change in the scene in case of a brief visual disruption, in particular if the visual disruption is long enough to clear the retinal afterimage or provoke contrast inversion in the afterimage [Mat06]. Rensink et al. [ROC97] observed that when changes to the scene are synchronized with measured blinks or saccades of an observer's eyes, e. g., exploiting *saccadic suppression*, scene changes can be introduced roughly every 250ms (assuming about 4 saccades and 0.2 blinks per second for a healthy observer [DUV02]). Additionally, Rensink et al. [ROC97] showed that blanking out the screen for 60-100ms is well suited to introduce a controllable visual disruption that can be used to stimulate and study change blindness in a visual scene.

Figure 5.2(bottom) illustrates in which order captured camera images can be displayed to an observer so that the resulting visual velocity is always twice ($g_v = 2$) as fast as using a veridical mapping. The blank display frames denote how change blindness can be used to hide a backward jump in time from the observer, in particular, without stimulating retinal motion detectors during reverse motion. Figure 5.2(center) provides an example for reducing visual velocity by 50% ($g_v = 0.5$), in which the blank frames conceal a forward jump in time. The maximal latency introduced by the two examples is defined by how often visual disruptions naturally occur or can be introduced with blank screens, i. e., how often change

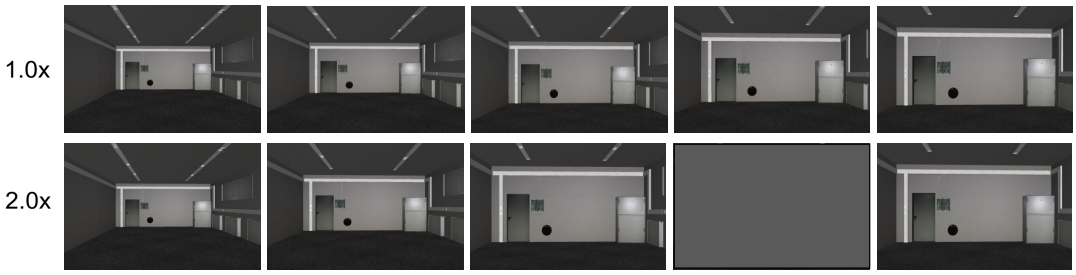


Figure 5.3: Illustration of screen space transformations: The top row shows camera images captured while the observer is moving forward. The bottom row shows an increased visual velocity with magnification and change blindness. The blank screen denotes an inter-stimulus interval.

blindness can be exploited. While the temporal scale factors shown in Figure 5.2 are easy to implement, variable visual velocities can be introduced by blending or morphing intermediate images between two camera frames (e. g., based on Mahajan et al. [MHM+09]).

5.3.2 Screen Space Transformations

It is quite common with video see-through HMDs that the built-in cameras have a slightly larger field of view than provided by the display optics. As a result, the visual angles have to be calibrated to provide a matching view of the real world as without the AR HMD. In the following, we denote the view angles of the camera images as *camera field of view* (CFOV), and that of the HMD as *display field of view* (DFOV). If the CFOV matches the DFOV (i. e., CFOV=DFOV), the viewport is mapped from the camera space onto the physical display in such a way that users perceive a "correct" perspective of the real world (assuming that we neglect other distortions of the camera or HMD such as pincushion distortion [KTCR09]). In case of CFOV<DFOV, the view of the real world appears magnified on the display, such as with a telephoto lens, because of the requirement for the camera image to fill a larger subtended angle of the HMD optics [KTCR09]. For CFOV>DFOV, the view of the real world is minified and compressed in a smaller visual field such as with a wide-angle lens [KTCR09]. Minification and magnification change several visual cues that provide distance information. In particular, the changed retinal size makes familiar objects appear closer or farther away [Lou07], and binocular convergence indicates a shift in distance [BPS12].

Controllable *minification* and *magnification* can be implemented for a fixed CFOV and DFOV as follows. With gains of $g_F \leq 1$ the used image region of the total CFOV can be changed as $g_F \cdot$ CFOV. Steinicke et al. [SBL+11] proposed the following equation to compute the amount of minification or magnification of the view:

$$m = \frac{\tan\left((g_F \cdot \mathrm{CFOV}) \cdot 0.5\right)}{\tan(\mathrm{DFOV} \cdot 0.5)} \tag{5.1}$$

for vertical fields of view with $m < 1$ denoting view magnification ($g_F \cdot$ CFOV>DFOV), and $m > 1$ denoting view minification ($g_F \cdot$ CFOV<DFOV). As described by Bruder et al. [BPS12], minification or magnification cause visual distance cues to be changed:

$$\widehat{D} = D \cdot m \tag{5.2}$$

with D the actual distance to a scene object and $\widehat{D}$ the resulting object distance along the view axis after minification or magnification.

Figure 5.3 shows an example in which the top row corresponds to the actual forward motion of an observer, and the bottom row shows effects of introducing additional magnification each frame. The AR view shown in the bottom row indicates an increased visual velocity, and makes use of change blindness inter-stimulus intervals (see Section 5.3.1) to revert the accumulated magnification in the view. While the results differ in geometric distortions, previous studies have shown that users tolerate a certain deviation in perspective cues with HMDs, and often are not even able to detect a discrepancy [SBK11, SBL+11]. Depending on the CFOV of the built-in cameras of the AR HMD, a similar effect can be achieved for minification.

5.3.3 Pixel Motion Transformations

While moving with a video see-through HMD, light intensities move over the visual field, which stimulate retinal motion detectors. Although the perceptual system uses different local and global approaches to filter out noise in retinal flow [HY95, Mat06], different approaches to manipulating visual information have been proposed that have the potential to subliminally change percepts of visual velocity (see Section 5.2). In order to introduce a flow field to the periphery of the AR view on a video see-through HMD, we implemented a GLSL shader to dynamically move pixels with or against the tracked self-motion direction using an approximate model of optic flow directions on the display surface. Scaled 2D optic flow direction vectors can be extracted from subsequent camera images (cf. [WC11]). With this approach, the final color of fragments is computed using time-varied weights:

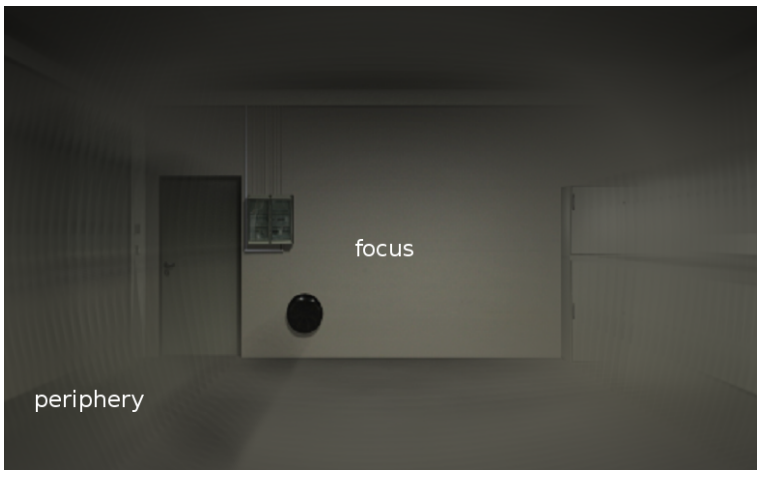


Figure 5.4: Example of pixel motion transformation with a faster or slower flow field introduced in the periphery of the observer's eyes.

$$\sum_{i=0}^{n-1} texture2D\left(cam, c + d \cdot (i+1)\right) \cdot \frac{n-i}{n} \cdot \left(\left(t + \frac{i}{n}\right) \% \frac{1}{2}\right) \tag{5.3}$$

along the 2D optic flow direction vector $d \in \mathbb{R}^2$ with pixel coordinate $c \in \mathbb{N}^2$, scaled frame time $t \in [0, 1]$, and $n \in \mathbb{N}$ (e. g., $n = 30$). Since the direction vectors d can be determined for each pixel using optic flow estimates between the current and last camera image, and their non-normalized length encodes the user's movement velocity, the pixel velocity over the periphery can be changed by $t = t + \alpha \cdot \Delta t$, with a speed factor $\alpha \in \mathbb{R}$.

Using this shader, selected pixels (e. g., only those in the periphery [BSWL12]) can be moved with or against the user's self-motion direction, thus changing the velocity of the retinal flow field (see Figure 5.4). The looped filter ensures that the peripheral flow field is perceived as continuous motion.

5.4 Experiment

In this section, we describe an experiment that we conducted to evaluate self-motion estimation in an AR environment based on a video see-through HMD. We analyze whether it is possible to change self-motion judgments with the visual augmentation techniques presented in Section 5.3. Although literature suggests that visual self-motion cues can dominate extraretinal cues [BPY75, BSWL12], it is unclear whether the proposed techniques can actually affect self-motion judgments. Studies investigating the effects of self-motion manipulations often used an adaptation procedure and measured blind-walking performance in pre- and post-tests [DPF+05, MTCRW07, RPAG95]. Here we used a condensed version of this method assuming that our techniques will have immediate effects on self-motion judgments. A single trial consisted of a stimulus phase with vision and an estimation phase without vision. We compared participants' blind walking distances towards targets directly after a stimulus phase with visual self-motion information. We varied the visual self-motion velocity by applying the augmentation algorithms with different parameters $g_v \in \mathbb{R}$ in the stimulus phase. During the estimation phase, the perceived self-motion velocity of the stimulus phase should influence the distance updating process. When the computer-mediated optic flow changes a participant's perceived self-motion velocity, the participant may walk farther or shorter than the unmediated target distance.

5.4.1 Experimental Design

We performed the experiment in a 12m×7m darkened laboratory room. As illustrated in Figure 5.1, participants wore a professional binocular NVIS nVisor MH60-V video see-through AR HMD for the stimulus presentation, which provides a resolution of 1280×1024 pixels with an update rate of 60Hz and a 60° diagonal field of view (modulo pincushion distortion [KTCR09]). The integrated VideoVision stereo cameras provide a resolution of 640×480 with an update rate of 60Hz and a slightly larger field of view of approximately 78°, which

we matched to the HMD. We measured an end-to-end latency of approximately 110ms with the video see-through HMD by evaluating the response time of photodiodes using A/D-converters. We tracked the HMD within the laboratory with a WorldViz Precision Position Tracking PPT-X4 active optical tracking system at an update rate of 60Hz. For rendering, system control and logging we used an Intel computer with 3.4GHz Core i7 processor, 16GB of main memory and Nvidia GeForce GTX 680 graphics card. The stimuli were introduced to the AR view using OpenCV, OpenGL, and GLSL. Participants judged their perceived self-motions via button presses on a Nintendo Wii remote controller.

5.4.2 Participants

10 male and 10 female (age 20-44, avg: 27.7) participants completed the experiment. All participants were students or members of the departments of computer science or psychology. All participants had normal or corrected to normal vision; 6 wore glasses and 2 contact lenses during the experiment. None of the participants reported known eye disorders, such as anisometropia, strong eye dominance, color or stereo blindness, and none of the participants reported disorders of the proprioceptive-vestibular systems. 9 of the participants had experience with HMDs. 9 had no experience with 3D games, 6 had some, and 5 had much experience. Participants were naïve to the experimental conditions. The total time per participant including pre-questionnaires, instructions, training, experiment, breaks, post-questionnaires, and debriefing was about 1 hour. Participants were allowed to take breaks at any time.

5.4.3 Material and Methods

For the experiments, we used a within-subjects design with a perception-action task based on the method of blind walking. Figure 5.5 illustrates the setup used during the experiment. At the beginning of each trial, participants pressed a button on a Wii remote controller, and the AR view was presented on the video see-through HMD. We instructed the participants to walk a distance of 3m straight at a convenient speed in our laboratory, during which they received visual feedback of their real self-motion via the AR HMD. Participants were instructed to walk towards and focus on a visual marker displayed at approximate eye height at a distance of 10m in the laboratory (see Figure 5.5). This instruction ensured that participants kept their visual focus in walking direction. Targets were presented as vertical poles and shown at 3m, 4m, and 5m distance after the initial 3m walking distance. After walking the initial 3m, the HMD turned black, while participants were instructed to walk the remaining distance to the targets without vision, and press a button on the Wii remote controller once they estimate that they have reached the middle between the vertical poles. We measured the walked distance along the ground plane between the start and end positions. The display remained black until the participant was guided back to the start position to limit feedback about walked distances between trials.

The procedure extends the traditional blind walking metric for distance judgments [LK03] with an initial visual stimulus phase. We use this phase to augment the view on the video

Figure 5.5: Experiment setup: participant walking in the direction of a target displayed at eye height for a distance indicated by the center of two vertical poles.

see-through HMD with the three optic flow transformations presented in Section 5.3. Each technique was parametrized to either double the optic flow speed with a gain of $g_v = 2.0$, provide a matching motion speed with a gain of $g_v = 1.0$, or half the optic flow speed with $g_v = 0.5$ (cf. Section 5.3). For the temporal and screen space transformations we applied $83.\overline{3}$ms inter-stimulus intervals, with $133.\overline{3}$ stimulus duration in between. For the pixel motion transformations, we limited manipulations to the periphery, providing an unmodified visual focus field of about 40 degrees.

When a computer-mediated optic flow technique changes a participant's perceived self-motion velocity, we would expect the participant to walk shorter or longer to the targets, depending on the stimuli. Differences in judged walking distances for gains of $g_v = 1.0$ provide insight into general effects of the AR hard- and software on the task. Relative differences for gains of $g_v = 2.0$ and $g_v = 0.5$ reveal the potential of the techniques to change self-motion percepts. We presented all independent variables randomly and uniformly distributed over all trials. Participants were instructed before the experiment to "ignore any and all visual modulations that may occur in the first part of the trials, and just walk to the target." We added this instruction to minimize potential experiment biases on the results that may stem from participants anticipating the hypotheses of the experiment.

We tested each of the three techniques (*temporal*, *screen space*, *pixel motion*) with the three gains $g_v \in \{0.5, 1.0, 2.0\}$ for the three distances 3m, 4m, and 5m. Additionally, we performed the trials for the three distances using the video see-through HMD without changing the AR view with optic flow transformations, which we refer to as the *camera view* condition. Moreover, we performed the trials in the real world without HMD as a baseline *real view*

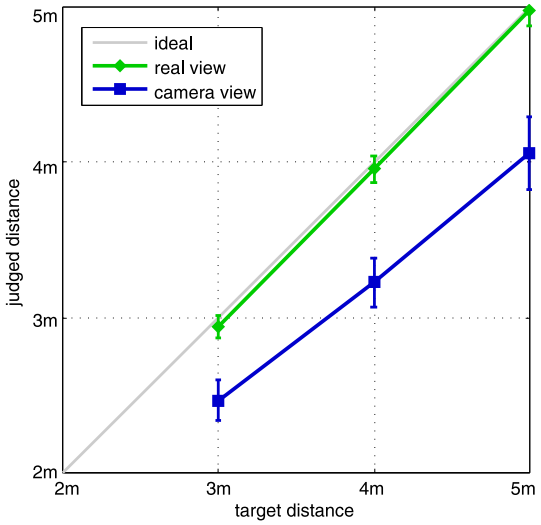


Figure 5.6: Results of judged walk distances in the real world and with the video see-through HMD. The green plot shows the results of the real view condition and the blue plot of the camera view condition. The vertical bars show the standard error.

condition, which we closely matched to the AR conditions by reducing the field of view using modified welding goggles, and by replacing the automatically triggered blind walking phases with auditive instructions to close the participant's eyes. The order of the trials was randomized between participants.

We measured effects of simulator sickness with the Kennedy-Lane Simulator Sickness Questionnaire (SSQ) before and after the experiment.

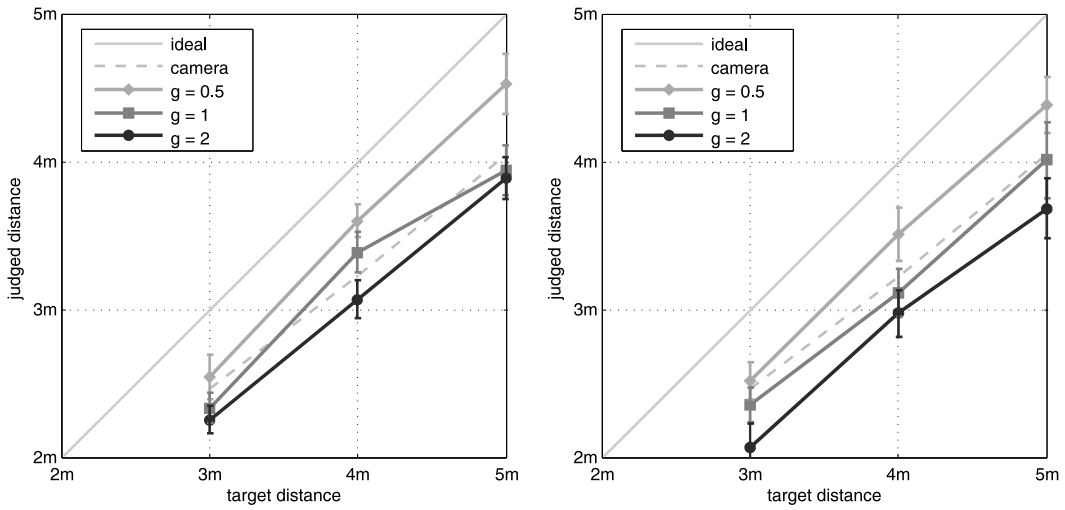
5.4.4 Results

The results of our experiments were normally distributed according to a Shapiro-Wilk test at the 5% level. The sphericity assumption was supported by Mauchly's test of sphericity at the 5% level, or degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. We analyzed the results with a repeated measure ANOVA and Tukey multiple comparisons at the 5% significance level (with Bonferonni correction).

Real view and camera view

Figure 5.6 shows the pooled results for the real view and camera view conditions, i. e., the differences between performing the task in the real world compared to with the AR HMD. The x-axis shows the target distances, whereas the y-axis shows the judged walk distances. In the real world, participants were very accurate at judging walk distances, as shown in the second column of Table 5.4.4. The walked distances with the video see-through HMD are shown in the third column of Table 5.4.4. Participants walked approximately 18% shorter in the camera view condition than without the video see-through HMD.

We found a significant main effect of condition ($F(1, 19)=28.44$, $p<.001$, $\eta_p^2=.60$) on the walked distances. As expected, we found a significant main effect of target distance ($F(1.325,$



(a) temporal **(b)** screen space

Figure 5.7: Results of the judged walk distances for (a) temporal and (b) screen space transformations. The vertical bars show the standard error.

25.173)=183.37, p<.001, η_p^2=.91) on the walked distances. We found a significant interaction effect between condition and target distance (F(1.496, 28.425)=4.33, p<.05, η_p^2=.19). Post-hoc tests revealed that the walked distances were significantly shorter for the camera view condition for all target distances.

Optic flow manipulation

The pooled results for the three tested techniques are shown in Figures 5.7 and 5.8. The x-axes show the target distances, and the y-axes show the absolute walked distances. For each technique, we plotted the results for the three applied gains $g_v \in \{0.5, 1.0, 2.0\}$ as well as the ideal and camera view results. Table 5.4.4 lists the mean absolute walked distances and relative to ideal judgments.

We found no significant main effect of technique (F(2, 38)=2.94, p<.66, η_p^2=.13) on the walked distances. We found a significant main effect of gain (F(2, 38)=45.86, p<.001, η_p^2=.71) on the walked distances, as well as a significant interaction effect between technique and

Table 5.1: Mean absolute and relative walked distances for the different target distances, view conditions, and applied gains.

	dist	real	camera	temporal			screen space			pixel motion		
		g_v=1	g_v=1	g_v=.5	g_v=1	g_v=2	g_v=.5	g_v=1	g_v=2	g_v=.5	g_v=1	g_v=2
absolute	3m	2.95	2.48	2.55	2.34	2.26	2.53	2.37	2.08	2.33	2.24	2.04
	4m	3.96	3.23	3.60	3.39	3.08	3.52	3.12	2.98	3.16	3.26	3.13
	5m	4.97	4.05	4.53	3.95	3.90	4.39	4.02	3.69	4.11	4.10	4.01
relative	3m	0.98	0.82	0.85	0.78	0.76	0.84	0.79	0.69	0.78	0.75	0.68
	4m	0.99	0.81	0.90	0.85	0.77	0.88	0.78	0.75	0.79	0.82	0.78
	5m	0.99	0.81	0.91	0.79	0.78	0.88	0.80	0.74	0.82	0.82	0.80

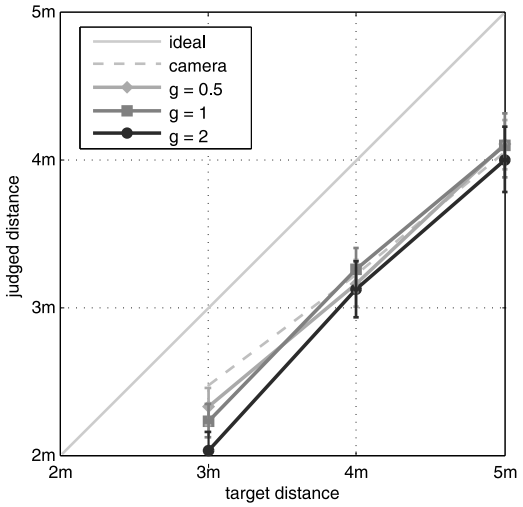


Figure 5.8: Results of the judged walk distances for pixel motion transformations. The vertical bars show the standard error.

gain (F(4, 76)=3.61, p<.01, η_p^2=.16). As expected, we found a significant main effect of target distance (F(1.332, 25.304)=228.585, p<.001, η_p^2=.92) on the walked distances. We observed no interaction effects between technique and target distance, as well as gain and target distance.

For temporal transformations, we found a significant main effect of gain (F(2, 38)=17.20, p<.001, η_p^2=.48) on the walked distances. Post-hoc tests revealed that the walked distances were significantly shorter (p<.02) for a gain of $g_v = 2.0$ compared to $g_v = 0.5$. Participants walked approximately 18% shorter for a gain of $g_v = 2.0$ compared to $g_v = 0.5$.

For screen space transformations, we found a significant main effect of gain (F(2, 38)=26.71, p<.001, η_p^2=.58) on the walked distances. Post-hoc tests revealed that the walked distances were significantly shorter (p<.03) for a gain of $g_v = 2.0$ compared to $g_v = 0.5$.

For pixel motion transformations, we found no significant main effect of gain (F(1.538, 29.213)=1.67, p<.21, η_p^2=.08) on the walked distances. Post-hoc tests revealed no significant differences in walked distances between the tested gains.

We observed no effect of the techniques on simulator sickness. SSQ scores before the experiment averaged to 7.11, with an average post-experiment score of 7.67.

5.4.5 Discussion

The results shown in Figure 5.6 indicate that participants significantly walked shorter with the AR HMD than in the real world. The results are in line with previous results in AR HMD environments [JW08, JSSE11, LDST13, SJK+07]. Although the field of view matched in both conditions, the results suggest that the weight of the HMD, low resolution, or latency in the video see-through condition may have caused a perceptual or motor difference between the conditions.

The results for temporal and screen space transformations show that judged walk distances

were significantly affected by the parametrization (see Figure 5.7). The significantly different walked distances for temporal and screen space transformations suggest that participants judged their self-motion as faster or slower depending on the applied gain, causing them to stop earlier or later for the fixed distances than without manipulation. As participants had to walk at least half of the distance without vision, the results show that the computer-mediated optic flow velocity affected how participants updated their distance to the target while walking without vision.

However, the walked distances show that visual self-motion speed estimates did not entirely dominate responses, or we would have seen participants walking precisely twice the distance for gains of $g_v = 0.5$, and half the distance for gains of $g_v = 2$. Still, participants' responses for $g_v = 0.5$ approximated ideal judgments in contrast to responses for natural optic flow velocities with $g_v = 1$, which suggests that computer-mediated optic flow can alleviate misperception of self-motion velocity in AR environments.

While temporal and screen space transformations require video see-through displays, we were particularly interested in the pixel motion transformations, since similar techniques could be transferred to optical see-through displays. However, effects of the tested parameters on judged walk distances were minimal (see Figure 5.8). It is possible that motion cues induced by the rendering technique could be interpreted by the visual system as external motion in the scene, rather than self-motion. As suggested by Johnston et al. [JBM99] this result could be explained by the interpretation of the visual system of multiple layers of motion information, in particular, due to the dominance of second-order motion information such as translations in a textured scene. However, this interpretation conflicts with perceptual experiments in IVEs by Bruder et al. [BSWL12], in which a similar peripheral stimulation has been found to have a strong influence on self-motion velocity estimates. The differences may be explained by the limitations of our hardware. In particular, limitations were the low resolution of the built-in cameras and the small field of view of our AR HMD (see Section 5.4.1), which could not stimulate a large peripheral view region. Moreover, our laboratory environment consisted mainly of gray-in-gray walls (see Figure 5.5), which may not have provided sufficient stimulation of retinal optic flow motion detectors in the peripheral view regions.

5.5 Conclusion

In this chapter, we analyzed self-motion perception in an AR environment and presented techniques to use computer-mediating reality to change optic flow self-motion cues. We introduced three different approaches to modifying optic flow velocity: temporal, screen space, and pixel motion transformations. We presented a psychophysical experiment which shows that participants wearing a video see-through HMD walked significantly shorter to a visual target after a phase with optic flow self-motion feedback than for the same task in the real world. This may be explained by a significant underestimation of target distances and/or overestimation of self-motion velocities while wearing the AR HMD. The experiment further showed that changing the visual self-motion velocity with the proposed computer-mediated

optic flow techniques had a significant effect on walked distances. In particular, for reduced optic flow velocities participants' responses approached ideal judgments. The results reveal that visually augmenting head-worn displays can be used to manipulate real-world self-motion perception, i. e., users may perceive their self-motion as faster or slower than it actually is. This shows the potential of such techniques to correct self-motion misperception, or to deliberately increase or decrease self-motion velocity estimates when desired by applications.

To summarize, in this chapter we have

(i) analyzed self-motion estimates with a head-worn video see-through display,

(ii) introduced *computer-mediated optic flow* to modify the perceived self-motion, and

(iii) evaluated the effects of three optic flow manipulation techniques on self-motion judgments.

In future work, we will analyze further approaches that may be used to change self-motion estimates using optic flow manipulations, in particular, using optical see-through display technologies. We aim to evaluate applications of computer-mediated optic flow for future AR setups. In particular, previous work suggests that changes in optic flow velocity can cause different locomotor behavior [MTCR+07, MTCRW07, RPAG95], including differences in muscular energy expenditure [GB08], which underlines the potential of visual self-motion manipulations for sports, rehabilitation, and training.

Chapter 6

6 Threefolded Motion Perception During Immersive Walkthroughs

Locomotion is one of the most fundamental processes in the real world, and its consideration in immersive virtual environments (IVEs) is of major importance for many application domains requiring immersive walkthroughs. From a simple physics perspective, such self-motion can be defined by the three components speed, distance, and time. Determining motions in the frame of reference of a human observer imposes a significant challenge to the perceptual processes in the human brain, and the resulting speed, distance, and time percepts are not always veridical. In previous work in the area of IVEs, these components were evaluated in separate experiments, i. e., using largely different hardware, software and protocols.

In this chapter, we analyze the perception of the three components of locomotion during immersive walkthroughs using the same setup and similar protocols. We conducted experiments in an Oculus Rift head-mounted display (HMD) environment which showed that participants largely underestimated virtual distances, slightly underestimated virtual speed, and we observed that participants slightly overestimated elapsed time.

6.1 Introduction

The motion of an observer or scene object in the real world or in a virtual environment (VE) is of great interest for many research and application fields. This includes computer-generated imagery, e. g., in movies or games, in which a sequence of individual images evokes the illusion of a moving picture [TFCRS11]. In the real world, humans move, for example, by walking or running, physical objects move and sometimes actuate each other, and, finally, the earth spins around itself as well as around the sun. From a simple physics perspective, such motions can be defined by three main components: (i) (linear or angular) speed and (ii) distances, as well as (iii) time. The interrelation between these components is given by the speed of motion, which is defined as the change in position or orientation of an object with respect to time:

$$s = \frac{d}{t} \tag{6.1}$$

with speed s, distance d, and time t. Motion can be observed by attaching a frame of reference to an object and measuring its change relative to another reference frame. As there is no absolute frame of reference, absolute motion cannot be determined; everything in the universe can be considered to be moving [BR12]. Determining motions in the frame of reference of a human observer imposes a significant challenge to the perceptual processes in the human brain, and the resulting percepts of motions are not always veridical [Ber00]. Misperception of speed, distances, and time has been observed for different forms of self-motion in the real world [Efr70, GGS10, MTHG10].

In the context of self-motion, walking is often regarded as the most basic and intuitive form of locomotion in an environment. Self-motion estimates from walking are often found to be more accurate than for other forms of motion [RL09]. This may be explained by adaptation and training since early childhood and evolutionary tuning of the human brain to the physical affordances for locomotion of the body [CB07]. While walking in the real world, sensory information such as vestibular, proprioceptive, and efferent copy signals as well as visual and auditive information create consistent multi-sensory cues that indicate one's own motion [Wer94].

However, a large body of studies have shown that spatial perception in IVEs differs from the real world. Empirical data shows that static and walked distances are often systematically over- or underestimated in IVEs (cf. [LK03] and [RVH13] for thorough reviews) even if the displayed world is modeled as an exact replica of the real laboratory in which the experiments are conducted in [IAR06]. Less empirical data exists on speed misperception in IVEs, which shows a tendency that visual speed during walking is underestimated [BSD$^+$05, BSWL12, DGS05, SBJ$^+$10].

We are not aware of any empirical data which has been collected for time misperception in IVEs. However, there is evidence that time perception can deviate from veridical judgments due to visual or auditive stimulation related to motion misperception [Gro08, RGK09, SGPB04]. Different causes of motion misperception in IVEs have been identified, including hardware characteristics [JSSE11, WCCRT09], rendering issues [TFCRS11], and miscalibration [KTCR09, WGTCR08].

Considering that self-motion perception is affected by different hard- and software factors, it is unfortunate that no comprehensive analysis to this day exists in which the different components were tested using the same setup. Hence, it remains unclear whether or not there is a correlation or causal relation between speed, distance, and time misperception in current-state IVEs.

Our main contributions in this chapter are:

- We compare self-motion estimation in the three components using the same setup and similar experimental designs.
- We introduce novel action-based two-alternative forced-choice methods for distance, speed, and time judgments.
- We provide empirical data on motion estimation that show researchers and practitioners

what they may expect when they present a virtual world to users with an Oculus Rift head-mounted display (HMD).

The chapter is structured as follows. In Section 6.2 we provide background information on self-motion perception in virtual reality (VR) setups. We describe the experiments in Section 6.3 and discuss the results in Section 6.4. Section 6.5 concludes the chapter.

6.2 Background

As stated above, research on distance, speed and time perception in IVEs is divided in separate experiments using largely different hardware, software and experimental protocols. In this section, we give an overview of the results for distance, speed, and time perception in IVEs.

6.2.1 Distance Perception

During self-motion in IVEs, different cues provide information about the travel distance with respect to the motion speed or time [Moh07]. Humans can use these cues to keep track of the distance they traveled, the remaining distance to a goal, or discriminate travel distance intervals [BL99]. Although humans are considerably good at making distance judgments in the real world, experiments in IVEs show that characteristic estimation errors occur such that distances are often severely overestimated for very short distances and underestimated for longer distances [LKG+93]. Different misperception effects were found over a large range of IVEs and experimental methods to measure distance estimation. While verbal estimates of the distance to a target can be used to assess distance perception, methods based on visually directed actions have been found to generally provide more accurate results [LK03]. The most widely used action-based method is *blind walking*, in which participants are asked to walk without vision to a previously seen target. Several experiments have shown that over medium range distances participants can accurately indicate distances using the blind walking method [RATY90]. Other action-based methods include *triangulated walking* and *timed imagined walking* [FLD97, KSS+09, PKC04]. Moreover, perceptual matching methods can be used, in which participants match the distance or size of a target to the distance or size of a reference object, respectively [LP08].

Although there is a large interest in solving the distance misperception problem, the reasons for this perceptual shift are still largely unknown, as are approaches to reduce such misperceptions. Kuhl et al. [KTCR09] observed that miscalibration of HMD optics is the main reason for distance misperception, although participants underestimated distances even for correctly calibrated HMDs [KBB+12]. Willemsen et al. [WCCRT09] compared HMD properties with natural viewing in the real world and observed that mechanical restrictions of HMDs can cause slight differences in distance judgments. Jones et al. [JSKB12, JSSE11] found that increasing the field of view of HMDs to approximate the visual angle of the human eyes helps alleviate distance misperception. Interrante et al. [IAR06] showed that the VE has an impact on distance judgments with underestimation being reduced if participants are immersed in

an accurate virtual replica of the real-world surroundings than in a different VE. Studies by Phillips et al. [PRI+09] further show that the visual rendering style affects distance judgments. They observed that distance underestimation was increased for a non-photorealistic rendering style than in a photorealistic scene.

6.2.2 Speed Perception

Different sensory motion cues support the perception of the speed of walking in an IVE (cf. [DGS05]). Visual motion information is often estimated as most reliable by the perceptual system, but can cause incorrect motion percepts. For example, in the illusion of linear vection [Ber00] observers feel their body moving although they are physically stationary because they are presented with large-field visual motion that resembles the motion pattern normally experienced during self-motion. Humans use such optic flow patterns to determine the speed of movement, although the speed of retinal motion signals is not uniquely related to movement speed. For any translational motion the visual velocity of any point in the scene depends on the distance of this point from the eye, i. e., points farther away move slower over the retina than points closer to the eye [BL99, War98]. Banton et al. [BSD+05] observed for participants walking on a treadmill with an HMD that optic flow fields at the speed of the treadmill were estimated as approximately 53% slower than their walking speed. Durgin et al. [DGS05] reported on a series of experiments with participants wearing HMDs while walking on a treadmill or over solid ground. Their results show that participants often estimated subtracted speeds of displayed optic flow fields as matching their walking speed. Steinicke et al. [SBJ+10] evaluated speed estimation of participants in an HMD environment with a real walking user interface in which they manipulated participants' self-motion speed in the VE compared to their walking speed in the real world. Their results show that on average participants underestimated their walking speed by approximately 7%. Bruder et al. [BSWL12, BWB+13] showed that visual illusions related to optic flow perception can change self-motion speed estimates in IVEs.

6.2.3 Time Perception

As discussed for distance and speed perception in IVEs, over- or underestimations have been well documented by researchers in different hard-, software and experimental protocols. In contrast, the perception of time has not been extensively researched in IVEs so far. Experimental studies of time perception in the field of psychology have well established that estimates of stimulus duration do not always match its veridical time interval, and are affected by a variety of factors [Efr70]. Since time cannot be directly measured at a given moment, the brain is often assumed to estimate time based on internal biological or psychological events, or external signals [Gro08]. The effect of exogenous cues (i. e., *zeitgebers*) from the local environment on endogenous biological clocks (e. g., circadian rhythms) is studied in the field of chronobiology [KM13]. It is possible that differences in exogenous time cues between the real world and IVEs have an effect on internal human time perception. In par-

ticular, system latency is known to change the perception of sensory synchronicity [SZM10] and can degrade the perceptual stability of the environment [AHJ+01].

Space and time are interdependent phenomena not only in physics but also in human perception [Gro08]. Helson keyed the term *tau effect* for the phenomenon that the variation of the time between spatial events can affect judgments of their spatial layout (cf. [HK31, JH82, SGPB04]). For instance, Helson and King [HK31] observed for a tactile estimation experiment that stimulating three equidistant surface points p_1, p_2, and p_3 with $||p_2 - p_1|| = ||p_3 - p_2||$ at points in time t_1, t_2, and t_3 for different durations $||t_2 - t_1|| > ||t_3 - t_2||$ that participants judge the distance between p_1 and p_2 as longer than between p_2 and p_3.

Conversely, Cohen et al. [CHS53] keyed the term *kappa effect* for the phenomenon that the variation of the spatial layout of events can affect judgments of their temporal layout (cf. [Gro08, RGK09]). They observed for a visual bisection task that three successive flashes at spatial points p_1, p_2, and p_3 for different distances $||p_2 - p_1|| > ||p_3 - p_2||$ with points in time t_1, t_2, and t_3 with $||t_2 - t_1|| = ||t_3 - t_2||$ that participants judge the duration between t_1 and t_2 as shorter than the duration between t_2 and t_3.

6.3 Psychophysical Experiments

In this section, we evaluate misperception while walking in an IVE. We describe three experiments that we conducted to evaluate distance, speed, and time perception based on similar measurement protocols to obtain judgments using the same IVE.

6.3.1 Participants

18 participants (6 female and 12 male, ages 19−38, M=25.0, SD=5.6) completed the distance and speed experiment, while 10 of them participated in the time experiment, which was performed one day later. The participants were students or members of the local departments of computer science, psychology, or human-computer media. Students obtained class credit for their participation. We confirmed stereoscopic vision of all participants via anaglyphic random-dot stereograms before the experiment. All of our participants had normal or corrected to normal vision. 2 participants wore glasses and 7 participants wore contact lenses during the experiment. We confirmed 20/20 visual acuity of all participants with a vision test based on a Snellen chart before the experiment. None of our participants reported a disorder of equilibrium. Two of our participants reported a slight red-green weakness. No other vision disorders have been reported by our participants. 10 participants had prior experience with HMDs and 8 of them had participated in an experiment involving HMDs before. All but two participants reported experience with 3D video games, and 13 participants reported experience with 3D stereoscopic cinema. The eye height of our participants ranged between 1.59−1.84m (M=1.70m, SD=0.06m) above the ground. We used this value to adjust the height of target markers in the VE as shown in Figure 6.1. We measured the interpupillary distances (IPDs) of our participants before the experiment [WGTCR08]. The IPDs of our

Figure 6.1: Experiment setup: participant walking with an HMD towards a virtual target displayed at eye height. The inset shows the participant's view in the virtual replica of the laboratory.

participants ranged between 5.9−6.8cm (M=6.38cm, SD=0.24cm). We used the IPD of each participant to provide a correct perspective on the HMD. All participants were naïve to the experimental conditions. The order of the distance and speed experiments was randomized and counterbalanced between participants. The total time per participant, including pre-questionnaires, instructions, experiment, breaks, post-questionnaires, and debriefing, was 2 hours. Participants wore the HMD for approximately 1 hour. They were allowed to take breaks at any time; short breaks after every 30 experiment trials were mandatory.

6.3.2 General Material

We performed the experiment in an 8m×14m darkened laboratory room. As illustrated in Figure 6.1, participants wore an Oculus Rift DK1 HMD for the stimulus presentation, which provides a resolution of 640×800 pixels per eye with a refresh rate of 60Hz and an approximately 110° diagonal field of view. We attached an active infrared marker to the HMD and tracked its position within the laboratory with a WorldViz Precision Position Tracking PPT-X4 active optical tracking system at an update rate of 60Hz. The head orientation was tracked with the inertial orientation tracker of the Oculus Rift HMD. We compensated for inertial orientation drift by incorporating the PPT optical heading plugin to improve the tracker output. The visual stimulus consisted of a virtual replica of our laboratory that we modeled with centimeter accuracy and textured with photos matching the look of the real laboratory (see Figure 6.1). For rendering, system control and logging we used an Intel computer with 3.4GHz Core i7 processor, 8GB of main memory and Nvidia Quadro 4000 graphics card. The stimuli were rendered with the Unity 3D Pro engine. In order to focus participants on the task, no communication between experimenter and participant was performed during

the experiment. Task instructions were presented via slides on the HMD. Participants judged their perceived self-motions via button presses on a Nintendo Wii remote controller.

6.3.3 General Methods

In the experiments, we make use of two-alternative forced choice tasks (2-AFCT) to determine effects of gains on percepts in IVEs [Fer08]. In this method, participants are asked to move to align an eccentric visual target with the body. During the movement, motion gains $g \in \mathbb{R}_0^+$ between the movement of the participant and the sensory motion feedback are varied between trials in a within-subjects design. If $g = 1$, the sensory feedback matches the participant's movement. For other gains, the sensory feedback is increased ($g > 1$) or decreased ($g < 1$) compared to the movement.

After the movement, the participant has to judge whether characteristics of the perceived motion were *slower* or *faster* / *longer* or *shorter* / *smaller* or *larger* than those of the participant's movement in a 2-AFCT [Fer08]. In these experiments, we use an adaptive staircase design, which starts with a discrepancy which is easy to detect [Cor62, Lee01]. For each following trial, the new gain is computed by the previous gain plus or minus a fixed step width (1-up-1-down) depending on the answer to the aforementioned 2-AFCT question. To eliminate response bias, we interleaved two staircases starting from a minimum and maximum gain [MC04]. By iterative refinement the interleaved staircase design converges on the point of subjective equality (PSE), i. e., the gain at which participants judge the virtual motion as identical to the physical movement, which allows us to identify possible systematic over- or underestimation (i. e., $g > 1$ or $g < 1$). All participants had to complete 30 trials in total. In the interleaved staircase design we started with a minimum and maximum gain of $g = 0.4$ and $g = 1.6$ and a step width of 0.2. All participants were able to identify the sensory discrepancies in the 2-AFCT experiment at these start gains. After two turns in response behavior (cf. [Lee01, MC04]) we halved the step width and computed the PSE as the mean of the last 10 trials.

6.3.4 Experiment E1: Distance Estimation

In this section, we describe the experiment that we conducted to evaluate the perception of walking distances in the IVE using a novel 2-AFCT approach for distance judgments.

Methods

The trials started with an initial stimulus phase in which participants were presented with a virtual view of the replica model of our laboratory on the HMD. As shown in Figure 6.1 we placed a marker at eye height at a target distance of 3m, 4m, or 5m from the participant's start position. After participants felt positive of having memorized the target distance, they pressed a button on the Wii remote controller, and the view on the HMD went black.

With no visual feedback of their self-motion in the VE, participants then walked in the direction of the previously seen target marker. Participants walked a physical distance that

was scaled relative to the virtual target distance. Therefore, we apply distance gains $g_d \in \mathbb{R}_0^+$ to determine a participant's physical walking distance in each trial, describing the relation between the physical walking distance and virtual target distance. For $g_d = 1$ physical Euclidean walking distances along the floor plane $D_p \in \mathbb{R}_0^+$ match one-to-one the virtual target distance $D_v \in \{3\text{m}, 4\text{m}, 5\text{m}\}$ with $D_p = D_v \cdot g_d$. In contrast, gains $g_d < 1$ result in participants walking shorter and $g_d > 1$ in longer distances in the real world. After walking the scaled physical distance, participants were asked to answer a 2-AFCT question: "Did you move *farther* or *shorter* than the virtual target distance?" The participants had to choose one of the two alternatives by pressing the up or down button on the Wii remote controller. If a participant cannot reliably discriminate between the virtual and real distance, the participant must guess, and will be correct on average in 50% of the trials. We applied gains in the range between $g_d = 0.4$ and $g_d = 1.6$, i. e., the walked distance differed by up to ±60% from the target distance.

After answering the question, participants were guided back to the start position of the next trial. Participants did not receive feedback about their walked distance. The gains converged on the PSE following the interleaved staircase approach described in Section 6.3.3. We measured a PSE for each participant for each of the three target distances, i. e., the gains at which the participants judge the physical and virtual distances as identical. The order in which the target distances were tested for each participant was given by Latin squares.

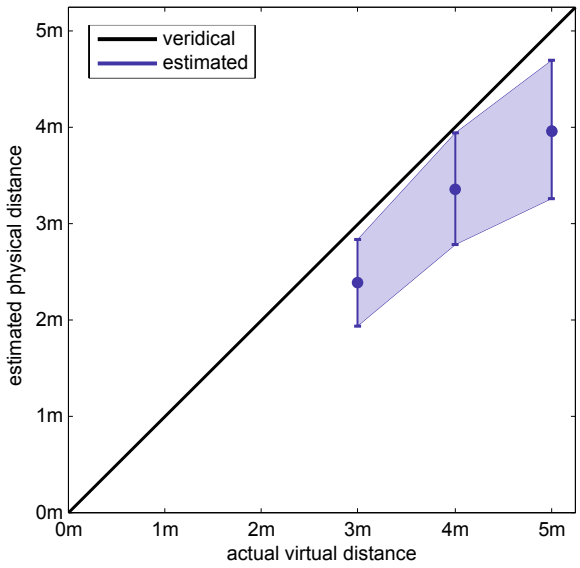


Figure 6.2: Results of the experiment for distance judgments. The horizontal axes show the actual motion components and the vertical axes show the self-motion judgments. The vertical bars and colored regions illustrate standard deviations of self-motion judgments.

Results

Figure 6.2 shows the results for the three tested within-subjects target distances. The x-axis shows the target distances, and the y-axis shows the mean judged PSEs of the participants. The vertical bars and colored regions illustrate the standard deviations.

Participants judged a physical distance as matching the virtual target distance that was approximately 2.39m (g_d=0.796, 20.4% decreased) for a target distance of 3m, 3.36m (g_d=0.839, 16.1% decreased) for a target distance of 4m, and 3.98m (g_d=0.795, 20.5% decreased) for a target distance of 5m. Over all target distances, participants judged an approximately 19.0% decreased physical distance as identical to the virtual distance. Considering the PSEs of individual participants, 15 of the 18 participants consistently judged a decreased physical distance to match the virtual distance, whereas 1 participant consistently judged an increased physical distance as correct, and 2 participants made mixed judgments.

Table 6.1 lists the judged PSEs and means over all participants for the three virtual target distances. We analyzed the results with a repeated measure ANOVA and paired-samples t-test. The results were normally distributed according to a Shapiro-Wilk test at the 5% level. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(2)$=9.44, p<.01), therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ϵ=.36). We found a significant difference between the judged PSEs over all participants and veridical results with $g_d = 1$ (t(17)=−6.25, p<.001). We found no significant main effect of target distance on the judged PSEs (F(1.384, 23.52)=1.636, p<.22, η_p^2=.088).

6.3.5 Experiment E2: Speed Estimation

In this section, we describe the experiment that we conducted to evaluate the perception of self-motion speed in the IVE.

Methods

The trials started with an initial startup phase in which participants had to increase their walking speed to either 1m/s, 1.25m/s, or 1.5m/s. During this phase, participants adjusted their self-motion to a target speed using a speedometer that was indicated on the HMD while no self-motion in the VE was shown. A walking speed of ca. 1.25m/s correlates to an average walking speed of HMD users according to Mohler et al. [MCWB07]. Accordingly, the speed of 1m/s correlates to slow walking, and 1.5m/s correlates to a brisk walking pace. If participants were unable to walk at a steady pace for at least 1m with the target speed (±0.1m/s) after an initial 2m straight distance the trial was repeated. Less than 10% of all trials had to be repeated.

After the initial startup distance, participants had to continue walking with the assumed pace for a duration of 3 seconds towards a virtual target marker that we displayed at eye height at 10m distance. During this time, participants received visual feedback of their self-motion in the virtual replica of the laboratory, which we scaled relative to their physical motion. While a user moves in the real world, applying gains to a user's virtual self-motion corresponds to

changes in the mapping from physical to virtual self-motion speed. Such differences were implemented based on user-centric coordinates as introduced by Steinicke et al. [SBJ+10]. Virtual self-motions can be scaled with speed gains $g_s \in \mathbb{R}_0^+$, describing the mapping from movements in the real world to motions of a head-referenced virtual camera. For $g_s = 1$ physical translations $T_p \in \mathbb{R}^3$ are mapped one-to-one to virtual translations $T_v \in \mathbb{R}^3$ with $T_v = T_p \cdot g_s$, i. e., the virtual scene remains stable considering the physical position change, whereas $g_s < 1$ result in slower and $g_s > 1$ in faster virtual self-motion speed. We applied gains in the range between $g_s = 0.4$ and $g_s = 1.6$ (cf. Section 6.3.3), i. e., the virtual speed differed by up to ±60% from the physical speed.

Following the stimulus phase, the participants were asked to answer a 2-AFCT question: "Did you move *faster* or *slower* in the virtual world than in the real world?" The participants had to choose one of the two alternatives by pressing the up or down button on the Wii remote controller. If participants cannot reliably discriminate between the virtual and real speed, they have to guess, and will be correct on average in 50% of the trials. After answering the question, participants were guided back to the start position of the next trial. The gains converged on the PSE following the interleaved staircase approach described in Section 6.3.3. We measured a PSE, i. e., the gain on which the interleaved staircase approach converged, for each participant for each of the three physical speeds. The order in which the physical speeds were tested was given by Latin squares.

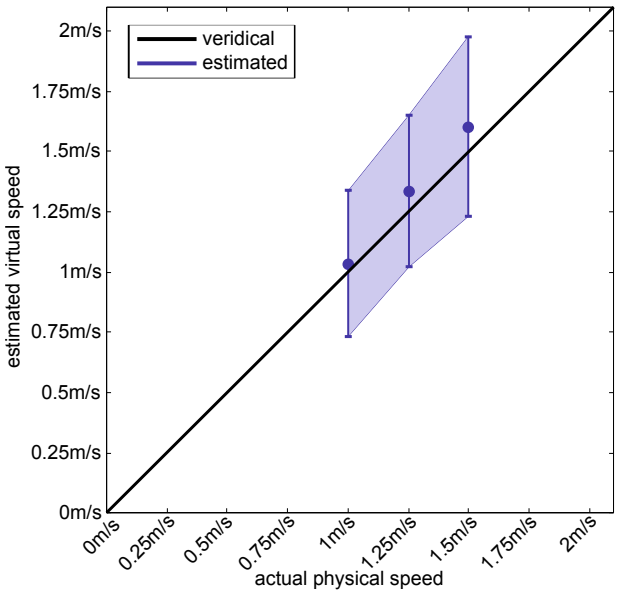


Figure 6.3: Results of the experiment for speed judgments. The horizontal axes show the actual motion components and the vertical axes show the self-motion judgments. The vertical bars and colored regions illustrate standard deviations of self-motion judgments.

Results

Figure 6.3 shows the results for the three tested within-subjects walking speeds. The x-axis shows the physical walking speed, and the y-axis shows the mean virtual walking speed that the participants judged as equivalent to their physical motion. The vertical bars and colored regions illustrate the standard deviations.

Participants judged a virtual self-motion speed as matching their physical walking speed that was approximately 1.03m/s (g_s=1.034, 3.4% increased) for a physical walking speed of 1m/s, 1.34m/s (g_s=1.069, 6.9% increased) for a physical walking speed of 1.25m/s, and 1.60m/s (g_s=1.067, 6.7% increased) for a physical walking speed of 1.5m/s. Over all physical walking speeds, participants judged an approximately 5.7% increased virtual speed as identical to their physical motion, which corresponds to an underestimation of virtual speeds. Considering the PSEs of individual participants, 9 of the 18 participants consistently judged increased virtual speeds as identical with their physical speed, whereas 6 participants consistently judged decreased virtual speeds as correct, and 3 participants made mixed judgments.

Table 6.1 lists the judged PSEs and means over all participants for the three physical walking speeds. We analyzed the results with a repeated measure ANOVA and paired-samples t-test. The results were normally distributed according to a Shapiro-Wilk test at the 5% level. The sphericity assumption was supported by Mauchly's test of sphericity at the 5% level. We did not find a significant difference between the judged PSEs over all participants and veridical results with $g_s = 1$ ($t(17)$=1.03, p<.32). We found no significant main effect of physical walking speed on the judged PSEs ($F(2, 34)$=.275, p<.77, η_p^2=.016).

6.3.6 Experiment E3: Time Estimation

In this section, we describe the experiment that we conducted to evaluate the perception of time in the IVE.

Methods

Since there is no clear notion of a discrepancy between elapsed time in the real and virtual world that could be compared simultaneously with a 2-AFCT, we measured values for time estimation separately in the IVE (i. e., with HMD) and in the real world (i. e., without HMD). The trials in the IVE started with an initial stimulus phase in which participants were presented with a view on the HMD to the virtual replica model of our laboratory. When they felt ready to start, participants pressed a button on the Wii remote controller, which caused a short acoustic signal to be heard, and then walked in the direction of a virtual target marker that we displayed at eye height at 10m distance. A second short acoustic signal was displayed after participants had walked over a time interval that we varied between experiment trials. Thereafter, participants answered a 2-AFCT question and were guided back to the start position of the next trial. We replicated the same setup for the experiment trials in the real world without the HMD.

For each trial, participants walked over a time interval that was scaled relative to a reference

time. Therefore, we apply time gains $g_t \in \mathbb{R}_0^+$ to determine a participant's walking interval in each trial, describing the relation between the trial's interval and the reference time. For $g_t = 1$ the trial's interval $T_v \in \mathbb{R}_0^+$ matches the reference time $T_p \in \{2\text{s}, 3\text{s}, 4\text{s}, 5\text{s}\}$ with $T_v = T_p \cdot g_t$. In contrast, gains $g_t < 1$ result in participants walking over a shorter interval and $g_t > 1$ in a longer interval relative to the reference time. We chose the reference times based on the common durations in the distance and speed experiments. After walking the trial's interval, participants were asked to answer a 2-AFCT question: "Did you move *longer* or *shorter* than # seconds?" with the # replaced by the corresponding reference time. The participants had to choose one of the two alternatives by pressing the up or down button on a Wii remote controller. If a participant cannot reliably discriminate between the elapsed and reference time, the participant must guess, and will be correct on average in 50% of the trials. We applied gains in the range between $g_t = 0.4$ and $g_t = 1.6$ (cf. Section 6.3.3), i. e., the walked interval differed by up to ±60% from the reference time.

After answering the question, participants were guided back to the start position of the next trial, i. e., they did not receive feedback about the elapsed time. The gains converged on the PSE following the interleaved staircase approach described in Section 6.3.3. We measured a PSE for each participant for each of the reference times, i. e., the gains at which the participants judge the elapsed and reference times as identical. PSEs that deviate from $g_t = 1$ indicate subjective time dilation or expansion for the different reference times. The order in which the reference times were tested was randomized for each participant.

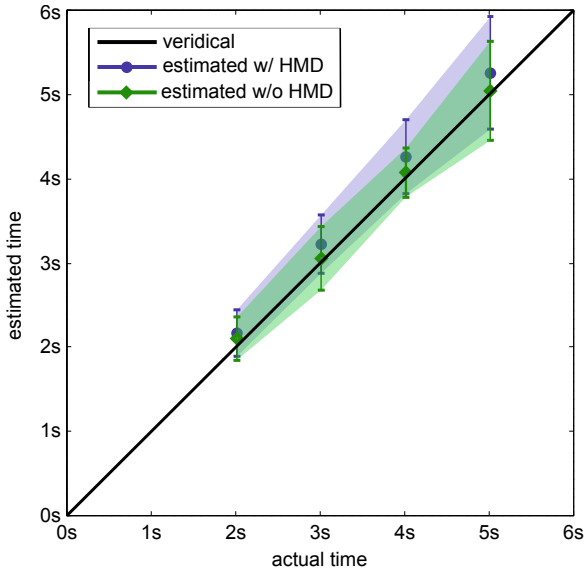


Figure 6.4: Results of the experiment for time judgments. The horizontal axes show the actual motion components and the vertical axes show the self-motion judgments. The vertical bars and colored regions illustrate standard deviations of self-motion judgments.

Results

Figure 6.4 shows the results for the tested within-subjects reference times. The x-axis shows the actual reference times, and the y-axis shows the judged times with and without HMD. The vertical bars and colored regions illustrate the standard deviations.

In the real world without using the HMD participants judged an elapsed time as matching the reference time that was approximately 2.10s (g_t=1.051, 5.1% longer) for a reference time of 2s, 3.05s (g_t=1.017, 1.7% longer) for a reference time of 3s, 4.06s (g_t=1.016, 1.6% longer) for a reference time of 4s, and 5.04s (g_t=1.007, 0.7% longer) for a reference time of 5s. In contrast, for the experiment trials in the IVE, participants judged an elapsed time as matching the reference time that was approximately 2.16s (g_t=1.081, 8.1% longer) for a reference time of 2s, 3.22s (g_t=1.073, 7.3% longer) for a reference time of 3s, 4.24s (g_t=1.061, 6.1% longer) for a reference time of 4s, and 5.23s (g_t=1.046, 4.6% longer) for a reference time of 5s.

Over all reference times, participants judged an elapsed time as identical to the reference time that was approximately 6.5% increased (g_t=1.065) in the IVE, and approximately 2.3% increased (g_t=1.023) in the real world. The difference between the PSEs in the real world condition and the IVE suggest that times were overestimated in the IVE by approximately 4.2% compared to estimates in the real world condition. Considering the PSEs of individual participants, 7 of the 10 participants judged a longer elapsed time in the IVE compared to the real world, whereas 3 participants judged a shorter elapsed time.

We analyzed the results with a repeated measure ANOVA and paired-samples t-tests. The

Table 6.1: PSEs of all participants for the tested speeds and distances.

subject	physical speed			virtual distance		
	1m/s	1.25m/s	1.5m/s	3m	4m	5m
s01	0.49	0.72	0.89	0.80	0.72	0.64
s02	1.39	1.37	0.84	0.72	0.88	0.73
s03	1.30	1.31	1.27	0.91	0.81	0.71
s04	1.43	0.87	0.91	0.81	0.92	0.89
s05	0.56	0.93	0.91	1.18	1.17	1.13
s06	1.19	1.22	1.07	0.69	0.70	0.71
s07	0.80	0.88	0.85	0.65	0.68	0.58
s08	1.21	1.41	1.50	0.71	0.99	0.93
s09	0.77	0.74	1.15	0.69	0.78	0.79
s10	0.87	0.89	0.98	0.72	0.93	0.97
s11	1.36	1.42	1.45	0.52	0.62	0.65
s12	1.11	1.33	1.32	1.04	0.95	0.77
s13	1.15	1.28	1.28	0.91	0.80	0.73
s14	1.13	1.13	1.27	0.82	0.77	0.68
s15	0.83	0.80	0.73	0.83	0.93	1.00
s16	1.35	1.09	1.08	0.85	1.02	0.87
s17	0.61	0.75	0.61	0.75	0.75	0.77
s18	1.07	1.11	1.09	0.73	0.68	0.76
mean	1.034	1.069	1.067	0.796	0.839	0.795

results were normally distributed according to a Shapiro-Wilk test at the 5% level. The sphericity assumption was supported by Mauchly's test of sphericity at the 5% level. We found a significant main effect of reference time on the judged PSEs in the real world ($F(3, 27)=224.65$, $p<.001$, $\eta_p^2=.961$) and in the IVE ($F(3, 27)=169.14$, $p<.001$, $\eta_p^2=.949$). We observed a trend for a difference between the judged PSEs in the real and virtual world ($t(9)=-1.98$, $p<.08$).

6.4 Discussion

The distance estimation experiment confirmed the common notion in the literature on spatial perception in IVEs that participants tend to underestimate target distances (cf. [IAR06, JSSE11, KTCR09, LJH07, LKG+93, LK03, MD05, WCCRT09]). The PSEs suggest that more than 80% of the participants underestimated the visual target distances in the experiment. We found a mean underestimation of approximately 20% over all distances and participants.

The speed estimation experiment showed that on average participants estimated a virtual self-motion speed as correct if it was up-scaled by approximately 5.7% from their physical self-motion. The PSEs suggest that half of the participants underestimated the visual speed cues from the virtual world compared to the proprioceptive-vestibular speed cues from the real world. This notion supports previous experimental results in the literature, which suggested a tendency towards underestimation of virtual walking speed in IVEs [BSD+05, BSWL12, DGS05, SBJ+10].

The time estimation experiment showed that participants were able to estimate time in the real world quite reliably without large errors, but judged times in the IVE were approximately 4.2% increased over the estimates in the real world. This apparent time dilation in IVEs is an interesting observation, which warrants future investigation. In particular, it is an open question which effect hardware and software factors have on time perception, e. g., compared to traditional computer graphics or game environments.

We performed our experiments in a state-of-the-art IVE with an Oculus Rift HMD and a large real walking space. In this IVE we found different magnitudes of biases in self-motion estimates regarding the three motion components. Pooled over all participants we observed a large underestimation of virtual distances, slight underestimation of virtual speed, and a slight overestimation of time. Figure 6.5 shows that the results for the motion components do not have to adhere to their mathematical relation described in Section 6.1. Further research is necessary to triangulate how this *self-motion perception triple* looks in different IVEs considering that other researchers found different magnitudes of biases as described in Section 6.2. Moreover, Figure 6.6 shows that the self-motion perception triples of individual users may differ from the means computed for a specific IVE, which should be carefully considered. In particular, the results show that all participants are rather accurate in judging elapsed time, whereas individual differences in distance and speed judgments can be observed. In particular, nine out of ten participants underestimate distances (except for the participant whose results are depicted in Figure 6.6(c)). Furthermore, the results highlight that six par-

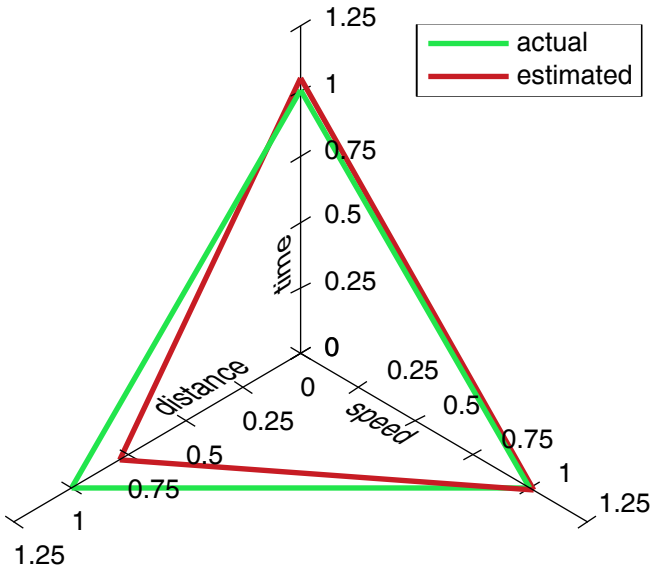


Figure 6.5: Self-motion perception triple: Plot of interrelations between the actual relations and estimated PSEs for speed, distance, and time pooled over all participants who completed all three experiments. In the pooled results a significant distance underestimation can be observed.

ticipants underestimate speed (cf. Figure 6.6(b),(f)-(j)), but that participants in case they overestimate speed, tend to highly overestimate speed (cf. Figure 6.6(a),(c),(d),(e)). In future experiments, such correlations may be considered in more detail.

It is important to determine the magnitude of common misperception in IVEs, in particular with current-state HMDs like the Oculus Rift. Such misperception effects greatly limit the applications of VR technologies in domains that require spatial perception that matches the real world. However, it is a challenging task to remedy the causes of these effects. In contrast, different approaches may be used to compensate for misperception by alleviating the effects in one of the motion components. Examples are magnification based on the field of view [KTCR09], increased optic flow [BSWL12], up-scaled travel distances [IRA07] etc. In particular, our PSEs suggest that distances and speed may be scaled such that users judge them as veridical, and exogenous cues may be scaled to provide a more accurate time perception. However, while these approaches provide advantages in the scope of a motion component, and may even lead to accurate impressions of self-motion, potential disadvantages may arise from side-effects of manipulations on the other motion components and cognitive processes, which should be considered.

6.5 Conclusion

In this chapter, we have analyzed the triple of self-motion speed, distance, and time perception in an IVE using the same setup and similar protocols. In psychophysical experiments in an Oculus Rift IVE, we measured PSEs indicating a bias in the three components of self-motion perception. The results show that virtual walking distances on average were underestimated

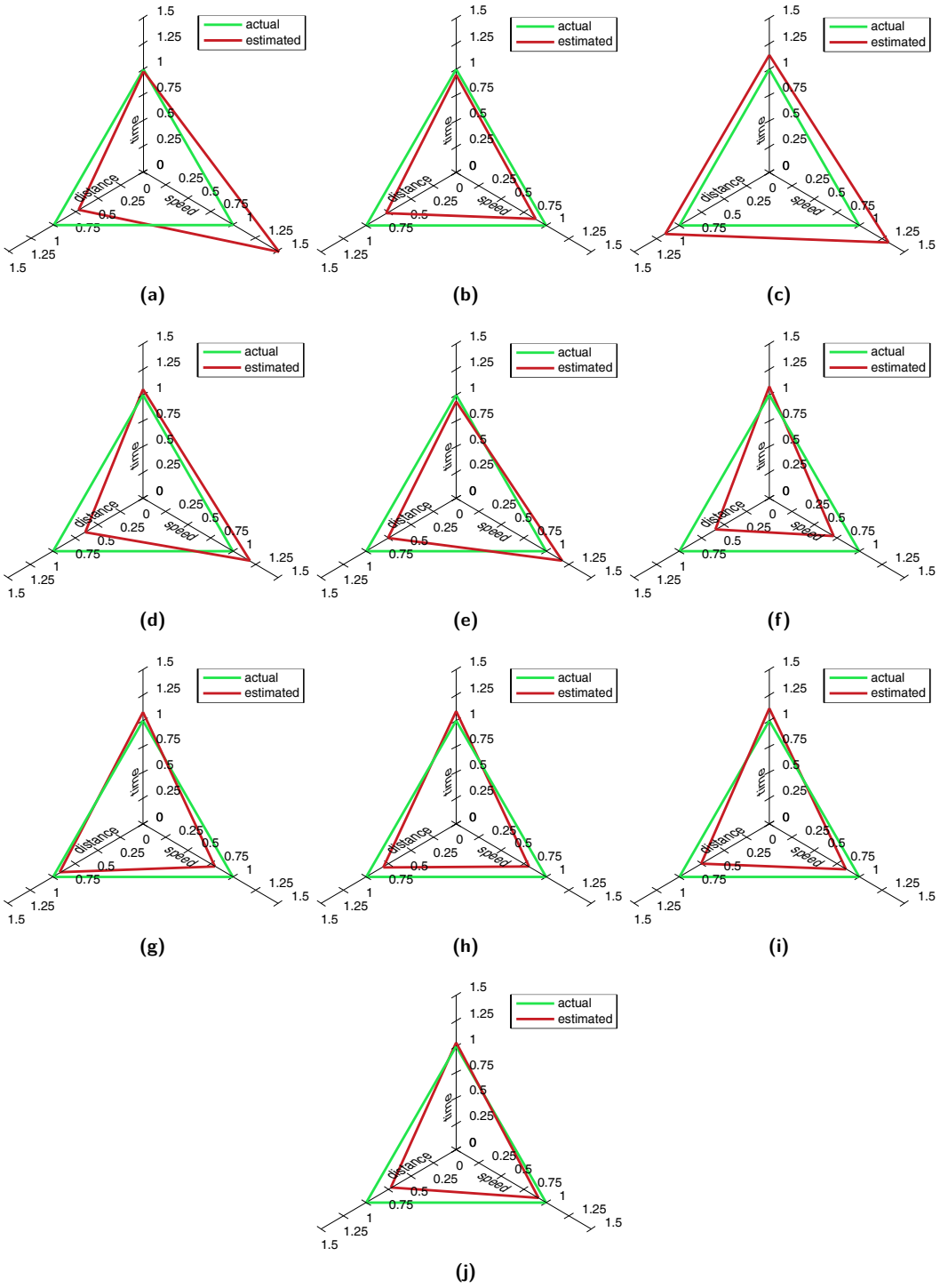


Figure 6.6: Individual self-motion perception triples for those participants who participated in all three experiments.

by 19.0%, virtual speeds on average were underestimated by 5.7%, and time was overestimated by 4.2% in the IVE.

As illustrated in Figure 6.5, differences in this self-motion triple indicate perceptual biases, which may be effected by VR hard- and software or individual differences, and should be triangulated over different IVEs and user groups. More research is necessary to understand the reasons, interrelations, and implications of such perceptual biases introduced by VR technologies in IVEs.

Part IV

Spatial Perception with Stereoscopic Displays

Chapter 7

7 Analyzing Effects of Geometric Rendering Parameters on Size and Distance Estimation in On-Axis Stereographics

Accurate perception of size and distance in a three-dimensional virtual environment is important for many applications. However, several experiments have revealed that spatial perception of virtual environments often deviates from the real world, even when the virtual scene is modeled as an accurate replica of a familiar physical environment. While previous research has elucidated various factors that can facilitate perceptual shifts, the effects of geometric rendering parameters on spatial cues are still not well understood.

In this chapter, we model and evaluate effects of spatial transformations caused by variations of the geometric field of view and the interpupillary distance in on-axis stereographic display environments. We evaluated different predictions in a psychophysical experiment in which participants were asked to judge distance and size properties of virtual objects placed in a realistic virtual scene. Our results suggest that variations in the geometric field of view have a strong influence on distance judgments, whereas variations in the geometric interpupillary distance mainly affect size judgments.

7.1 Introduction

Modern virtual reality (VR) display technologies enable users to experience a three-dimensional virtual environment (VE) from an egocentric perspective. Such immersive viewing experiences have an enormous potential for a variety of application domains, in which accurate spatial perception during design, exploration or review of virtual models and scenes is required. Head-mounted displays (HMDs) and immersive projection technologies are often used in order to provide a user with near-natural feedback of virtual content, as if the user

is present in the virtual scene. In particular, modern real-time rendering systems can create compelling immersive experiences offering most of the spatial visual cues we can find in real-world views, including perspective, interposition, lighting and shadows [TFCRS11]. However, distance and size perception are often biased in such environments, causing users to over- or underestimate spatial relations in virtual scenes to a much higher magnitude than can be observed in similar scenes in the real world [LK03, TWG+04, IRLA07].

Obviously, issues with visual rendering have been suggested as a potential source for biased spatial perception. In order for a virtual scene to be displayed stereoscopically on a binocular HMD, the computer graphics system must determine which part of the scene should be displayed where on the two screens. In 3D computer graphics, planar geometric projections are typically applied, which make use of a straightforward mapping of graphical entities within a 3D 'view' region, i. e., the *view frustum*, to a 2D image plane. During the rendering process, objects inside the view frustum are projected onto the 2D image plane; objects outside the view frustum are omitted. The exact shape of each view frustum in on-axis stereographic displays (as used in many HMDs) is a symmetric truncated rectangular pyramid. The opening angle at the top of the pyramid, often denoted as *geometric field of view* (GFOV) [MRE85], should match the *display's field of view* (DFOV) for the imagery to be projected in a geometrically correct way [SBK11]. Another important characteristic of the human visual system is the *interpupillary distance* (IPD), which describes the horizontal separation of the eyes that ranges from 5.77cm to 6.96cm (median: 6.32cm) in adult males (according to Woodson [Woo81]). Since the viewpoints of both eyes are horizontally separated, each eye receives a slightly different retinal image. The brain interprets the binocular inputs and fuses the images, resulting in the impression of a solid space and the perception of depth [Cut97]. Typically, the user's IPD is measured and then applied to the *geometric interpupillary distance* (GIPD) used for stereoscopic rendering, assuming that the HMD's display units are correctly adjusted in front of the user's eyes.

Both geometric rendering parameters, GFOV and GIPD, have to be defined in all on-axis 3D stereoscopic visualization systems. At the same time, they are particularly prone to calibration errors and therefore bear a high risk of accidentally skewing a user's perception in immersive VEs. Common sources for such errors are naïvely applying manufacturer specifications (e. g., the FOV of built-in displays in head-mounted devices [KTCR09, KBB+12]) without verification of the physical display characteristics, or by using population means to approximate a user's IPD. Although slight errors in such rendering parameters are quite common in VR environments, it is still not entirely clear as to what effects these discrepancies have on distance and size cues. Moreover, it has been found that when users have direct control over a rendering parameter, they often try to use it to compensate for given perceptual biases that may have been introduced by miscalibration of other parameters [SBL+11]. It remains a challenging question how rendering parameters are related regarding particular cues, and if they could be used to address perceptual biases.

The motivation of this work is to compare mathematical models for selected cues that are dominated by the two rendering parameters GFOV and GIPD in terms of their mutual

influence on size and distance perception in realistic VEs [TWG+04, IRLA07, WCCRT09]. We describe the effects both rendering parameters have on the theoretical models, and then compare the predictions to relative size and distance judgments collected in a psychophysical experiment.

The chapter is structured as follows: Section 7.2 provides background information about spatial perception and the cues that are primarily affected by the rendering parameters GFOV and GIPD. In Section 7.3, we describe how the considered cue models are influenced by changes in both stereoscopic rendering parameters. In Section 7.4, we describe a psychophysical experiment in which we evaluate how participant responses correlate with the predictions of these models when varying the rendering parameters, and discuss the results. Section 7.5 concludes the chapter.

7.2 Background

In this section, we summarize background information about selected mathematical models for size and distance cues in the scope of the two stereoscopic rendering parameters GFOV and GIPD.

7.2.1 Egocentric Perspective

Human visual perception of distance, size and spatial relationships relies on prior knowledge (i. e., our lifelong experience in the physical world), individual preferences and neurophysical properties, as well as on given visual stimuli. Among the most important of these are perspective cues [Gol09], including object retinal size scaled by distance, object distance as a function of relative position to the ground plane and sky (both of which extend toward a visible horizon), convergence of parallel lines in vanishing points, and many more [TFCRS11].

Emmert's Law [Lou07] provides a simple approximation of the inherent assumptions of the perceptual system regarding size, distance, and retinal size. Emmert observed that afterimages, although having a constant retinal size, appeared to be larger if the viewed background was farther away. In other words, for the same retinal size, an object's perceived size will depend on its perceived distance. This phenomenon is illustrated in Figure 7.1(a). In order to judge an object's size, the visual system needs to evaluate the object's distance from the viewpoint or vice versa. The perceived size $\tilde{S_0} \in \mathbb{R}^+$ of an object with linear size S_0 is proportional to the product of its perceived distance $\tilde{D_0} \in \mathbb{R}^+$ (D_0 is linear distance) and angular size $\Theta \in \mathbb{R}^+$, and can be expressed in simple form as

$$\tilde{S_0} \sim \tilde{D_0} \cdot \tan\left(\frac{\Theta}{2}\right). \tag{7.1}$$

The phenomenon of perceived-size constancy [Gil51] denotes the tendency of an object to maintain the same perceived size, even if its retinal size changes, i. e., if the object moves farther away or approaches the observer. In such cases, e. g., for objects with familiar size,

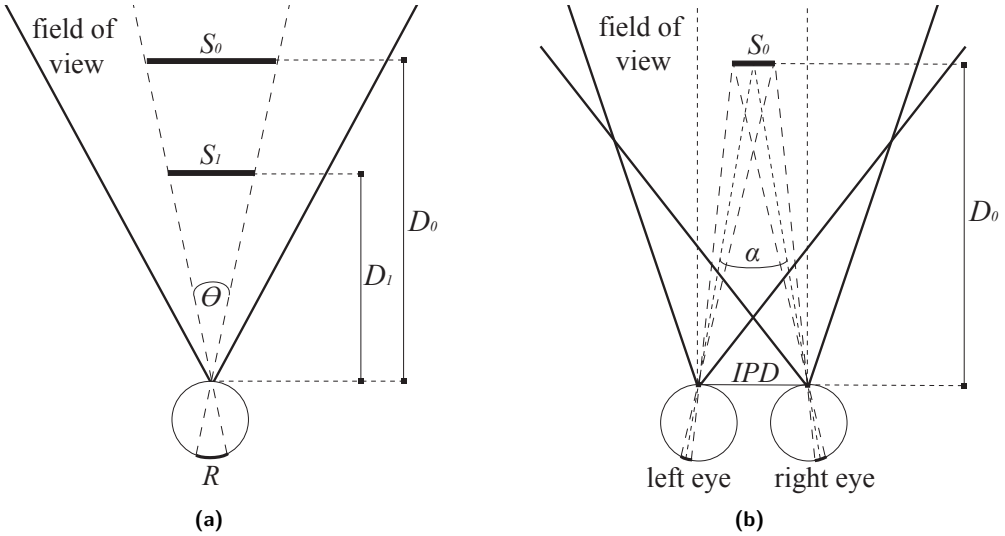


Figure 7.1: Illustration of (a) visual size-distance ambiguity with object size S_i being a linear function of the distance D_i in case of a constant retinal size R with corresponding angular size Θ, and (b) object size and distance as a function of eye convergence.

the perceived distance of an object can be described using the above relation. This size-distance-invariance hypothesis can describe many properties of size and distance perception, including cases of apparent misperception. However, Murray et al [MBK06] have recently replaced the retinal size in the relation by the perceived retinal size. This updated hypothesis states that the ratio of perceived linear size to perceived distance is not necessarily a simple function of the visual angle, but that the visual angle can be subject to perceptual biases as well. These findings underline that the controversial size-distance invariance hypothesis and some other aspects of visual perception are still not well understood [MT99].

7.2.2 Stereopsis

While the visual system faces an inherent ambiguity of size and distance in case of monocular vision (given unfamiliar objects or objects of uncertain size), the binocular configuration of the eyes provides humans with absolute stereoscopic cues. As mentioned in Section 7.1, since the eyes are horizontally separated by an interpupillary distance, the brain receives two slightly different images of a scene. The perceptual system can then make use of these cues, in particular, of binocular disparity and convergence. The former refers to differences in the retinal image locations when light from an object is projected into the left and right eye of an observer. Solving the disparity correspondence problem, the brain relates retinal image contents from the two eyes to one another, allowing the perceptual system to compute the distance of seen objects using simple triangulation (see Figure 7.1(b)). Further, when focusing on an object, the eyes need to rotate toward that object to bring it to the fovea of each retina. Turning the eyes inward when fixating on a closer object leads to larger convergence angles. The convergence state of the eyes, changed by extrinsic muscle exertion,

thus provides an absolute cue about the distance of an object from an observer. In a simplified setting, the distance $D_0 \in \mathbb{R}^+$ of an object can be computed from the convergence angle using the following equation:

$$D_0 = \frac{\text{IPD}}{2 \cdot \tan(\frac{\alpha}{2})}, \tag{7.2}$$

with the user's IPD and the convergence angle $\alpha \in \mathbb{R}^+$. Figure 7.1(b) shows how size and distance ambiguities can be resolved with binocular viewing.

Stereoacuity is naturally limited [SSG00]. Researchers assume a conservative threshold of about 10 seconds of arc (approx. 0.003 degrees) [PGG+10] for stereoacuity. Referring to Equation 7.2, the maximum distance at which stereopsis may produce usable data would be approx. 1.24km. Beyond this distance, the human visual system cannot sufficiently differentiate binocular information. However, in VR environments the angular resolution of a pixel on a display surface may act as an artificial cut-off to the capabilities of human natural vision. That is, although the retina may be able to respond to much smaller visual stimuli, the size of the smallest digital element of a visual display may prevent this capacity from being exploited. To which degree stereopsis effectively contributes to size or distance judgments in immersive VEs depends on various factors, including user characteristics, virtual scene, display technology, and many more. Cue reliabilities and the resulting weights during cue integration ultimately determine the impact of stereopsis [Ern06]. The effects of stereoscopic display and IPD on distance judgments in realistic scenes have been found less important than predicted by Equation 7.2 [WGTCR08, CRWGT05].

7.3 Cue Conflicts in On-Axis Stereographics

In this section, we discuss the relationship between the field of view and interpupillary distance in *on-axis* stereoscopic rendering configurations, which are used in many HMD settings. In contrast, projection screens use *off-axis* stereoscopic rendering, which accounts for the situation that only one display surface exists to present the left and right eye views. In the simplest case of an on-axis binocular display design, the display surfaces are oriented orthogonally to the parallel optical axes for both views, which intersect with the display surfaces in their center. If we disregard optical distortions (e. g., pincushion distortion), the binocular configurations of such displays can be illustrated as shown in Figure 7.2. We assume a coordinate system in which the camera is represented by a position, look direction vector, up vector, and strafe vector. As illustrated in Figure 7.2, the frustum of the camera model for HMDs is delimited by the near and far clipping planes, as well as by the size of the virtual image, which appears at the display optics' accommodation distance from the eyes. The size of the display surface defines the horizontal and vertical geometric fields of view[1]. The camera frustums for the left and right eye are separated by the geometric interpupillary distance along the strafe vector.

[1] If not stated otherwise, we refer to the GFOV as the vertical geometric field of view since the horizontal field of view can be computed using the fixed aspect ratio of a display surface.

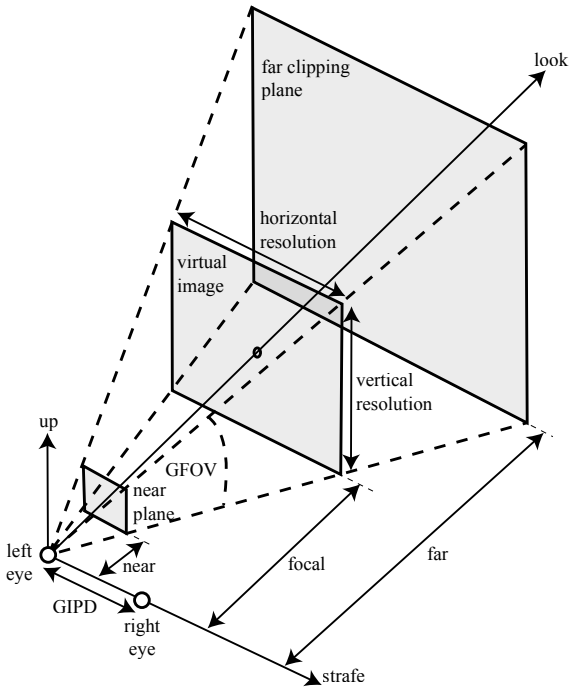


Figure 7.2: Idealized binocular camera model in three-dimensional computer graphics for head-mounted displays [RH94]. For better legibility, only the view frustum for the left eye camera object is shown. The right eye camera frustum results from a translation by GIPD along the strafe vector.

7.3.1 Introducing GFOV and GIPD Conflicts

When displaying computer generated images on a physical display, we have to distinguish between the virtual rendering setup and the physical display setup. In order to provide a view to a virtual scene on a head-referenced display surface that matches what a user would see in a corresponding real-world scene, the computer graphics rendering system has to be calibrated to the physical display characteristics. In particular, the GFOV in the rendering environment has to be set to the visual angle covered by the display units in front of the user's eyes. The interpupillary distance of the user has to be applied to the binocular camera model as shown in Figure 7.2. In case of any discrepancy, various size and distance cues are affected, and can cause perceptual shifts.

Field of View Gains

The DFOV refers to the visual angles subtended by a display unit in front of the user's eyes (cf. Section 7.1), whereas the GFOV refers to the visual angles of the view frustum in the computer graphics rendering model illustrated in Figure 7.2. If GFOV matches DFOV, the virtual space is mapped onto the physical display in such a way that users see a natural perspective. Mapping differences can be described via GFOV gains $g_{_F} \in \mathbb{R}^+$ as

$$\text{GFOV} = g_{_F} \cdot \text{DFOV}. \tag{7.3}$$

If the GFOV differs from the DFOV with $g_F \neq 1$, the retinal size of displayed objects is scaled. For $g_F > 1$, the virtual image is rendered over a larger visual field, but compressed onto a smaller physical display surface. For $g_F < 1$, a smaller visual field is rendered, and up-scaled onto a larger physical display surface.

Interpupillary Distance Gains

Figure 7.2 illustrates the displacement of the eye points in strafe direction by GIPD. Assuming the physical display characteristics are correctly applied to the left and right eye camera frustums, differences between users only occur in the GIPD between the centers of projection of the two cameras. If anthropometric population means are applied as GIPD but deviate from a user's individual IPD, this either results in an increased or decreased GIPD. This relative difference can be described via GIPD gains $g_I \in \mathbb{R}^+$ as

$$\text{GIPD} = g_I \cdot \text{IPD}. \tag{7.4}$$

For $g_I = 1$, GIPD and the user's IPD match, thus providing a natural perspective. Changes to GIPD with $g_I \neq 1$ lead to systematic changes of convergence angles (see Section 7.2.2), but have limited effects on retinal size. Increasing the GIPD over the user's IPD (with $g_I > 1$) results in up-scaled convergence angles, i. e., virtual objects should appear closer, whereas decreasing the GIPD (with $g_I < 1$) results in down-scaled convergence angles. For $g_I = 0$ the left and right images are exactly the same.

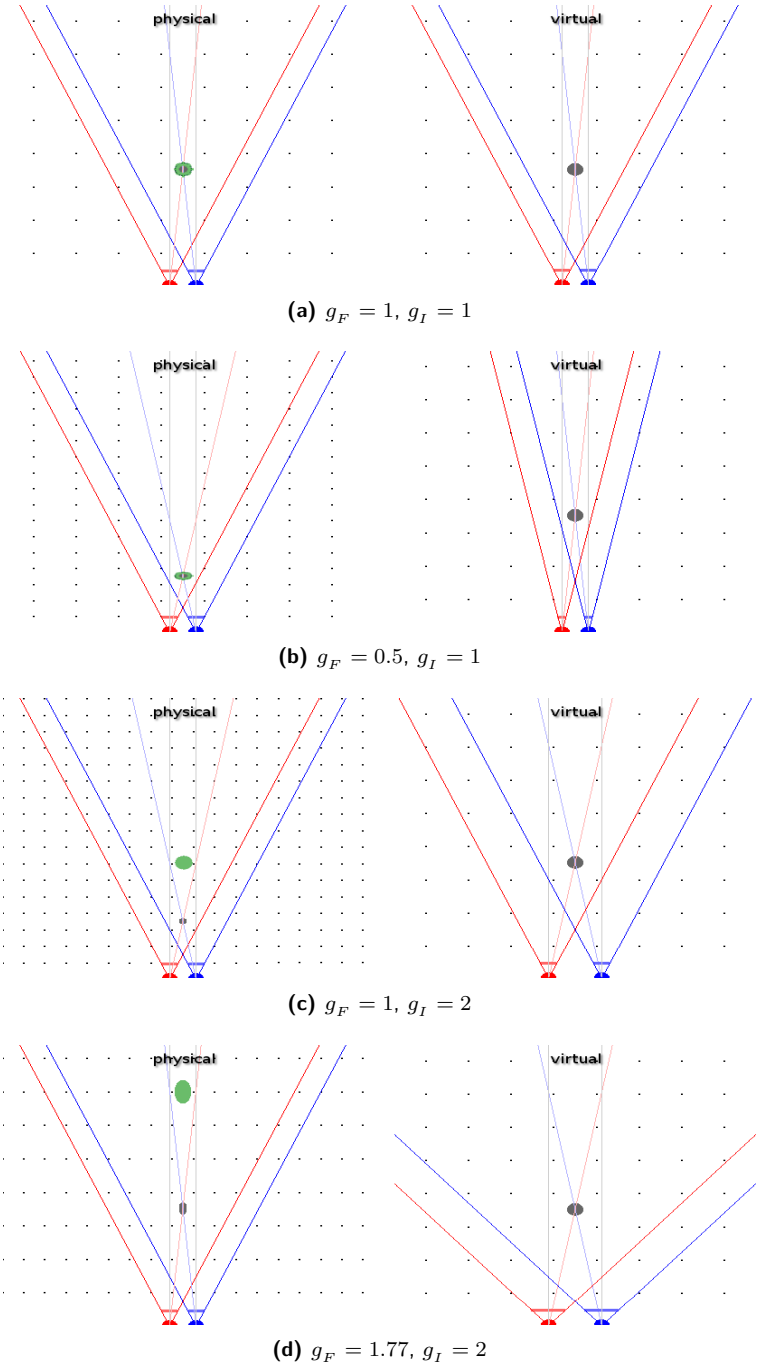
7.3.2 Effects of GFOV and GIPD Conflicts

In the following, we describe effects on distance and size that occur in theory when the geometric field of view and interpupillary distance of the rendering cameras do not comply with the physical configuration. Although virtual renderings often deviate from the physical display properties, it is difficult to estimate how discrepancies may influence the perceived size or distance of objects in binocular displays. In order to compute effects introduced by GFOV and GIPD gains on size and distance cues of stereoscopically displayed objects, we developed an OpenGL simulation for on-axis stereoscopic rendering. In our test environment, we use the following procedure:

1. We render a scene for both eyes using a virtual camera configuration (with GFOV and GIPD gains applied), i. e., the virtual scene is projected onto the image surfaces using planar perspective projections (see Figure 7.3, images in right column).

2. We compute the transformation that is introduced by displaying the planar rendered images for the left and right eye on the physical display surfaces; the FOV and IPD in the physical configuration can differ from the virtual configuration (see Figure 7.3, images in left column).

3a. By testing ray intersections from the eye positions to the positions of objects that are



(a) $g_F = 1$, $g_I = 1$

(b) $g_F = 0.5$, $g_I = 1$

(c) $g_F = 1$, $g_I = 2$

(d) $g_F = 1.77$, $g_I = 2$

Figure 7.3: Representations of cue conflict situations: (a) matching size and distance cues with calibrated GFOV and GIPD, (b) effects of reducing the GFOV by 50%, (c) effects of rendering the virtual view with twice the IPD, and (d) effects of gains $g_F = 1.77$ and $g_I = 2$. We computed the position and size of the gray objects by ray intersection from convergence angles (see Step 3a); the dot pattern illustrates spatial transformations in case of dominance of stereoscopic cues. In contrast, the green objects show the prediction from perceived-size constancy (see Step 3b).

displayed on the left and right physical displays in front of the user's eyes, we compute the object's size and distance from stereopsis as described in Section 7.2.2 (see Figure 7.3, gray objects).

3b. From the retinal size of the object on the physical display surfaces in front of the user's eyes, we compute the object size and distance as predicted solely by perceived-size constancy (see Section 7.2.1), i. e., without considering convergence angles (see Figure 7.3, green objects).

As illustrated in Figure 7.3, changing GFOV and GIPD from the physical configuration introduces perceptual conflicts between the two cues. Figure 7.3(a) shows natural viewing with matching rendering and display configurations, (b) shows effects of decreasing the GFOV by a factor of $g_{_F} = 0.5$ when rendering the virtual view, (c) shows effects of increasing the GIPD by a factor of $g_{_I} = 2$ when rendering the virtual view, and (d) shows effects of a GFOV gain of $g_{_F} = 1.77$ and a GIPD gain of $g_{_I} = 2$. The left eye frustums are shown in red, whereas the right eye frustums are shown in blue. The examples show that applying GFOV and GIPD gains both cause changes to the convergence angles of virtual objects, and thus the size and distance cues that can be derived from stereopsis, whereas only GFOV gains have a strong effect on retinal size. Figure 7.3(d) indicates that if distance cues from stereopsis dominate percepts, it should even be possible to provide a controlled cue conflict stimulus, in which perceptual effects of changing GFOV may be reversed by changing GIPD. However, if retinal size dominates percepts due to perceived-size constancy (see Section 7.2.1), there should be limited to no effect on distance perception by changing GIPD. In the following, we describe the mathematical effects of GFOV and GIPD gains on the cues, before we describe our psychophysical evaluation in Section 7.4.

Field of View Gains

We determined the differences in convergence cues that are introduced by changing GFOV gains. As described in Section 7.2, increasing convergence angles should result in objects being perceived closer to the observer, as well as smaller, whereas decreasing convergence angles should have reverse effects. As illustrated in Figure 7.3(b), a gain $g_{_F} = 0.5$ results in the distance to objects being reduced by 50%, whereas the object size appears non-uniformly scaled. The relation between virtual and displayed object distance for convergence cues can be expressed with scaling factor m_f as

$$\hat{D_0} = D_0 \cdot m_f, \quad m_f := \frac{\tan(g_{_F} \cdot \mathrm{DFOV}/2)}{\tan(\mathrm{DFOV}/2)}, \tag{7.5}$$

with D_0 being the virtual object distance, and $\hat{D_0}$ the resulting object distance from convergence cues when shown on the physical display. The size of an object results as

$$\hat{S_s} = S_s, \quad \hat{S_l} = S_l \cdot m_f, \tag{7.6}$$

with S_s the virtual size in strafe direction, S_l the virtual size in look direction (cp. Figures 7.2 and 7.3), and $\hat{S}_s$ as well as $\hat{S}_l$ the respective object size dimensions that can be derived from convergence cues when shown on the physical display. In particular, the virtual scene should appear non-uniformly scaled along the look direction in front of the user's eyes.

On the other hand, perspective changes with GFOV gains also modify the retinal size of a displayed object. As predicted by Emmert's Law, a systematic increase or decrease of retinal size changes the perceived size or distance of seen objects (see Section 7.2.1). Kuhl et al. [KTCR09] and Steinicke et al. [SBL+11] observed that changes of GFOVs have an impact on the perceived distance to virtual objects, whereas effects on perceived size have not been reported in the literature. In previous work, researchers mainly studied how distance underestimation with HMDs can be compensated by applying GFOV gains $g_{_F} > 1$. The predicted distance in the case of perceived-size constancy (see Section 7.2.1) can be described with the following coefficient [SBL+11]:

$$\hat{D}_0 = D_0 \cdot m_f. \tag{7.7}$$

The size of an object results as

$$\hat{S}_s = S_s, \quad \hat{S}_l = S_l \cdot m_f, \tag{7.8}$$

which matches the results from convergence cues. Perceived-size constancy may apply to object features in planes perpendicular to the view direction, but object features may appear stretched or compressed along the view direction.

Interpupillary Distance Gains

With GIPD gains, the distance between the cameras for the left and right eye can be changed leading to altered perspective and convergence cues. As illustrated in Figure 7.3(c), a gain $g_{_I} = 2$ results in the distance to an object being reduced by 50%, as well as its size being uniformly scaled by 50%. The relation between virtual and displayed object distance for convergence cues can be expressed as

$$\hat{D}_0 = D_0 \cdot g_{_I}^{-1}, \tag{7.9}$$

with D_0 the virtual target object distance, and $\hat{D}_0$ the resulting object distance from convergence cues when shown on the physical display. The size of an object results as

$$\hat{S} = S \cdot g_{_I}^{-1}, \tag{7.10}$$

with S the object size in the virtual scene, and $\hat{S}$ the resulting object size from convergence cues when shown on the physical display.

On the other hand, changing the GIPD also changes the Euclidean distance of the cameras to virtual objects, and defines how much the cameras see of the sides of objects (cp. Fig-

ure 7.1(b)). Thus, changing the GIPD also has an effect on the retinal size of virtual objects. However, this depends to a large part on the geometry of the virtual object. In a simple case, for a spherical object, the predicted distance in the case of perceived-size constancy results as

$$\hat{D_0} = \sqrt{D_0^2 + (g_r^2 - 1) \cdot (\text{IPD}/2)^2} \tag{7.11}$$

The size of a spherical object in simple form results as

$$\hat{S} = S, \tag{7.12}$$

with S the virtual size and $\hat{S}$ the resulting object size that can be derived from the retinal size in the case of perceived-size constancy.

Observations and Questions

From the above explanations for convergence cues, we can see that GFOV gains introduce a non-uniform scaling of the scene in look direction, whereas GIPD gains introduce a uniform scaling of the scene in front of the user's eyes. Indeed, these computational results broadly follow experimental observations in the literature [KTCR09, SBL+11, YOO06]. Moreover, as illustrated in Figure 7.3(d), one can compute pairs of GFOV and GIPD that–theoretically–exactly compensate for individual effects. However, Figures 7.3(c) and (d) also show that the perceptual system may need to deal with a cue conflict in the presence of perceived-size constancy (e. g., for objects with familiar size). The main questions that arise are, how such conflicts are resolved, and how much each of the rendering parameters GFOV and GIPD will contribute to the size and distance percepts.

7.4 Perceptual Experiment

In this section, we present an experiment, in which we have investigated the effects of changes to the GFOV and GIPD on relative size and distance judgments.

7.4.1 Experiment Design

As visual stimulus, we used a virtual hallway scene of 3.8m×2.5m×35m (in width, length, and height), which was rendered with Crytek's CryEngine 3. We used a split screen design of a virtual hallway (see Figure 7.4), with the left-hand side view being rendered with one pair of GFOV and GIPD, and the right-hand side view being rendered with another. We did not use the stereoscopic rendering facilities of the CryEngine 3, but added an interface to our own software, with which we handled the generation of the split-screen stereo pair and were able to provide accurate on-axis stereo graphics. In both virtual scenes, we placed a virtual avatar (Caucasian male, 1.85m height) as focus object. We considered the distance of the avatar from the observer as between-subjects variable and tested three distances: 4m, 6m and 8m.

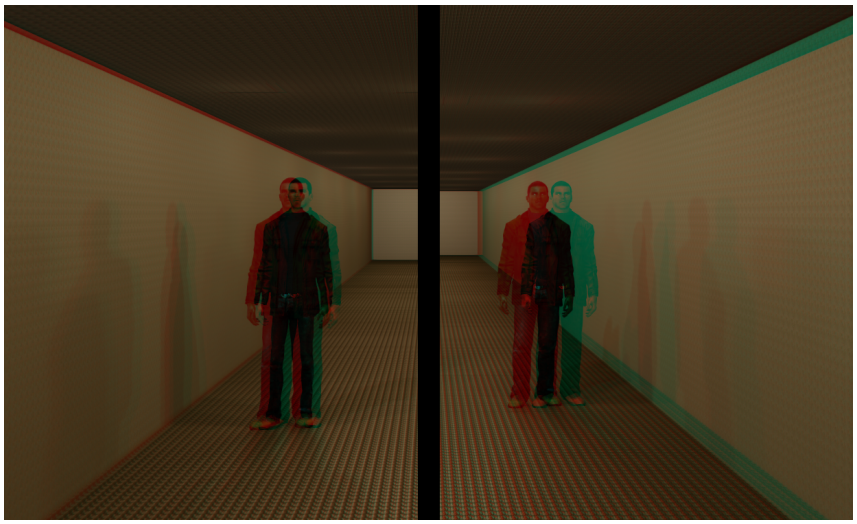


Figure 7.4: Illustration of the split-screen visual stimulus used in the experiment (here with red-cyan anaglyphs). Participants had to compare size and distance of the avatars displayed in the left and right view.

Hardware Setup

The participants were equipped with an NVIS nVisor SX60 HMD with a resolution of 1280×1024 pixels per eye at a refresh rate of 60Hz for the visual stimulus presentation. The nVisor HMD uses parallel symmetric on-axis display optics and has a diagonal FOV of 60 degrees, which we used as the basis for the GFOV transformations during visual stimuli generation. Views on the real world were blocked with a black light shield. During the experiment, the participant's head was oriented in view direction along the virtual hallway and presented at the participant's eye height. The experiment did not involve head movements of the participants in order to provide only static size and distance cues, but no motion cues.

We used an Intel computer (Core i7 processors, 6GB of RAM and an Nvidia Quadro 4000 graphics card) for rendering, system control, and logging. A standard keyboard served as means for the participants to enter their perceived size and distance judgments. The participants received all instructions on slides presented on the HMD. There was no communication between experimenter and participants during the experiment in order to focus participants on the task.

Participants

22 male and 17 female participants (age 18−44, $M = 23.49$, $SD = 4.93$) completed the experiment. Participants were students or members of the departments of computer science, psychology or human-computer-media. All had normal or corrected to normal vision; 7 wore glasses and 9 contact lenses during the experiment. We tested visual acuity of all participants with a simple vision test prior to the experiment based on a classic Snellen chart. We measured the interpupillary distances of our participants using the mirror method described by Willemsen et al. [WGTCR08]. The IPDs of our participants ranged between 5.0−7.1cm

(mean 6.14, SD 0.48). The eye height of our participants ranged between 1.51−1.87m (mean 1.66, SD 0.1). 26 of the participants had no experience with 3D games, 3 had some, and 10 had much experience. All participants were naïve to the experimental conditions. 31 of the participants had much experience with 3D stereoscopic cinema, 7 had some, and 1 had no experience. 4 of the participants had much experience with HMDs, 3 had some, and 32 had no experience. 3 participants had participated in experiments involving HMDs before.

We verified all participants' ability to see stereoscopically prior to the experiment by asking participants to look at an anaglyphic random-dot stereogram, and report the type of 3D object that was shown. Students obtained class credit for their participation. The total time per participant, including pre-questionnaires, instructions, training trials, experiment, breaks, post-questionnaires, and debriefing, was 1 hour. Participants wore the HMD for approx. 45 minutes. They were allowed to take breaks at any time; short breaks after every 50 trials were mandatory to rest the eyes for a few moments.

7.4.2 Methods

We used the method of constant stimuli in a two-alternative forced-choice (2AFC) task [Fer08]. In the method of constant stimuli, the applied gains are not related from one trial to the next but presented randomly and uniformly distributed. We applied GFOV gains $g_F \in \{0.5, 0.75, 1.0, 1.25, 1.5\}$ relative to the DFOV of the HMD, and GIPD gains $g_I \in \{0.0, 1.0, 2.0, 3.0, 4.0\}$ relative to the IPD of the participant. We varied combinations of GFOV and GIPD gains in the left and right views in the split screen design. We tested all combinations of GFOV and GIPD gains against all other combinations for all participants in random order.

In order to investigate the mutual impact of GFOV and GIPD on size and distance perception, participants had to answer two 2AFC questions at each trial. They had to choose between one of two possible responses: "Does the left or right avatar appear closer to you?" and "Does the left or right avatar appear smaller?"; responses like "I can't tell." were not allowed. Hence, if participants cannot detect the signal, they are forced to guess, and will be correct in 50% of the trials. The gain at which the participant favors one response in half of the trials corresponds to the point of subjective equality (PSE), at which the participant judges the size or distance of the avatars that are displayed with different rendering parameters as identical. As the GFOV and GIPD gains decrease or increase from this value, the ability of the participant to detect the differences in distances or sizes increases, resulting in psychometric curves for the discrimination performance. In order to avoid participants directly comparing the renderings of the virtual scene in subsequent trials, we displayed a blank image for 200ms between the renderings as a short interstimulus interval. Additionally, participants filled out the Kennedy-Lane simulator sickness questionnaire (SSQ) immediately before and after the experiment, further the Slater-Usoh-Steed (SUS) presence questionnaire, and a demographic questionnaire.

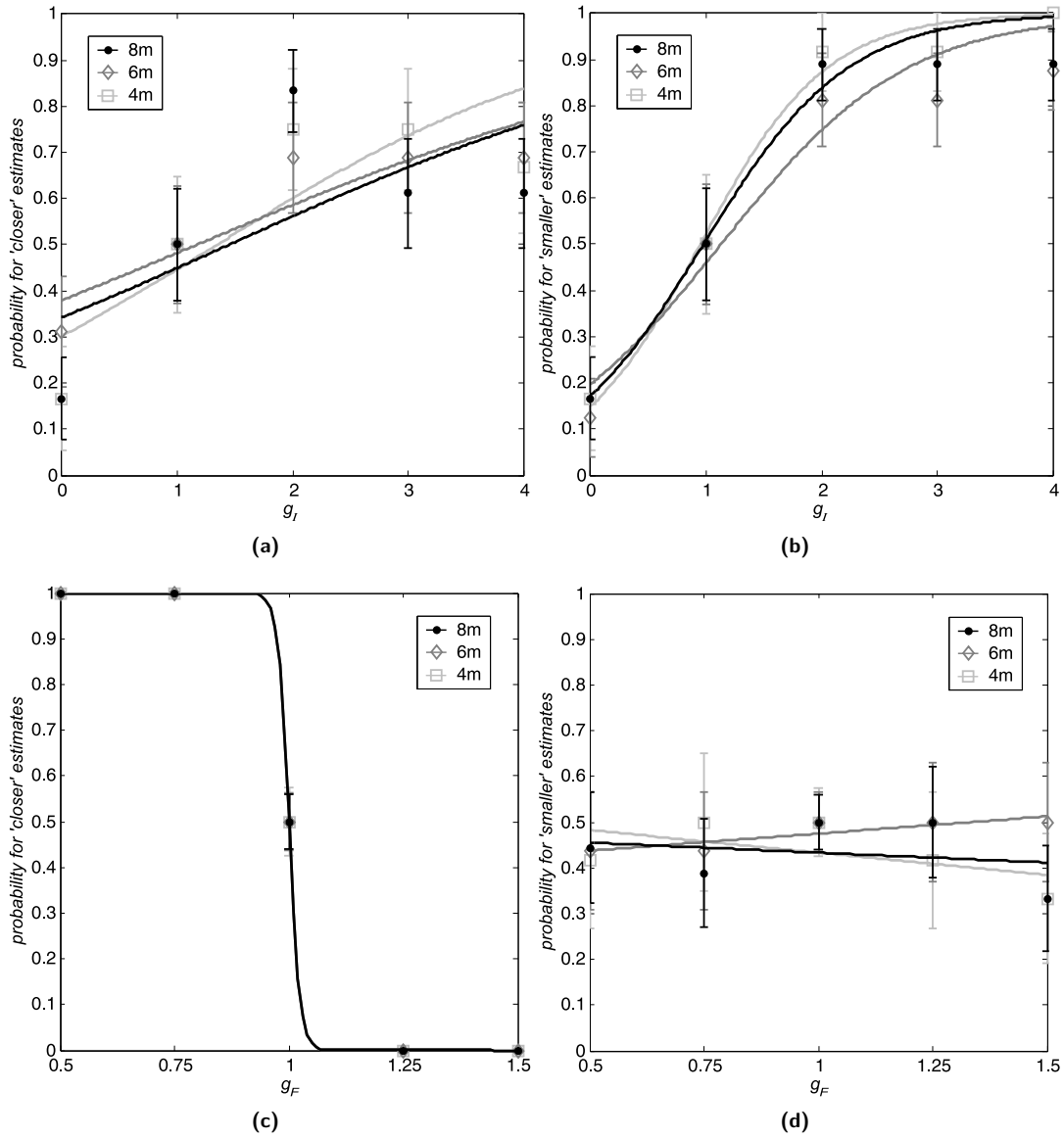


Figure 7.5: (a+b) Results of $g_F = 1$ in both views, $g_I = 1$ in one view, with $g_I \in \{0, 1, 2, 3, 4\}$ on the x-axis varied in the other view. The y-axis shows the probability that participants judged the target object displayed with the varied g_I as (a) closer to the observer or (b) smaller. (c+d) Results of $g_I = 1$ in both views, $g_F = 1$ in one view, with $g_F \in \{0.5, 0.75, 1.0, 1.25, 1.5\}$ on the x-axis varied in the other view. The y-axis shows the probability that participants judged the target object displayed with the varied g_F as (c) closer to the observer or (d) smaller.

7.4.3 Results

We pooled the data over all participants for the two 2AFC tasks in the three between-subjects groups. We had to exclude the data of one participant from the 8m target distance group, who showed strong simulator sickness symptoms and made inconsistent judgments over a large part of the experiment. The participants reported low mean SUS presence scores (mean

2.62, SD 0.97). SSQ mean scores increased from 7.11 (SD 8.60) before the experiment to 28.95 (SD 21.10) after the experiment.

Figure 7.5 shows the pooled results of the participants for the three tested target object distances (8m, 6m, and 4m). In Figures 7.5(a) and (b) we plotted the effects of varying the g_I in one view, with $g_I = 1$ in the other view, and $g_F = 1$ in both views, on distance and size judgments, respectively. The y-axis shows the probability that participants judged the avatar displayed with the varied g_I as (a) closer to the observer or (b) smaller. In Figures 7.5(c) and (d) we plotted the effects of varying the g_F in one view, with $g_F = 1$ in the other view, and $g_I = 1$ in both views, on distance and size judgments, respectively. The y-axis shows the probability that participants judged the avatar displayed with the varied g_F as (c) closer to the observer or (d) smaller. The solid black functions show the results for the 8m target object distance, the dark gray functions for the 6m distance, and the light gray functions for the 4m distance. The error bars show the standard error. The sigmoid psychometric functions are fitted to the data with $f(x) = \frac{1}{1+e^{ax+b}}$, $a, b \in \mathbb{R}$. The chi-square goodness of fit for the functions is shown in Table 7.1.

Figure 7.6 shows the pooled results of the participants in the 6m target distance condition[2]. We plotted the participants' responses when seeing one view with $g_F = 1$ and $g_I = 1$, as well as the other view varied with $g_F \in \{0.5, 0.75, 1, 1.25, 1.5\}$ on the x-axis and $g_I \in \{0, 1, 2, 3, 4\}$ on the y-axis. The color gradients show the probability that participants judged the target object as (a) closer to the observer or as (b) smaller.

Table 7.1: χ^2 goodness of fit for the psychometric functions for the three target distances plotted in Figure 7.5.

Fig.	8m	6m	4m
7.5(a)	0.82	0.09	0.37
7.5(b)	0.18	0.22	0.21
7.5(c)	1e-14	1e-14	1e-14
7.5(d)	0.13	0.01	0.10

7.4.4 Discussion

Results indicate that the GFOV (within the tested range) had a strong effect on distance judgments. Larger GFOV gains g_F resulted in objects being judged as farther away from the observer (see Figure 7.5(c)), which is in line with the predictions of the models described in Section 7.3.2. The GIPD (within the tested range) also showed an effect on distance judgments. Larger GIPD gains g_I resulted in objects being judged as closer to the observer (see Figure 7.5(a)), although distance discrimination performance appears to be less influenced by the GIPD than by the GFOV (see Figure 7.6(a)). Figure 7.6(a) further reveals that the tested GIPD gains had only a slight effect on distance judgments when set in direct relation to the tested range of GFOV gains. These results suggest that distance perception in the

[2]Plots for 4m and 8m distances showed the same qualitative distribution.

tested realistic scene relies less on convergence cues than predicted in Section 7.3.2 (cp. Figure 7.3(d)). We observed no consistent effect of the chosen object distances on relative distance judgments.

Results further show an effect of the GIPD on size judgments. Larger GIPD gains g_I resulted in objects being judged as smaller (see Figure 7.5(b)), which correlates with the predictions of the convergence cue model described in Section 7.3.2. We observed no consistent effect of GFOV gains on size judgments (cp. Figures 7.5(d) and 7.6(b)). We found no consistent effect of the chosen object distances on relative size judgments.

Our results indicate that for cue conflicts introduced with GFOV and GIPD gains in realistic virtual scenes, human distance and size perception differ from the predictions of the models for reduced-cue visual stimuli (i. e., stereopsis and retinal size) described in Section 7.3.2. In particular, for a typical range of miscalibrated GFOVs, the results indicate a strong effect on relative distance perception, with little effect on relative size perception. In contrast, for a typical range of miscalibrated GIPDs, the results indicate only a slight effect on relative distance perception, but a stronger effect on relative size perception. Figure 7.6(a) moreover suggests that it may be possible to provide a controlled cue conflict stimulus for distance cues by balancing GFOV and GIPD gains, although merely within a small range.

Limitations

These are interesting results since they suggest that GFOV and GIPD gains have different effects in realistic scenes than for reduced-cue visual stimuli. However, we have to consider that these quantitative results are likely dependent on a variety of other cues that were specific to the visual stimulus used in our experiment (e. g., retinal size, stereopsis, textures, angle of declination, and height-in-the-picture). Another limitation of the results may be caused by the split-screen design of the visual stimulus. Although the split-screen allowed participants to directly compare two different renderings with no temporal distortion, this directness could have introduced possible cross-effects due to individual perceptual requirements of alternately viewing the left and right hallways. Finally, the gain ranges we chose for the GFOV and GIPD may have imposed a limitation on our results. It is not unlikely that, for much narrower gain ranges (see Figure 7.6(a)), we may have found a clearer relationship between GFOV and GIPD for distance perception.

7.5 Conclusion

We have investigated the effects of changing the geometric field of view and geometric interpupillary distance when displaying a virtual scene on physical displays. We have described selected size and distance cues that can be derived from egocentric perspective and stereopsis, and we have described the effects of manipulating field of view and interpupillary distance on these cues. In a psychophysical experiment, we have evaluated the mutual impact of the two parameters on size and distance perception and set the results in relation to the models for the cues. In this work, we made one step toward understanding what happens if incorrect GFOVs

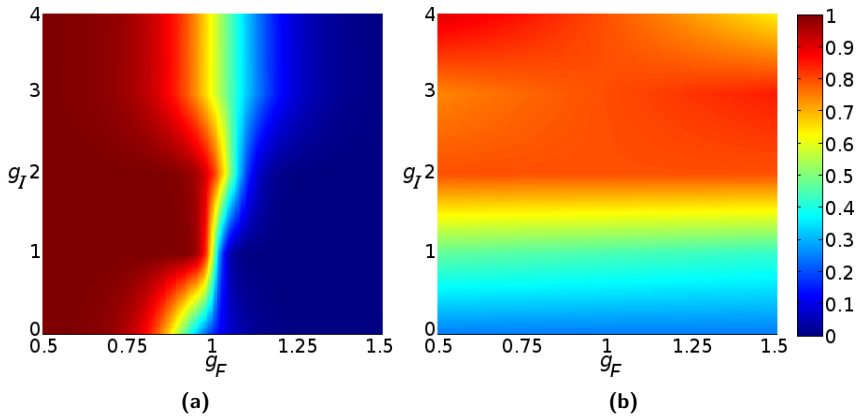


Figure 7.6: Result plots of $g_F = 1$ and $g_I = 1$ in one view, with $g_F \in \{0.5, 0.75, 1, 1.25, 1.5\}$ and $g_I \in \{0, 1, 2, 3, 4\}$ varied in the other view. The colors show the probability that participants judged the target object as (a) closer to the observer or (b) smaller.

or GIPDs are applied in on-axis stereoscopic display environments. Future work includes to further disentangle the effects of miscalibration of immersive VR displays on different spatial cues and further evaluate resulting cue conflicts. In particular, there are many more cues that we did not consider in this work, some of which are changed by GFOV or GIPD gains (e. g., angle of declination [KTCR09] or accommodation blur [HCOB10]).

Chapter 8

8 CAVE Size Matters: Effects of Screen Distance and Parallax on Distance Estimation in Large Immersive Display Setups

When walking within a CAVE-like system, accommodation distance, parallax and angular resolution vary according to the distance between the user and the projection walls which can alter spatial perception. As these systems get bigger, there is a need to assess the main factors influencing spatial perception in order to better design immersive projection systems and virtual reality applications. In this chapter, we present two experiments which analyze distance perception when considering the distance towards the projection screens and parallax as main factors. Both experiments were conducted in a large immersive projection system with up to ten-meter interaction space. The first experiment showed that both the screen distance and parallax have a strong asymmetric effect on distance judgments. We observed increased underestimation for positive parallax conditions and slight distance overestimation for negative and zero parallax conditions. The second experiment further analyzed the factors contributing to these effects and confirmed the observed effects of the first experiment with a high-resolution projection setup providing twice the angular resolution and improved accommodative stimuli. In conclusion, our results suggest that space is the most important characteristic for distance perception, optimally requiring about 6 to 7-meter distance around the user, and virtual objects with high demands on accurate spatial perception should be displayed at zero or negative parallax.

8.1 Introduction

Immersive virtual reality (VR) systems can provide users with a sense of feeling present in the displayed virtual environment (VE) similar to perceiving an environment in the real

world [Sla09]. Recent advances in hardware technologies make it possible to build immersive projection environments (IPE), such as CAVEs [CNSD+92], with a large room-sized interactive workspace. Such IPEs support natural embodied forms of interaction with the displayed virtual world in the egocentric frame of reference of the observer, including real walking. These affordances facilitate spatial impressions of the VE that are important for exploration and review in a wide range of application domains such as architecture and engineering in which users benefit from experiencing the VE at real scale.

For such applications, it is essential to facilitate a veridical impression of the spatial layout, e. g., sizes, distances, and interrelations, within the perceived virtual world. Modern real-time rendering systems can create compelling immersive experiences offering most of the spatial visual cues we can find in the real world, including perspective, interposition, lighting, and shadows [TFCRS11]. However, distance and size perception are often biased in such environments, causing users to overestimate or underestimate spatial relations [IRLA07, LK03, TWG+04]. The particular factors influencing a user's distance estimates in IPEs are not yet clearly identified, and large portions of the observed misperception effects still cannot be explained [LK03, RVH13].

Although IPEs differ from the real world in many respects, issues with the visual rendering and display technologies have naturally been suggested as a potential cause of non-veridical spatial perception. One of the potential suspects for such misperception is the accommodation-convergence conflict [HGAB08]. In the real world, accommodation and convergence are coupled together and provide distance cues up to a distance of about six meters [CRWGT05, CV95, WGTCR08]. In stereoscopic display systems, the observer accommodates to the distance of the display surface to perceive objects without blur, whereas the convergence angle depends on parallax. Three parallax conditions are considered: *negative parallax* (object in front of the display), *zero parallax* (object on the display) and *positive parallax* (object behind the display) [Bou99]. With negative or positive parallax the user's visual system is confronted with conflicting depth information and might be misguided by the accommodative information [DM96].

In this chapter, we present two experiments in which we have quantified egocentric distance perception in an IPE with an interaction space up to ten meters. In such large IPEs, the accommodation distance, the accommodation-convergence mismatch and the angular resolution vary largely depending on where the user is standing and where virtual objects are displayed. In comparison to previous distance perception research in IPEs, which were restricted to small negative parallaxes due to limited interaction workspaces, to our knowledge, we detail the first analysis of distance perception in an IPE that supports to display virtual objects with such a large negative parallax.

Our main contributions are:

- We analyze the role of stereoscopic parallax, angular resolution, accommodative stimuli and retinal size for screen and target distances ranging from 1 to 9 meters.
- Our results reveal a strong asymmetric effect of screen distance and stereoscopic parallax

on distance estimation.

- We discuss the contributing factors and implications for future immersive projection setups.

The remainder of this chapter is structured as follows. Section 8.2 presents background information on distance perception. Sections 8.3 and 8.4 describe the psychophysical experiments in which we assessed distance perception. Implications and guidelines are discussed in Section 8.5. Section 8.6 concludes the chapter.

8.2 Background

In this section, we summarize information about distance cues in the scope of the conducted experiments.

8.2.1 Stereopsis

The binocular configuration of human eyes provides the brain with two views of a scene from laterally separated positions at a fixed interpupillary distance (IPD). Solving the disparity correspondence problem, the brain may relate retinal image contents from the two eyes to one another, computing the distance to seen objects via triangulation. Further, when focusing on an object, the eyes need to rotate toward that object to bring it to the fovea of each retina. The convergence state of the eyes, changed by extrinsic muscle exertion, provides an absolute cue about the distance to an object. In a simplified setting, the distance $D_0 \in \mathbb{R}^+$ of an object can be computed from the user's IPD $\in \mathbb{R}^+$ and the convergence angle $\alpha \in \mathbb{R}^+$ [BPS12] (see Figure 8.1):

$$D_0 = \frac{\text{IPD}}{2 \cdot \tan(\frac{\alpha}{2})}. \tag{8.1}$$

Stereoacuity is naturally limited with a conservative threshold of about 10 seconds of arc [PGG+10]. Considering this threshold and referring to Equation 8.1, the maximum distance at which stereopsis may produce usable data would be ca. 1.24km. However, in IPEs the *angular resolution* of pixels on a screen acts as an artificial cut-off to the theoretical capabilities of human vision.

8.2.2 Accommodation

The human eye can alter its optical power to hold objects at different distances into sharp focus on the retina. When an object is fixated by the eye, the ciliary muscles are adjusted such that a sharp image is perceived on the retina. The state of the ciliary muscles provides an absolute cue about the distance to the focused object. However, full accommodation response depends on the *accommodative stimulus*, i. e., responses of ciliary muscles differ between fuzzy and high-contrast stimuli [FC88].

The role of accommodation in distance perception is controversial. While a body of early

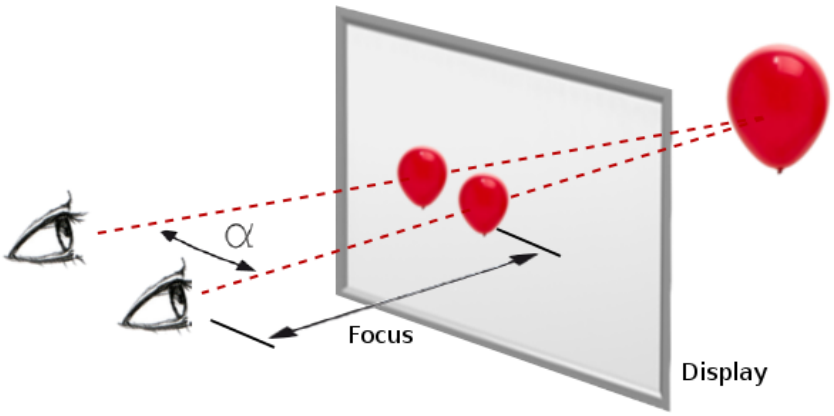


Figure 8.1: Illustration of object distance as a function of eye convergence angle α and screen distance from accommodation with stereoscopic displays.

work found negligible effects of accommodation on human distance perception, many of these results nowadays have to be reconsidered with improved experimental analyses [FC88]. While accommodation can be leveraged as a distance cue, the influence of the accommodative information declines with age [Pie93], distance [CV95] and with decreasing cue reliability [HGAB08].

8.2.3 Accommodation-Convergence Conflict

Oculomotor responses of accommodation and convergence usually co-vary to provide a sharply focused view. Accommodation is mainly driven by retinal blur (monocular cue), and convergence by retinal disparity (binocular cue). However, with stereoscopic displays, the physiologically coupled oculomotor processes of convergence and accommodation are dissociated. Observers fixate an object with the same convergence as in natural viewing, but the eyes focus on the screen and not the object, which can bias distance estimation and can cause visual discomfort [HGAB08]. Loomis and Knapp [LK03] and Renner et al. [RVH13] provide thorough reviews of the literature on effects of visual conflicts on distance estimation in IPEs. Although the reported studies are based on different materials and methods, their results agree that users tend to underestimate egocentric distances in vista space in IPEs [ATd+10, GDP+10, KSS+09]. In particular, Piryankova et al. [PdK+13] observed distance underestimation over multiple immersive large screen displays, as well as an interaction effect with the distance to the displayed target.

We have to note that most of these studies focused on virtual objects with positive parallax. In such situations, larger distances to objects correlate with smaller convergence angles, but also with reduced angular resolution and diminished accommodation responses due to more blur in retinal images. In contrast, objects displayed with negative parallax cause the conflict to reverse its sign, and accommodation responses benefit from the reduced blur in retinal images [BSS13b]. Considering that objects displayed near zero parallax approximate viewing as in the real world, it is a challenging question whether zero parallax defines a singularity of

optimal distance estimation. So far, the effects of the sign of the conflict and blur in retinal images due to low angular resolution are not yet clear.

8.3 Psychophysical Experiment

In this section we describe the experiment which we conducted to analyze the interrelations between the egocentric distance to the projection screen (i. e., screen distance $D_s \in \mathbb{R}^+$) and the distance to a visual target (i. e., target distance $D_t \in \mathbb{R}^+$) in terms of distance judgments measured with a triangulated pointing method [KSS+09].

8.3.1 Material

The experiment was conducted in a $9.6\text{m} \times 3\text{m} \times 3.1\text{m}$ (width, depth, and height) 4-sided IPE (see Figure 8.2) equipped with 16 Barco Galaxy projectors at 15MPixels resolution in total. The fourth wall was closed during the experiment using an opaque black light shield to avoid distractions and external cues [MCM14]. The pixel size for the side walls was $1.56\text{mm} \times 2.56\text{mm}$ and for the front wall $1.36\text{mm} \times 1.47\text{mm}$. For visual display, system control and logging we used a cluster of 7 HP Z400 with 1×7 Nvidia Quadro FX 5000 and 2 HP Z420 with 1×2 Nvidia Quadro 5000 graphics cards. The VE was rendered using the Unity 3D Pro game engine with the MiddleVR plugin for multi-surface rendering. Participants wore shutter glasses (Volfoni ActivEyes Pro Radiofrequency) for stereoscopic visual stimulus presentation. The shutter glasses were tracked with 6 degrees of freedom passive markers using an ART optical tracking system with 16 cameras at an update rate of 60Hz. An ART Flystick 2 was used for the pointing task.

8.3.2 Protocol

Participants had to judge the distance to a seen virtual target object using the method of blind triangulated pointing, which we adapted to the configuration of our projection setup. Similar to previously introduced procedures [FLD97, KSS+09], participants held the Flystick as they observed the object. When participants were ready to judge the distance to the object, they had to close their eyes, trigger the button of the Flystick to fade to black the rendered scene, take two steps to the left or right, and point the Flystick to the object (see Figure 8.2). Participants were instructed to point at the target as accurately as possible while performing the side stepping at a reasonable speed to reduce effects of decreased precision caused by changes in the remembered position of the target over time [PGJ01]. Participants received no feedback about their pointing accuracy in order to minimize the effects of perception-action motor recalibration in the response method while assessing distance perception.

The visual stimulus consisted of a virtual scene as shown in Figure 8.2, i. e., an all-gray virtual world, and virtual balloons were chosen as targets for the distance estimation task. Traditional helium party balloons in the real world have a standardized size of 28cm, thus providing known retinal size cues [SCRT+15]. Helium balloons are one of the few objects in

the real world that occur floating in mid-air [SDO06].

Instructions were provided on a computer screen prior to the experiment. In order to focus participants on the tasks, no communication between experimenter and participant was performed during the experiment after the initial training phase, in which we ensured that participants correctly understood the task. After each trial, a new starting position was shown on the floor of the IPE to determine the start position and orientation of the next trial. We instructed participants always to point to the center axis of the virtual balloon with an outstretched arm with their dominant hand. The round shape of the target balloon has the benefit that pointing towards its center is independent of the pointing angle, i. e., rotationally invariant, which would be confounded using a traditional flat target.

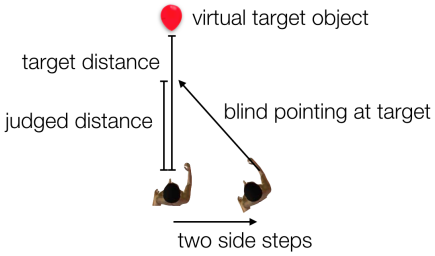
8.3.3 Methods

We followed a repeated measures within-subjects design. The independent variables were the screen distance (D_s) and the distance to the virtual target object (D_t). We instructed participants to assume different positions in the IPE while standing upright (D_s). These positions were at 1m to 9m distance in 2m steps from a side wall of the IPE. Virtual target objects were rendered at approximate eye height at distances of 1m to 9m in 2m steps (D_t). To avoid bias, each target was displayed either on the left or right side wall of the projection setup; the order was counterbalanced. During each trial, users stood at a fixed distance $D_s \in \{1, 3, 5, 7, 9\}$m to a projection wall, which defines the screen distance, while virtual target objects were placed at a fixed distance $D_t \in \{1, 3, 5, 7, 9\}$m from the participant. In particular, for each screen distance we tested one condition in which the virtual target object was centered around zero parallax, i. e., where $D_t = D_s$. Participants were guided to the positions in the immersive setup via virtual markers that we projected on the floor between trials.

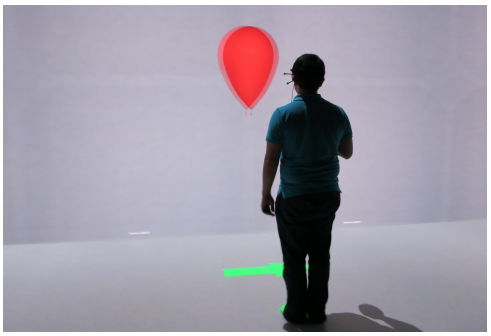
The experiment was divided into *two main blocks*: In the *first block*, participants performed the triangulated blind pointing task (eyes-closed block), while in a *second block*, we measured the ability of participants to accurately and precisely point to the 3D targets (eyes-opened block). Therefore, participants had to complete the triangulated pointing trials with open eyes, i. e., they observed a distant object, performed two side-steps, and pointed at its position without closing their eyes. We measured this ground truth pointing data to analyze pointing behavior and to calibrate the results of the first part of the experiment.

The screen distances were balanced using a Latin Squared design. For each screen distance, the order of the distances towards the virtual object was randomized. For each combination, there were 4 repetitions (two at each side wall). In summary, participants completed 5 (screen distances) × 5 (target distances) × 2 (side walls) × 2 (repetitions) × 2 (experiment bocks) = 200 trials, as well as 5 training trials for each block of the experiment, which were excluded from the analysis. Participants were allowed to take a short break at any time between trials. A short break between the two blocks of the experiment was mandatory.

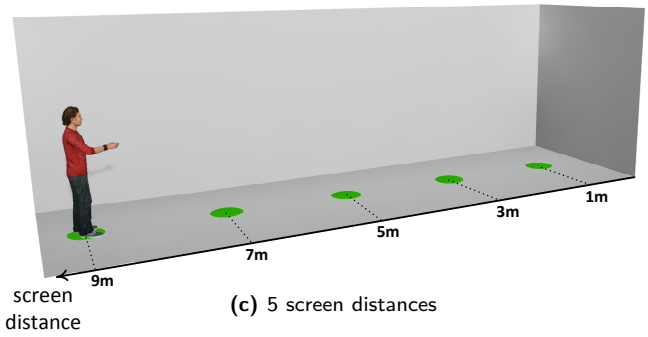
The dependent variable was the distance estimate. From the initial view direction to the target object, as well as the position and pointing direction after the participant performed



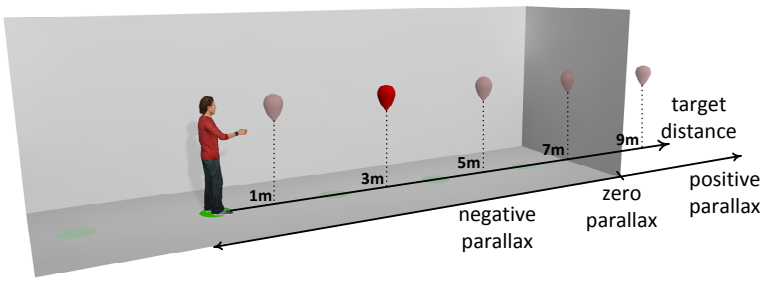
(a) blind triangulated pointing task



(b) participant during experiment



(c) 5 screen distances



(d) 5 target distances

Figure 8.2: Experimental design: (a) Illustration of the blind triangulated pointing task, (b) performed by a participant during the experiment with visual indicator of side stepping direction. Participants performed the pointing task counterbalanced facing the left and right side wall in the display setup at (c) 5 different screen distances (green circles) and for each position estimated the distance of virtual objects at (d) 5 target distances (red balloons), resulting in 5×5 conditions in total with half of the targets displayed with negative and positive parallax, respectively.

the side-steps, we computed the judged distance to the perceived position of the virtual target [FLD97, KSS+09].

Considering the previous results in the literature [ATd+10, GDP+10, KSS+09] and the distance cues described in Section 8.2, our hypotheses were:

H1 No underestimation nor overestimation of the distance to objects at zero parallax.

H2 Underestimation of the distance to objects exhibiting positive parallax.

H3 Overestimation of the distance to objects exhibiting negative parallax.

H4 More accurate distance estimation for longer screen distances.

Furthermore, we collected demographic information with a questionnaire before the experiment and measured the participants' sense of presence with the Slater-Usoh-Steed (SUS) questionnaire [UCAS99], as well as simulator sickness with the Kennedy-Lane SSQ [KLBL93] before and after the experiment. The total time per participant including pre-questionnaires, instructions, training, experiment, breaks, and debriefing was 1 hour. Participants were immersed in the VE for about 45 minutes.

8.3.4 Participants

We recruited 15 participants for our experiment, 13 male and 2 female (aged from 23 to 38, $M = 28.1$). The participants were students or professionals in computer science or engineering. All participants reported that they were right-handed, which we confirmed with the Lateral Preference Inventory questionnaire [Cor93]. Six participants wore glasses and three wore contact lenses during the experiment. We measured each participant's visual acuity before the experiment using a Snellen chart. 13 participants had at least 20/20 visual acuity and 2 participants had 20/30. None of the participants reported known vision disorders, such as color or night blindness, dyschromatopsia, or a known displacement of balance. 13 participants reported previous experience with 3D stereoscopy (rating scale 0 = yes, 4 = no, $M = 1.67$, $SD = 1.45$). 10 participants had participated in a study in the immersive projection setup before. Using the technique proposed by Willemsen et al. [WGTCR08] we measured the IPD of each participant before the experiment started ($M = 6.49$cm, $SD = .29$cm). Moreover, we measured the eye height of each participant ($M = 1.65$m, $SD = .063$m).

The data from one of the users was not considered in the analysis due to a hardware malfunction.

8.3.5 Results

First, we analyzed the results of the eyes-opened block. We observed angular errors of $M = 0.45$ degrees ($SD = 1.19$ degrees) in pointing performance with eyes open. The results show that participants achieved overall high pointing accuracy and precision in the considered range of target distances with no significant bias. The results show that the system was working properly and the protocol is valid.

Regarding the eyes-closed block, Figures 8.3a and 8.3b show the pooled results for the screen distances $D_s \in \{1, 3, 5, 7, 9\}$m with the standard error of the mean. We found no significant difference between the results for the left and right side wall of the immersive projection setup, so we pooled the responses. The x-axes show the actual target distances $D_t \in \{1, 3, 5, 7, 9\}$m, the y-axes show the judged target distances. The gray lines show the distribution of judged

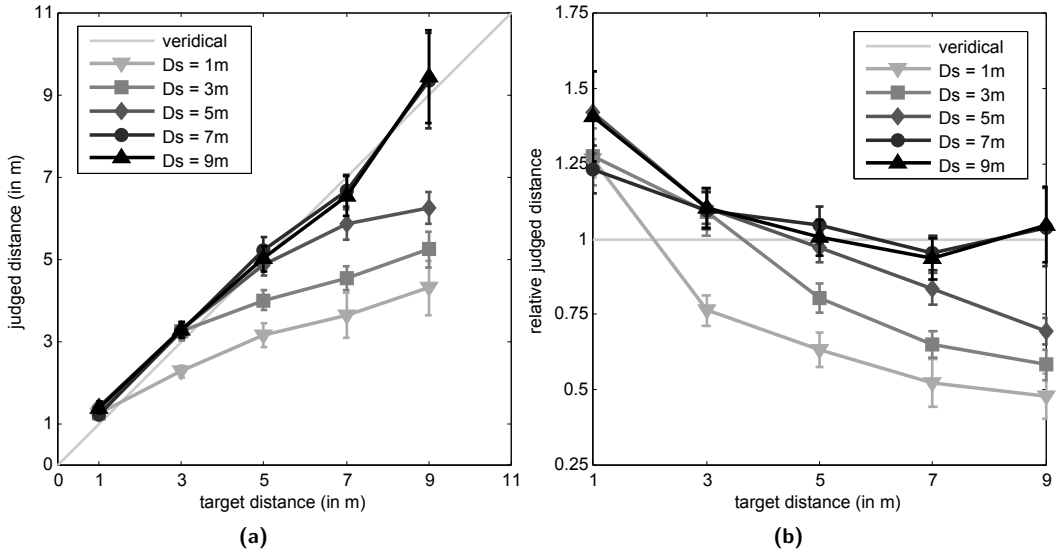


Figure 8.3: Pooled results of the judged distances for the different screen distances (D_s) in the first experiment. The x-axes show the actual distance to the target object. The y-axes show the (a) absolute and (b) relative judged distance. The light to dark gray lines show the results for $D_s \in \{1, 3, 5, 7, 9\}$ meters, respectively.

distances $D_j \in \mathbb{R}_0^+$ in the different conditions. We computed relative judged distances as D_j/D_t, i. e., values near 1 indicate ideal results, whereas values >1 indicate overestimation, and values <1 underestimation.

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. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated. In the following, we report statistics for both absolute and relative judgments to account for the two mainly used types of analyses in the distance estimation literature.

We found a significant main effect of screen distance on absolute distance judgments, $F(1.99, 25.84) = 27.367$, $p < .001$, $\eta_p^2 = .678$, and on relative distance judgments, $F(2.11, 27.40) = 22.749$, $p < .001$, $\eta_p^2 = .636$. Post-hoc tests showed that the judged distances between each two screen distances were significantly different ($p < .05$), except between $D_s = 5\text{m}$ and $D_s = 7\text{m}$, and between $D_s = 7\text{m}$ and $D_s = 9\text{m}$. In addition, we found a significant main effect of target distance on absolute distance judgments, $F(1.30, 16.87) = 67.094$, $p < .001$, $\eta_p^2 = .838$, and on relative distance judgments, $F(1.31, 17.02) = 21.933$, $p < .001$, $\eta_p^2 = .628$. Post-hoc tests showed that the judged distances between each two target distances were significantly different ($p < .05$), except between $D_t = 7\text{m}$ and $D_t = 9\text{m}$. Moreover, we found a significant interaction effect between screen distance and target distance on absolute distance judgments, $F(16, 208) = 10.499$, $p < .001$, $\eta_p^2 = .447$, and on relative distance judgments, $F(16, 208) = 5.819$, $p < .001$, $\eta_p^2 = .309$. Post-hoc tests showed that objects with zero and negative parallax exhibited similar distance estimations for each D_s. For example,

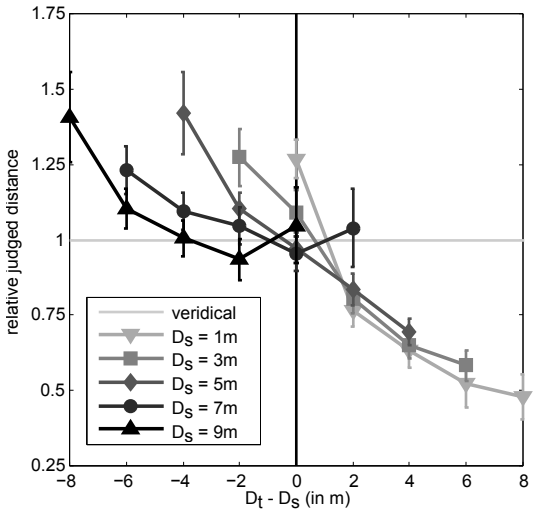


Figure 8.4: The x-axis shows the difference between target and screen distances (i. e., $D_t - D_s$). The y-axis shows the relative judged distance in the first experiment. The light to dark gray lines show the results for $D_s \in \{1, 3, 5, 7, 9\}$ meters, respectively.

at $D_t = 1$, there are no significant differences among each D_s, this is also true for $D_t = 3$ when $D_s \geq 3$, for $D_t = 5$ when $D_s \geq 5$, and for $D_t = 7$ when $D_s \geq 7$. The only exception is $D_t = 9$ in which there is no significant difference among $D_s = 7$ and $D_s = 9$. In contrast, for objects exhibiting positive parallax, post-hoc tests showed significant differences (all $p < .05$) among all screen distances.

Finally, we analyzed the results considering the difference between target and screen distances ($D_t - D_s$). Figure 8.4 shows the pooled data. We compared pooled distance judgments for targets at zero parallax, positive parallax, and negative parallax. We observed a main effect of parallax on relative distance judgments, $F(2, 26) = 41.106$, $p < .001$, $\eta_p^2 = .760$. Post-hoc tests showed that relative distance judgments were significantly closer to veridical for zero parallax than for positive parallax, $t(13) = 8.849$, $p < .001$. Moreover, relative distance judgments significantly differed between positive and negative parallax, $t(13) = 6.833$, $p < .001$. We found a trend in relative distance judgments between zero parallax and negative parallax, $t(13) = 1.931$, $p = .076$.

Questionnaires

We measured a mean SSQ-score of $M = 17.6$ ($SD = 14.8$) before the experiment, and a mean SSQ-score of $M = 24.3$ ($SD = 20.2$) after the experiment. This increase in simulator sickness symptoms was not significant, $t(13) = 1.26$, $p = .23$. The mean SUS-score for the reported sense of feeling present in the VE was $M = 4.2$ ($SD = .78$), which indicates a reasonably high level of presence [UCAS99].

8.3.6 Discussion

In line with our Hypothesis H1, we observed a singularity for objects displayed at zero parallax (see Figure 8.4), for which participants on average were significantly more accurate at distance judgments than for objects displayed with positive parallax. However, we only found a trend for a difference between negative and zero parallax. Supporting our hypotheses H2 and H3, we found that participants on average overestimated distances to objects with negative parallax, but showed an underestimation for longer distances. Furthermore, the magnitude of underestimation was higher than that of the observed overestimation (see Figure 8.3a).

The results also reveal an interaction effect between the screen distance and the distance to a virtual target object in terms of a user's distance judgments. Distance judgments are strongly affected by the position of a user in a CAVE-like immersive setup. Post-hoc tests showed that for each target distance D_t distance judgments are similar when the virtual object exhibits zero or positive parallax ($D_t \geq D_s$), with the only exception of $D_t = 9$ for $D_s = 7$ (see Figure 8.3a).

Interpretation and Limitations

Our results show an effect of viewing distance from the projection screen on distance judgments as well as an interaction effect with stereoscopic parallax. One possible explanation for the results is related to the accommodation-convergence mismatch. As stereoscopic parallax increases or decreases from zero parallax, the accommodation-convergence mismatch usually increases (i. e., in case of full accommodation responses, cf. Section 8.2), which is characterized by the convergence cue indicating the distance to the virtual object, whereas the accommodation cue indicates the distance to the physical screen. Since the difference between these indicated distances is signed, this may explain why objects with negative parallax were overestimated and objects with positive parallax underestimated. Moreover, at screen distances $D_s \geq 7$ the accommodation cue might be less taken into account for distance estimates [CV95], which may explain why judged distances were more accurate for longer screen distances.

However, further experimentation is needed to understand if the accommodation-convergence mismatch is the actual cause of the observed effects. There are other possible explanations for the effects: First, changes in the angular resolution may explain some of the observed underestimation. A low angular resolution may act as an artificial cut-off to binocular distance cues (see Section 8.2.1) and may reduce accommodation responses (see Section 8.2.2). The angular resolutions in our experiment ranged from 5.36×8.8 arcmin/px for $D_s = 1\text{m}$ to 0.59×0.97 arcmin/px for $D_s = 9\text{m}$ (constant pixel size of $2.6\text{cm} \times 1.56\text{cm}$). The closer the user is to the screen, the lower is the angular resolution. Second, the retinal size (projection) of the virtual stimuli is proportional to the screen distance and inversely proportional to the target distance. Objects at negative parallax take up a larger screen space on the projection wall than objects at positive parallax, i. e., these objects are represented by more pixels on the wall. In the worst case scenario $D_s = 1\text{m}$ and $D_t = 9\text{m}$ the projected size

is approximately 2.6cm × 1.5cm. Furthermore, the sparse depth cues in the visual stimuli in our controlled experiment might have reduced overall precision in distance judgments, and peripheral vision of the bezels of the projection setup might have had an additional effect on the results.

8.4 Second Psychophysical Experiment

We conducted a second experiment to verify the experimental results in an IPE with higher angular resolution and to investigate which factors caused the strong distance underestimation for screen distances up to three meters in the first experiment (cf. Figure 8.3).

8.4.1 Material and Methods

As previously discussed, the angular resolution of visual stimuli is dramatically reduced in IPEs for short screen distances. In order to evaluate this potential factor, the second experiment was conducted on the main screen of our IPE, which provided *twice the angular resolution*, given the same configuration as in the first experiment while it allowed us to test screen distances of up to three meters.

In addition to the doubled resolution, which may benefit accommodation responses and thus increase accommodation-convergence conflicts, we also considered the influence of two different stimulus types with an additional factor in the experiment (see Figure 8.5):

- We performed the experiment with target objects textured with a *checkerboard* pattern, which supports accommodation responses due to salient luminance edges (cf. Section 8.2).
- As a baseline we tested the *plain* target objects used in the first experiment, for which features of the target object appeared less sharp at large target distances, thus potentially limiting accommodation responses.

Furthermore, the influence of *retinal size* was taken into account in the experiment design. Retinal size provides information about the absolute distance to an object of known size, or information about the relative distance between objects of unknown size [Pal99]. Moreover, retinal size is related to the screen space that objects take up on the projection walls and thus the number of pixels that are available to represent the objects. We evaluated the potential influence of retinal size on distance judgments with an additional factor, in which we introduced conflicts between the retinal size and the distance of the target object (see Figure 8.5):

- We applied a cyclopean *constant* scaling, i. e., the retinal size remained constant while the distance to the object increased.
- We tested effects of a *reverse* correlation between retinal size and object distance, i. e., the retinal size increased with object distance.

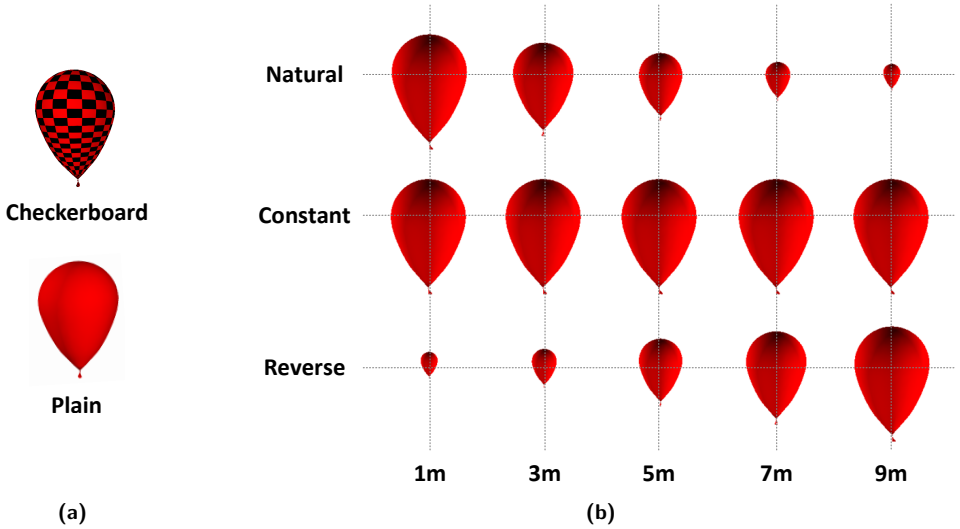


Figure 8.5: Visual stimuli used in the second experiment: (a) target objects displayed with checkerboard pattern and plain color and (b) retinal size conditions. Each cell represents the projection of the visual stimuli on the projection screen. The leftmost column corresponds to the target at $D_t = 1\text{m}$ and the rightmost column to $D_t = 9\text{m}$. (Top) Natural: retinal size decreases with target distance. (Center) Constant: retinal size does not change with target distance. (Bottom) Reverse: retinal size increases with target distance. The retinal size of the target at $D_t = 1\text{m}$ in the reverse condition matches the retinal size of the target at $D_t = 9\text{m}$ in the natural condition.

- As a baseline we tested the *natural* condition as in the first experiment, in which retinal size decreased with distance.

The same retinal sizes were tested in the natural and reverse conditions, but the order of presentation was reversed, i. e., the retinal size of the closest target object in the reverse condition matched the retinal size of the farthest target object in the natural condition and vice versa.

To assess distance perception, we used a blind triangulated pointing task as in the first experiment with a mixed factorial design. The independent variables were the screen distance $D_s \in \{0.8, 1.7, 2.6\}\text{m}$, the target distance $D_t \in \{1, 3, 5, 7, 9\}\text{m}$, the accommodative pattern of the virtual target object (plain, checkerboard) and the retinal size (natural, constant, reverse). We tested the accommodative pattern as a between-subjects variable, the others were within-subjects variables.

To avoid lateral preference effects, participants performed the trials once moving to the left and once to the right for each combination as in the first experiment. A virtual arrow was displayed on the floor indicating the direction of movement (see Figure 8.2). Similar to the first experiment, the screen distance was counterbalanced using a Latin Squared design, while the remaining factors were fully randomized for each screen distance. In summary, the experiment used a 3 (screen distances) $\times$ 5 (target distances) $\times$ 2 (sides) $\times$ 3 (retinal size conditions) $\times$ 2 (accommodative patterns) factorial design, resulting in a total of 180 trials.

The hypotheses for the second experiment were:

H5 More accurate distance judgments on the doubled-resolution projection wall.

H6 Effect of retinal size on distance estimation in the experiment.

H7 Less accurate distance judgments for objects with checkerboard pattern.

The experiment protocol was designed to match that of the first experiment. We collected demographic information with a questionnaire before the experiment, measured simulator sickness with SSQs before and after the experiment, and presence with a SUS questionnaire. The experiment took 1 hour per participant to complete.

8.4.2 Participants

We recruited 18 participants for our experiment. 9 participants were shown the checkerboard targets and 9 participants were shown the plain targets. 14 of the participants were male, 4 were female (ages 19−37, $M = 25.6$). The participants were either students or professionals in computer science or engineering. 16 participants were right-handed, 2 participants were left-handed. 8 participants wore glasses. None of the participants reported known vision disorders. We measured the IPD ($M = 6.4$cm, $SD = 0.32$cm) and the eye height ($M = 1.63$m, $SD = 0.08$m) of the participants. Most of the participants reported previous experience with 3D stereoscopy (rating scale $0 =$ no, $4 =$ yes, $M = 2.06$, $SD = 0.9$).

16 participants had at least 20/20 visual acuity and 2 participants had 20/30, which we measured using a standard Snellen chart before the experiment. Additionally, we displayed virtual Snellen charts with the corresponding size for users standing at a distance of 0.8m, 1.7m, or 2.6m from the projection wall. Although the resolution of the projection wall was significantly increased over the first experiment, we measured that the visual acuity of our participants differed from the real world. Participants at 0.8m distance to the projection wall reached a visual acuity of 20/60 or worse, at 1.7m distance they reached 20/50, and at 2.6m they reached 20/40 on average.

8.4.3 Results

Figure 8.6 shows the pooled results for the different conditions with the standard error of the mean over all participants. We found no significant difference between the left and right side movement and pooled the responses. The x-axes show the actual target distances $D_t \in \{1, 3, 5, 7, 9\}$m, the y-axes show the judged target distances $D_j \in \mathbb{R}_0^+$. The gray lines show the distribution of judged distances for the screen distances $D_s \in \{0.8, 1.7, 2.6\}$m.

The results were normally distributed according to a Shapiro-Wilk test at the 5% level. We analyzed the results with a mixed factorial ANOVA and Tukey multiple comparisons at the 5% significance level. Degrees of freedom were corrected using Greenhouse-Geisser

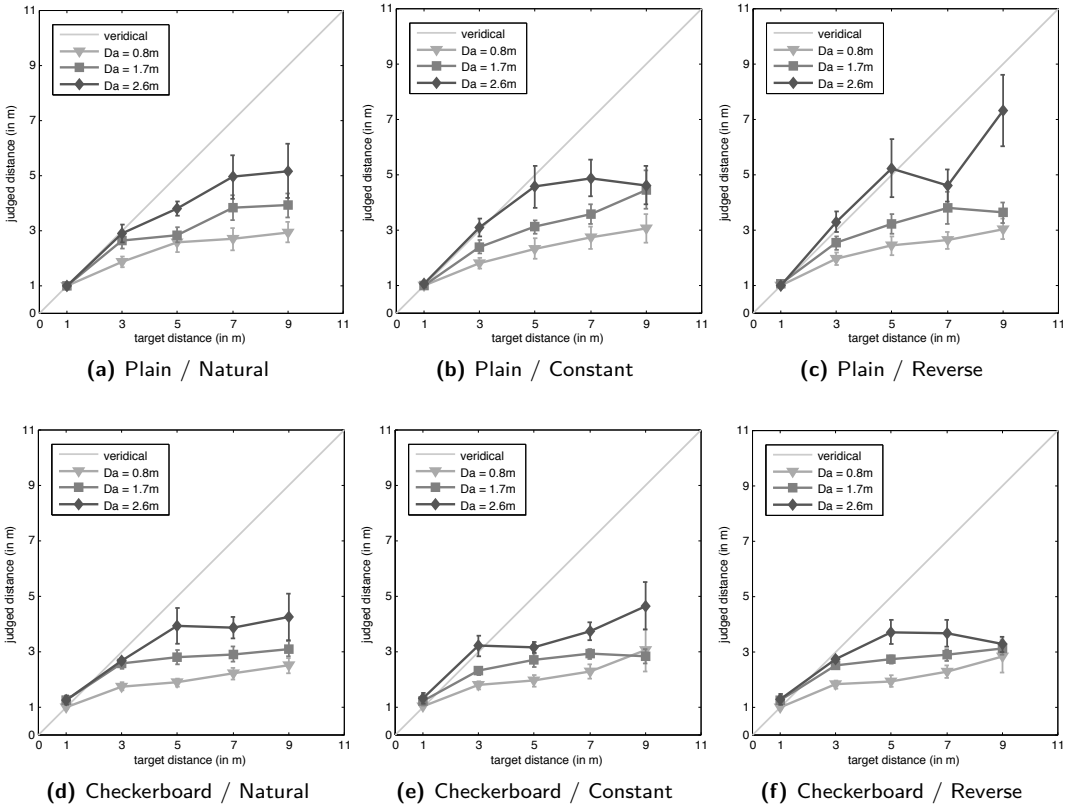
estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated.

We found a significant main effect of screen distance on absolute distance judgments, $F(2,32)=28.856$, $p<.001$, $\eta_p^2=.643$, and on relative distance judgments, $F(2,32)=30.354$, $p<.001$, $\eta_p^2=.655$. Post-hoc tests showed that the judged distances between each two screen distances were significantly different ($p<.01$). In addition, we also found a significant main effect of target distance on absolute distance judgments, $F(1.59,25.34)=60.499$, $p<.001$, $\eta_p^2=.791$, and on relative distance judgments, $F(1.29,20.62)=62.397$, $p<.001$, $\eta_p^2=.796$. Post-hoc tests showed that the judged distances between each two target distances were significantly different ($p<.05$). In contrast, we found no significant main effect of retinal size on absolute distance judgments, $F(1.44,23.01)=.970$, $p=.39$, $\eta_p^2=.057$, or on relative distance judgments, $F(1.35,21.54)=1.388$, $p=.26$, $\eta_p^2=.080$. Finally, with independent-samples t-tests we found no significant difference between checkerboard and plain target objects (between groups independent variable) on absolute distance judgments, $t(16)=1.339$, $p=.20$, or on relative distance judgments, $t(16)=.557$, $p=.59$. We found a significant interaction effect between screen distance and target distance on absolute distance judgments, $F(3.00,47.95)=8.670$, $p<.001$, $\eta_p^2=.351$, and on relative distance judgments, $F(2.88,46.08)=4.516$, $p<.01$, $\eta_p^2=.220$, as in the first experiment. No other two-way interaction effects were found.

We found a three-way interaction effect between retinal size, screen distance, and accommodative pattern on absolute distance judgments, $F(1.803,28.853)=5.224$, $p<.05$, $\eta_p^2=.246$, and a trend for relative judgments, $F(2.080,33.286)=3.089$, $p=.057$, $\eta_p^2=.162$. Post-hoc tests for the absolute distance estimation showed that the configuration for reverse, plain and $D_s=2.7$m was significantly different compared to all other conditions ($p<.05$). Furthermore, for the plain condition changes in the screen distance always resulted in significant differences in absolute distance judgments, even when comparing different levels of retinal size. For the checkerboard condition, this was not always the case. For example, for the reverse retinal size condition, only significant differences were found between $D_s=2.7$m and $D_s\in\{0.8,1.7\}$m, but no significant differences were found between $D_s=0.8$m and $D_s=1.7$m. Also, no significant differences were found between $D_s=0.8$m and $D_s=1.7$m for the checkerboard, constant retinal size condition.

Questionnaires

We measured a mean SSQ-score of $M=16.8$ ($SD=22.1$) before the experiment, and a mean SSQ-score of $M=26.0$ ($SD=30.3$) after the experiment. This increase in simulator sickness symptoms was not significant, $t(17)=1.87$, $p=.08$. The mean SUS-score for the reported sense of feeling present in the VE was $M=4.1$ ($SD=.89$), corresponding to a reasonably high level of presence [UCAS99].



(a) Plain / Natural **(b)** Plain / Constant **(c)** Plain / Reverse

(d) Checkerboard / Natural **(e)** Checkerboard / Constant **(f)** Checkerboard / Reverse

Figure 8.6: Pooled results of the judged distances for each condition in the second experiment. The x-axes show the actual distance to the target object while the y-axes show the absolute judged distance. Each plot shows the effect of the screen and target distance for each combination of accommodative pattern and retinal size.

8.4.4 Discussion

The doubled angular resolution in the second experiment did not remedy distance underestimation. In contrast to our Hypothesis H5, distance judgments were not improved over the results of the first experiment due to the increased resolution. While the resolution did not reach the same visual acuity in our participants as in the real world (cf. Section 8.4.2) it appears unlikely that a resolution with full visual acuity due to ultra high-resolution displays would result in veridical distance perception.

Our results further showed that the different retinal sizes did not have a strong effect on absolute or relative distance estimation in the experiment. We found no significant difference between the conditions with natural and constant retinal size over the tested distances. Even when we reversed the natural change in retinal size over distance we found that distance estimates were less precise but accuracy was still affected by underestimation in line with the results of the first experiment. In contrast to our Hypothesis H6, the results indicate that retinal size was not a major contributing factor for distance underestimation in this experiment.

In contrast to our Hypothesis H7, we found no significant difference in distance estimation

due to a checkerboard pattern on the target objects. We expected the checkerboard pattern to provide more luminance differences and thus enable improved accommodation responses, which in turn should consolidate the accommodation-convergence conflicts. However, the results suggest that accommodation responses might have generally been quite good in the experiments, even when we only tested a plain target object.

In summary, we conclude that the distance underestimation that we observed in the experiments was not a direct result of the resolution, nor of the plain target stimuli, and neither were they largely affected by retinal size. Additionally, we conducted the second experiment on the main wall of the projection setup, at which peripheral vision of the bezels of the projection screens was reduced, which also did not reveal an effect on distance estimation. Overall, our results consolidate the finding that screen distance and parallax are the main factors affecting distance estimation to targets up to nine meters distance from the user in large CAVE-like systems. The results of the controlled experiment provide novel insights into distance perception of objects in reduced-cue environments, and open up new research questions on how these asymmetric effects apply to rich-cue environments, in which the perception of multiple salient objects at different distances may be integrated into spatial percepts of interposition, perceived affordances and behavior.

8.5 Implications and Guidelines

The take-home message of the described experiments is that egocentric distance estimation of a user standing in a CAVE-like system depends to a large degree on the distance to the projection wall on which a virtual object is displayed as well as the stereoscopic parallax of that object. Generally, optimal distance estimation was observed when objects were displayed centered around zero parallax, but virtual objects that were displayed within the physical confines of the projection walls of a CAVE-like system also showed no distance underestimation. In contrast, our results show that egocentric distances to objects that are displayed far behind a projection wall will likely be misperceived up to magnitudes of 50%. As far as we know, this is the first time that these differences and asymmetric properties could be established for the tested egocentric distances in a large immersive projection setup.

We found no difference in distance estimation between a medium- and high-resolution projection wall, i. e., we expect that the observed underestimation cannot be remedied simply by increasing the resolution in future IPEs. We suggest the following implications and guidelines for future CAVE-like setups:

- Space is the most important characteristic for distance perception in IPEs, optimally providing about 6−7m distance around a user.

- Virtual objects with high demands on accurate spatial perception should be presented at zero or negative parallax.

8.6 Conclusion

In this chapter, we presented the first study of egocentric distance perception with view on screen distance and stereoscopic parallax in a large ten-meter immersive projection setup. Our two conducted experiments reveal that the spatial impression of rendered virtual objects greatly depends on the relative position of the user with respect to the virtual objects and the projection screens.

We found a large impact of how far a user is standing away from the projection walls, and whether a virtual object is displayed within a CAVE-like setup with negative parallax or outside with positive parallax. Our results show that neither the amount of luminance cues for accommodation responses, nor retinal size or angular resolution effected the observed main underestimation.

Distance estimation benefits from zero and negative parallaxes. Although our experiment revealed distance overestimation for close objects at negative parallaxes, the magnitude of this effect was limited compared to the large underestimation we observed for large positive parallaxes. We discussed implications and guidelines for the development of CAVE-like systems, showing that space is the most important requirement for distance perception.

Future work should focus on the effects of negative parallax on visual fatigue and novel visualization displays. First, while distance perception benefits from larger CAVE-like systems, it should be considered that objects displayed with negative parallax are likely to increase visual fatigue as compared with positive parallax [QWLL13, SWF+15]. However, no studies have yet addressed the effects of negative parallax in large CAVE-like systems. Moreover, we have to consider that virtual objects presented in CAVE-like systems might not always be displayed on one projection wall only, but might instead spread over multiple walls or the floor, e. g., when objects are not floating in mid-air, which provides multiple cues and imposes additional conflicts due to multiple different parallaxes for the same object or neighboring objects. The effects of such conflicts on visual fatigue and spatial perception remain to be investigated. Second, a possible explanation for the effects is the accommodation-convergence conflict, which is an inherent limitation of current-state stereoscopic displays, but first prototypes exist which have the potential to alleviate these conflicts in the future. In particular, light field displays [LH96] may provide a viable alternative to traditional displays once the supported depth range is improved and the computational complexity can be handled.

Bibliography

[AHJ+01] R. S. Allison, L. R. Harris, M. Jenkin, U. Jasiobedzka, and J. E. Zacher. Tolerance of temporal delay in virtual environments. In *Proceedings of IEEE Virtual Reality (VR)*, pages 247–253, 2001.

[AKLG04] M. Avraamides, R. Klatzky, J. Loomis, and R. Golledge. Use of cognitive versus perceptual heading during imagined locomotion depends on the response mode. *Psychological Science*, 15(6):403–408, 2004.

[AST09] B. Avery, C. Sandor, and B. H. Thomas. Improving spatial perception for augmented reality x-ray vision. In *Proceedings of IEEE Virtual Reality (VR)*, pages 79–82, 2009.

[ATd+10] I. V. Alexandrova, P. T. Teneva, S. de la Rosa, U. Kloos, H. H. Bülthoff, and B. J. Mohler. Egocentric distance judgments in a large screen display immersive virtual environment. In *Proceedings of the ACM Symposium on Applied Perception in Graphics and Visualization (APGV)*, pages 57–60, 2010.

[Bad12] A. D. Baddeley. Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63(1):1–29, 2012.

[BDA+05] O. Beauchet, V. Dubost, K. Aminian, R. Gonthier, and R. W. Kressig. Dual-task-related gait changes in the elderly: Does the type of cognitive task matter? *Journal of Motor Behavior*, 37:259–264, 2005.

[Ber00] A. Berthoz. *The Brain's Sense of Movement*. Harvard University Press, Cambridge, Massachusetts, 2000.

[BF07] H. Benko and S. Feiner. Balloon selection: A multi-finger technique for accurate low-fatigue 3D selection. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pages 79–86, 2007.

[BH74] A. D. Baddeley and G. J. Hitch. Working memory. *G. H. Bower (Ed.): The psychology of learning and motivation: Advances in research and theory*, 8:47—89, 1974.

[BIB+11] F. Berard, J. Ip, M. Benovoy, D. El-Shimy, J. R. Blum, and J. R. Cooperstock. Did "minority report" get it wrong? Superiority of the mouse over 3D input devices in a 3D placement task. In *Proceedings of International Conference on Human-Computer Interaction (INTERACT)*, pages 400–414, 2011.

[BIL00] R. J. Bertin, I. Israël, and M. Lappe. Perception of two-dimensional, simulated ego-motion trajectories from optic flow. *Vision Research*, 40(21):2951–2971, 2000.

[BKH97] D. A. Bowman, D. Koller, and L. F. Hodges. Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control Techniques. In *Proceedings of IEEE Virtual Reality Annual International Symposium (VRAIS)*, pages 45–52, 1997.

[BL99] F. Bremmer and M. Lappe. The use of optical velocities for distance discrimination and reproduction during visually simulated self-motion. *Experimental Brain Research*, 127(1):33–42, 1999.

[Bou99] P. Bourke. Calculating Stereo Pairs (`http://paulbourke.net/stereographics/stereorender/`), 1999.

[BPS12] G. Bruder, A. Pusch, and F. Steinicke. Analyzing effects of geometric rendering parameters on size and distance estimation in on-axis stereographics. In *Proceedings of ACM Symposium on Applied Perception (SAP)*, pages 111–118, 2012.

[BPY75] A. Berthoz, B. Pavard, and L. R. Young. Perception of linear horizontal self-motion induced by peripheral vision (linearvection): basic characteristics and visual-vestibular interactions. *Experimental Brain Research*, 23:471–489, 1975.

[BR06] J. L. Boakes and G. T. Rab. *Human Walking*, chapter Muscle Activity During Walking, pages 33–51. Lippincott Williams and Wilkins, 2006.

[BR12] O. Bottema and B. Roth. *Theoretical Kinematics*. Dover Publications, 2012.

[BRK07] M. R. K. Baumann, D. Rösler, and J. F. Krems. Situation awareness and secondary task performance while driving. *Engineering Psychology and Cognitive Ergonomics: Lecture Notes in Computer Science (LNCS)*, 4562:256–263, 2007.

[BS02] L. Bouguila and M. Sato. Virtual locomotion system for large-scale virtual environment. In *Proceedings of IEEE Virtual Reality (VR)*, pages 291–292, 2002.

[BSD+05] T. Banton, J. Stefanucci, F. Durgin, A. Fass, and D. Proffitt. The perception of walking speed in a virtual environment. *Presence: Teleoperators and Virtual Environments*, 14(4):394–406, 2005.

[BSH09] G. Bruder, F. Steinicke, and K. Hinrichs. Arch-Explore: a natural user interface for immersive architectural walkthroughs. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pages 75–82, 2009.

[BSS13a] G. Bruder, F. Steinicke, and W. Stuerzlinger. Effects of visual conflicts on 3D selection task performance in stereoscopic display environments. In *Proceedings of ACM Symposium on 3D User Interfaces (3DUI)*, pages 115–118, 2013.

[BSS13b] G. Bruder, F. Steinicke, and W. Stuerzlinger. Touching the Void Revisited: Analyses of Touch Behavior On and Above Tabletop Surfaces. In *Proceedings of International Conference on Human-Computer Interaction (INTERACT)*, pages 278–296, 2013.

[BSW11] G. Bruder, F. Steinicke, and P. Wieland. Self-motion illusions in immersive virtual reality environments. In *Proceedings of IEEE Virtual Reality (VR)*, pages 39–46, 2011.

[BSWL12] G. Bruder, F. Steinicke, P. Wieland, and M. Lappe. Tuning Self-Motion Perception in Virtual Reality with Visual Illusions. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18(7):1068–1078, 2012.

[BWB06] H. Benko, A. D. Wilson, and P. Baudisch. Precise selection techniques for multi-touch screens. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1263–1272, 2006.

[BWB08] H. Benko, A. D. Wilson, and R. Balakrishnan. Sphere: multi-touch interactions on a spherical display. In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*, pages 77–86, 2008.

[BWB+13] G. Bruder, P. Wieland, B. Bolte, M. Lappe, and F. Steinicke. Going with the flow: Modifying self-motion perception with computer-mediated optic flow. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 67–74, 2013.

[CB07] J. T. Choi and A. J. Bastian. Adaptation reveals independent control networks for human walking. *Nature Neuroscience*, 10:1055–1062, 2007.

[CGBL98] S. Chance, F. Gaunet, A. Beall, and J. Loomis. Locomotion mode affects updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration. *Presence*, 7(2):168—178, 1998.

[CHS53] J. Cohen, C. E. M. Hansel, and J. D. Sylvester. A New Phenomenon in Time Judgment. *Nature*, 172:901, 1953.

[CKC+10] L.-W. Chan, H.-S. Kao, M. Y. Chen, M.-S. Lee, J. Hsu, and Y.-P. Hung. Touching the void: Direct-touch interaction for intangible displays. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, pages 2625–2634, 2010.

[CNSD+92] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The

CAVE, Audio Visual Experience Automatic Virtual Environment. *Communications of the ACM*, 35(6):64–72, 1992.

[Cor62] T. N. Cornsweet. The Staircase-Method in Psychophysics. *The American Journal of Psychology*, 75(3):485–491, 1962.

[Cor93] S. Coren. The Lateral Preference Inventory for measurement of handedness, footedness, eyedness, and earedness: Norms for young adults. *Bulletin of the Psychonomic Society*, 31(1):1–3, 1993.

[CRWGT05] S. H. Creem-Regehr, P. Willemsen, A. A. Gooch, and W. B. Thompson. The influence of restricted viewing conditions on egocentric distance perception: implications for real and virtual environments. *Perception*, 34(2):191–204, 2005.

[Cut97] J. E. Cutting. How the eye measures reality and virtual reality. *Behavior Research Methods, Instruments, and Computers*, 29(1):27–36, 1997.

[CV95] J. E. Cutting and P. E. Vishton. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein and S. J. Rogers, editors, *Handbook of perception and cognition*, volume Perception of space and motion, pages 69–117. Academic Press, San Diego, CA, USA, 1995.

[DAK07] J. C. Dean, N. B. Alexander, and A.D. Kuo. The effect of lateral stabilization on walking in young and old adults. *IEEE Transactions on Biomedical Engineering (TBME)*, 54(11):1919–1926, 2007.

[DB78] J. Dichgans and T. Brandt. Visual Vestibular Interaction: Effects on Self-Motion Perception and Postural Control. In R. Held, H.W. Leibowitz, and H.L. Teuber, editors, *Perception. Handbook of Sensory Physiology*, volume 8, pages 755–804, Berlin, Heidelberg, New York, 1978. Springer.

[DCJH13] B. R. De Araújo, G. Casiez, J. A. Jorge, and M. Hachet. Mockup Builder: 3D modeling on and above the surface. *Computers & Graphics*, 37:165–178, 2013.

[DGS05] F. H. Durgin, K. Gigone, and R. Scott. Perception of visual speed while moving. *Journal of Experimental Psychology: Human Perception and Performance*, 31(2):339–353, 2005.

[DKK07] A. Y. Dvorkin, R. V. Kenyon, and E. A. Keshner. Reaching within a dynamic virtual environment. *Journal of NeuroEngineering and Rehabilitation*, 4(23):182–186, 2007.

[dKOD08] J.-B. de la Rivière, C. Kervégant, E. Orvain, and N. Dittlo. CubTile: A multi-touch cubic interface. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology (VRST)*, pages 69–72, 2008.

[dlRDO+10] J.-B. de la Rivière, N. Dittlo, E. Orvain, C. Kervégant, M. Courtois, and T. Da Luz. iliGHT 3D touch: A multiview multitouch surface for 3D content visualization and viewpoint sharing. In *Proceedings of ACM International Conference on Interactive Tabletops and Surfaces (ITS)*, pages 312–312, 2010.

[DM96] D. Drascic and P. Milgram. Perceptual issues in augmented reality. In M. T. Bolas, S. S. Fisher, and J. O. Merritt, editors, *Stereoscopic Displays and Virtual Reality Systems III*, volume 2653, pages 123–134. SPIE, Bellingham, WA, USA, 1996.

[DPF+05] F. H. Durgin, A. Pelah, L. F. Fox, J. Lewis, R. Kane, and K. A. Walley. Self-motion perception during locomotor recalibration: more than meets the eye. *Journal of Experimental Psychology: Human Perception and Performance*, 31(3):398–419, 2005.

[DUV02] S. M. Domhoefer, P. J. A. Unema, and B. M. Velichkovsky. Blinks, blanks and saccades: How blind we really are for relevant visual events. *Progress in Brain Research*, 140:119–131, 2002.

[DW93] C. J. Duffy and R. H. Wurtz. An Illusory Transformation of Optic Flow Fields. *Vision Research*, 33(11):1481–1490, 1993.

[EB04] M. O. Ernst and H. H. Bülthoff. Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4):162–169, 2004.

[Efr70] R. Efron. Effect of stimulus duration on perceptual onset and offset latencies. *Perception & Psychophysics*, 8(4):231–234, 1970.

[Ern06] M. O. Ernst. *Human Body Perception From The Inside Out*, chapter A Bayesian View on Multimodal Cue Integration, pages 105–131. New York, NY: Oxford University Press, 2006.

[FAH91] W. T. Freeman, E. H. Adelson, and D. J. Heeger. Motion Without Movement. *ACM SIGGRAPH Computer Graphics*, 25(4):27–30, 1991.

[FC88] K. Fisher and K. J. Ciuffreda. Accommodation and apparent distance. *Perception*, 17(5):609–621, 1988.

[Fer08] J. Ferwerda. SIGGRAPH Core: Psychophysics 101: How to run perception experiments in computer graphics. In *Proceedings of ACM Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 2008.

[FHW11] I. Fründ, N. Haenel, and F. Wichmann. Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, 11(6):1–19, 2011.

[Fit54] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47:381–391, 1954.

[FLD97] S. S. Fukusima, J. M. Loomis, and J. A. Da Silva. Visual perception of egocentric distance as assessed by triangulation. *Journal of Expimental Psychology: Human Perreption and Pedormance*, 23(1):86–100, 1997.

[GB08] P. Guerin and B.G. Bardy. Optical modulation of locomotion and energy expenditure at preferred transition speed. *Experimental Brain Research*, 189(4):393–402, 2008.

[GC93] A. S. Gevins and B. C. Cutillo. Neuroelectric evidence for distributed processing in human working memory. *Electroencephalography and Clinical Neurophysiology*, 87:128–143, 1993.

[GCE08] L. Geniva, R. Chua, and J. T. Enns. Attention for perception and action: task interference for action planning, but not for online control. *Experimental Brain Research*, 185(4):709–717, 2008.

[GDP+10] T. Y. Grechkin, T. Dat Nguyen, J. M. Plumert, J. F. Cremer, and J. K. Kearney. How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Transactions on Applied Perception (TAP)*, 7:1–18, 2010.

[GGS10] R. Gibb, R. Gray, and L. Scharff. *Aviation Visual Perception: Research, Misperception and Mishaps*. Ashgate, 2010.

[Gib50] J. J. Gibson. *The Perception of the Visual World*. Riverside Press, Cambridge, England, 1950.

[Gie97] M. A. Giese. *A Dynamical Model for the Perceptual Organization of Apparent Motion*. PhD thesis, Ruhr-University Bochum, 1997.

[Gil51] A. S. Gilinsky. Perceived size and distance in visual space. *Psychological Review*, 58:460–482, 1951.

[GL99] A. Grigo and M. Lappe. Dynamical use of different sources of information in heading detection from retinal flow. *Journal of the Optical Society of America A*, 16(9):2079–2091, 1999.

[GNRH05] H. Groenda, F. Nowak, P. Rößler, and U. D. Hanebeck. Telepresence techniques for controlling avatar motion in first person games. In *Proceedings of the International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN)*, pages 44–53, 2005.

[Gol09] E. B. Goldstein. *Sensation and Perception*. Cengage Learning EMEA, 2009.

[Gro08] S. Grondin. *Psychology of Time*. Emerald Group Publishing, 2008.

[GW07] T. Grossman and D. Wigdor. Going deeper: A taxonomy of 3D on the tabletop. In *Proceedings of IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP)*, pages 137–144, 2007.

[HAH02] R. Hosman, S. Advani, and N. Haeck. Integrated design of flight simulator motion cueing systems. In *Proceedings of the Royal Aeronautical Society Conference on Flight Simulation*, pages 1–12, 2002.

[HBCd11] M. Hachet, B. Bossavit, A. Cohe, and J.-B. de la Rivière. Toucheo: multitouch and stereo combined in a seamless workspace. In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*, pages 587–592, 2011.

[HCC07] M. Hancock, S. Carpendale, and A. Cockburn. Shallow-depth 3D interaction: design and evaluation of one-, two- and three-touch techniques. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1147–1156, 2007.

[HCOB10] R. T. Held, E. A. Cooper, J. F. O'Brien, and M. S. Banks. Using blur to affect perceived distance and size. *ACM Transactions on Graphics (TOG)*, 29(2):1–16, 2010.

[HCRC88] A. D. Hall, J. B. Cunningham, R. P. Roache, and J. W. Cox. Factors affecting performance using touchentry systems: Tactual recognition fields and system accuracy. *Journal of Applied Psychology*, 4:711–720, 1988.

[HGAB08] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, 8(3):1–30, 2008.

[HIW+09] O. Hilliges, S. Izadi, A. D. Wilson, S. Hodges, A. Garcia-Mendoza, and A. Butz. Interactions in the air: Adding further depth to interactive tabletops. In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*, pages 139–148, 2009.

[HK31] H. Helson and S. M. King. The Tau effect: An example of psychological relativity. *Journal of Experimental Psychology*, 14:202–217, 1931.

[Hol02] J. Hollerbach. Locomotion interfaces. In *Handbook of Virtual Environments: Design, Implementation, and Applications*, pages 239–254. Lawrence Erlbaum Associates, 2002.

[HY95] Y. Hermush and Y. Yeshurun. Spatial-gradient limit on perception of multiple motion. *Perception*, 24(11):1247–1256, 1995.

[IAR06] V. Interrante, L. Anderson, and B. Ries. Distance perception in immersive virtual environments, revisited. In *Proceedings of IEEE Virtual Reality (VR)*, pages 3–10, 2006.

[IHT06] H. Iwata, Y. Hiroaki, and H. Tomioka. Powered shoes. *SIGGRAPH Emerging Technologies*, page 28, 2006.

[iMU13] iMUTS - Interscopic Multi-Touch-Surfaces (`http://imuts.uni-muenster.de/`), 2013.

[InS13] InSTInCT - Touch-based interfaces for Interaction with 3D Content (`http://anr-instinct.cap-sciences.net/`), 2013.

[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.

[IRA07] V. Interrante, B. Riesand, and L. 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 (3DUI)*, pages 167–170, 2007.

[IRLA07] V. Interrante, B. Ries, J. Lindquist, and L. Anderson. Elucidating the Factors that can Facilitate Veridical Spatial Perception in Immersive Virtual Environments. In *Proceedings of IEEE Virtual Reality (VR)*, pages 11–18, 2007.

[IYFN05] H. Iwata, H. Yano, H. Fukushima, and H. Noma. CirculaFloor. *IEEE Computer Graphics and Applications*, 25(1):64–67, 2005.

[JBM99] A. Johnston, C. P. Benton, and P. W. McOwan. Induced motion at texture-defined motion boundaries. *Proceedings of the Royal Society B: Biological Sciences*, 266(1436):2441–2450, 1999.

[JH82] B. Jones and Y. L. Huang. Space-time dependencies in psychophysical judgment of extent and duration: Algebraic models of the tau and kappa effects. *Psychological Bulletin*, 91(1):128–142, 1982.

[JSKB12] J. A. Jones, E. A. Suma, D. M. Krum, and M. Bolas. Comparability of Narrow and Wide Field-Of-View Head-Mounted Displays for Medium-Field Distance Judgments. In *Proceedings of ACM Symposium on Applied Perception (SAP)*, page 119, 2012.

[JSSE11] J. A. Jones, J. E. Swan II, G. Singh, and S. R. Ellis. Peripheral visual information and its effect on distance judgments in virtual and augmented environments. In *Proceedings of ACM Symposium on Applied Perception in Graphics and Visualization (APGV)*, pages 29–35, 2011.

[JW08] C. J. Jerome and B. G. Witmer. The Perception and Estimation of Egocentric Distance in Real and Augmented Reality Environments. Technical Report 1230, U.S. Army Research Institute for the Behavioral and Social Sciences, 2008.

[KBB+12] F. Kellner, B. Bolte, G. Bruder, U. Rautenberg, F. Steinicke, M. Lappe, and R. Koch. Geometric Calibration of Head-Mounted Displays and its Effects on Distance Estimation. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18(4):589–596, 2012.

[KBMF05] L. Kohli, E. Burns, D. Miller, and H. Fuchs. Combining passive haptics with redirected walking. In *Proceedings of ACM Conference on Augmented Tele-Existence (ICAT)*, volume 157, pages 253–254, 2005.

[KLBL93] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, 3(3):203–220, 1993.

[Kle01] S. Klein. Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, 63(8):1421–1455, 2001.

[KM13] A. Kramer and M. Merrow. *Circadian Clocks*. Handbook of Experimental Pharmacology 217. Springer Verlag, 2013.

[KSF10] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual Issues in Augmented Reality Revisited. In *Proceedings of IEEE Symposium on Mixed and Augmented Reality (ISMAR)*, pages 3–12, 2010.

[KSS+09] E. Klein, J. E. Swan, G. S. Schmidt, M. A. Livingston, and O. G. Staadt. Measurement protocols for medium-field distance perception in large-screen immersive displays. In *Proceedings of IEEE Virtual Reality (VR)*, pages 107–113, 2009.

[KTCR09] S. A. Kuhl, W. B. Thompson, and S. H. Creem-Regehr. HMD calibration and its effects on distance judgments. *ACM Transactions on Applied Perception (TAP)*, 6(3):1–19, 2009.

[LAPD06] M. Q. Liu, F. C. Anderson, M. G. Pandy, and S. L. Delp. Muscles that support the body also modulate forward progression during walking. *Journal of Biomechanics*, 39:2623–2630, 2006.

[LBv99] M. Lappe, F. Bremmer, and A.V. van den Berg. Perception of self-motion from visual flow. *Trends in Cognitive Sciences*, 3(9):329–336, 1999.

[LCE08] G. Liu, R Chua, and J. T. Enns. Attention for perception and action: Task interference for action planning, but not for online control. *Experimental Brain Research*, 185:709–717, 2008.

[LDST13] M. A. Livingston, A. Dey, C. Sandor, and B. H. Thomas. *Human Factors in Augmented Reality Environments*, chapter Pursuit of "X-Ray Vision" for Augmented Reality, pages 67–107. Springer New York, 2013.

[Lee01] M. R. Leek. Adaptive procedures in psychophysical research. *Perception & Psychophysics*, 63(8):1279–1292, 2001.

[LH96] M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of ACM Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 31–42, 1996.

[LJH07] M. Lappe, M. R. Jenkin, and L. R. Harris. Travel distance estimation from visual motion by leaky path integration. *Experimental Brain Research*, 180:35–48, 2007.

[LK03] J. M. Loomis and J. M. Knapp. Visual perception of egocentric distance in real and virtual environments. In L. Hettinger and M. Haas, editors, *Virtual and Adaptive Environments*, pages 21–46. Erlbaum, 2003.

[LKG+93] J. M. Loomis, R. L. Klatzky, R. G. Golledge, J. G. Cicinelli, J. W. Pellegrino, and P. A. Fry. Nonvisual navigation by blind and sighted: Assessment of path integration ability. *Journal of Experimental Psychology: General*, 122(1):73–91, 1993.

[LL75] D. N. Lee and J. R. Lishman. Visual proprioceptive control of stance. *Journal of Human Movement Studies*, 1:87–95, 1975.

[Lou07] L. Lou. Apparent Afterimage Size, Emmerts Law, and Oculomotor Adjustment. *Perception*, 36(8):1214–1228, 2007.

[LP08] J. M. Loomis and J. W. Philbeck. *Embodiment, Ego-Space, and Action*, chapter Measuring Spatial Perception with Spatial Updating and Action, pages 1–43. Psychology Press, 2008.

[LR95] M. Lappe and J. P. Rauschecker. An illusory transformation in a model of optic flow processing. *Vision Research*, 35:1619–1631, 1995.

[Mat06] G. Mather. Two-stroke: A new illusion of visual motion based on the time course of neural responses in the human visual system. *Vision Research*, 46(13):2015–2018, 2006.

[MBK06] S. O. Murray, H. Boyaci, and D. Kersten. The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience*, 9(3):429–434, 2006.

[MC04] N. A. Macmillan and C. D. Creelman. *Detection Theory: A User's Guide*. Psychology Press, 2004.

[MCG10] A. Martinet, G. Casiez, and G. Grisoni. The design and evaluation of 3D positioning techniques for multi-touch displays. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pages 115–118, 2010.

[MCM14] E. Marsh, J.-R. Chardonnet, and F. Merienne. Virtual distance estimation in a CAVE. In *Proceedings of Spatial Cognition IX International Conference*, pages 397–412, 2014.

[MCWB07] B. J. Mohler, J. L. Campos, M. B. Weyel, and H. H. Bülthoff. Gait parameters while walking in a head-mounted display virtual environment and the real world. In *Proceedings of Eurographics Symposium on Virtual Environments (EGVE)*, pages 85–88, 2007.

[MD05] R. Messing and F. H. Durgin. Distance perception and the visual horizon in head-mounted displays. *ACM Transactions on Applied Perception (TAP)*, 3(2):234–250, 2005.

[MHG12] J. P. McIntire, P. R. Havig, and E. E. Geiselman. What is 3D good for? A review of human performance on stereoscopic 3D displays. *Proceedings of the SPIE, Head- and Helmet-Mounted Displays XVII*, 8383:1–13, 2012.

[MHM+09] D. Mahajan, F.-C. Huang, W. Matusik, R. Ramamoorthi, and P. Belhumeur. Moving gradients: A path-based method for plausible image interpolation. In *Proceedings of ACM Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 42:1–42:11, 2009.

[MHZS13] W. E. Marsh, T. Hantel, C. Zetzsche, and K. Schill. Is the user trained? Assessing performance and cognitive resource demands in the Virtusphere. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pages 15–22, 2013.

[MI08] I. S. MacKenzie and P. Isokoski. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1633–1636, 2008.

[MISH00] M. E. Morris, R. Iansek, F. Smithson, and F. Huxham. Postural instability in parkinson's disease: A comparison with and without an added task. *Gait Posture*, 12:205–216, 2000.

[MKDO13] W. E. Marsh, J. W. Kelly, V. J. Dark, and J. H. Oliver. Cognitive demands of semi-natural virtual locomotion. *Presence: Teleoperators and Virtual Environments*, 22(3):216–234, 2013.

[MMD+87] C. L. MacKenzie, R. G. Marteniuka, C. Dugasa, D. Liskea, and B. Eickmeiera. Three-dimensional movement trajectories in Fitts' task: Implications for control. *Journal of Experimental Psychology A*, 39(4):629–647, 1987.

[MOB03] A. P. Mapp, H. Ono, and R. Barbeito. What does the dominant eye dominate? A brief and somewhat contentious review. *Perception & Psychophysics*, 65(2):310–317, 2003.

[Moh07] B. J. Mohler. *The Effect of Feedback Within a Virtual Environment on Human Distance Perception and Adaptation*. ProQuest, 2007.

[MRE85] M. McGreevy, C. Ratzlaff, and S. Ellis. Virtual space and two-dimensional effects in perspective displays. In *Proceedings of Annual Conference on Manual Control*, pages 29:1–29:14, 1985.

[MT99] M. Mon-Williams and J. R. Tresilian. The size-distance paradox is a cognitive phenomenon. *Experimental Brain Research*, 126:578–582, 1999.

[MTCR+07] B. J. Mohler, W. B. Thompson, S. H. Creem-Regehr, H. L. Pick, Jr., and W. H. Warren, Jr. Visual flow influences gait transition speed and preferred walking speed. *Experimental Brain Research*, 181(2):221–228, 2007.

[MTCRW07] B. J. Mohler, W. B. Thompson, S. H. Creem-Regehr, and P. Willemsen. Calibration of locomotion resulting from visual motion in a treadmill-based virtual environment. *ACM Transactions on Applied Perception (TAP)*, 4(1):1–15, 2007.

[MTHG10] B. Mao, Z. Tian, H. Huang, and Z. Gao. *Traffic and Transportation Studies*. American Society of Civil Engineers (ASCE), 2010.

[MTSH+10] C. Müller-Tomfelde, J. Schöning, J. Hook, T. Bartindale, D. Schmidt, P. Oliver, F. Echtler, N. Motamedi, P. Brandl, and U. Zadow. Building interactive multi-touch surfaces. In C. Müller-Tomfelde, editor, *Tabletops - Horizontal Interactive Displays*, Human-Computer Interaction Series, pages 27–49. Springer, 2010.

[NHS04] N. Nitzsche, U. Hanebeck, and G. Schmidt. Motion compression for telepresent walking in large target environments. *Presence*, 13(1):44–60, 2004.

[NSE+11] C. T. Neth, J. L. Souman, D. Engel, U. Kloos, H. H. Bülthoff, and B. J. Mohler. Velocity-dependent dynamic curvature gain for redirected walking. In *Proceedings of IEEE Virtual Reality (VR)*, pages 151–158, 2011.

[NZL+10] N. K. Nadkarni, K. Zabjek, B. Lee, W. E. McIlroy, and S. E. Black. Effect of working memory and spatial attention tasks on gait in healthy young and older adults. *Motor Control*, 14(2):195–210, 2010.

[Pal99] S. E. Palmer. *Vision Science: Photons to Phenomenology*. MIT Press, 1999.

[PdK+13] I. V. Piryankova, S. de la Rosa, U. Kloos, H. H. Bülthoff, and B. J. Mohler. Egocentric distance perception in large screen immersive displays. *Elsevier Displays*, 34(2):153–164, 2013.

[PFC+97] J. Pierce, A. Forsberg, M. Conway, S. Hong, R. Zeleznik, and M. Mine. Image Plane Interaction Techniques in 3D Immersive Environments. In *Proceedings of ACM Symposium on Interactive 3D Graphics (i3D)*, pages 39–44, 1997.

[PFFB90] J. Pailhous, A. M. Fernandez, M. Fluckiger, and B. Bamberger. Unintentional modulations of human gait by optical flow. *Behavioural Brain Research*, 38:275–281, 1990.

[PFW11] T. C. Peck, H. Fuchs, and M. C. Whitton. An evaluation of navigational ability comparing redirected free exploration with distractors to walking-in-place and joystick locomotion interfaces. In *Proceedings of IEEE Virtual Reality (VR)*, pages 56–62, 2011.

[PGG+10] S. Palmisano, B. Gillam, D. G. Govan, R. S. Allison, and J. M. Harris. Stereoscopic perception of real depths at large distances. *Journal of Vision*, 10(6):1–16, 2010.

[PGJ01] C. C. Pagano, R. P. Grutzmacher, and J. C. Jenkins. Comparing verbal and reaching responses to visually perceived egocentric distances. *Ecological Psychology*, 13(3):197–226, 2001.

[PH12] S. J. Pritchard and S. T. Hammett. The effect of luminance on simulated driving speed. *Vision Research*, 52(1):54–60, 2012.

[PHH12] D. Pyryeskin, M. Hancock, and J. Hoey. Comparing elicited gestures to designer-created gestures for selection above a multitouch surface. In *Proceedings of ACM International Conference on Interactive Tabletops and Surfaces (ITS)*, pages 1–10, 2012.

[Pie93] B. Pierscionek. In vitro alteration of human lens curvatures by radial stretching. *Experimental Eye Research*, 57:629–635, 1993.

[PKC04] J. M. Plumert, J. K. Kearney, and J. F. Cremer. Distance perception in real and virtual environments. In *Proceedings of ACM Symposium on Applied Perception in Graphics and Visualization (APGV)*, pages 27–34, 2004.

[PRI+09] L. Phillips, B. Ries, V. Interrante, M. Kaeding, and L. Anderson. Distance Perception in NPR Immersive Virtual Environments, Revisited. In *Proceedings of ACM Symposium on Applied Perception in Graphics and Visualization (APGV)*, pages 11–14, 2009.

[PWF08] T. C. Peck, M. C. Whitton, and H. Fuchs. Evaluation of reorientation techniques for walking in large virtual environments. In *Proceedings of IEEE Virtual Reality (VR)*, pages 121–128, 2008.

[PWS88] R. L. Potter, L. J. Weldon, and B. Shneiderman. Improving the accuracy of touch screens: an experimental evaluation of three strategies. In *Proceedings*

of ACM Conference on Human Factors in Computing Systems (CHI), pages 27–32, 1988.

[QWLL13] Z. Qian, X. Wang, C. Lan, and W. Li. Analysis of fatigue with 3D TV based on EEG. In *Proceedings of IEEE International Conference on Orange Technologies (ICOT)*, pages 306–309, 2013.

[RATY90] J. J. Rieser, D. H. Ashmead, C. R. Taylor, and G. A. Youngquist. Visual perception and the guidance of locomotion without vision to previously seen targets. *Perception*, 19:675–689, 1990.

[Raz05] S. Razzaque. *Redirected Walking*. PhD thesis, University of North Carolina, Chapel Hill, 2005.

[RDD+05] J. P. Regnaux, D. David, O. Daniel, D. B. Smail, M. Combeaud, and B. Bussel. Evidence for cognitive processes involved in the control of steady state of walking in healthy subjects and after cerebral damage. *Neurorehabilitation and Neural Repair*, 19(2):125–132, 2005.

[RDH09] J. L. Reisman, P. L. Davidson, and J. Y. Han. A screen-space formulation for 2D and 3D direct manipulation. In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*, pages 69–78, 2009.

[RGK09] M.-E. Roussel, S. Grondin, and P. Killeen. Spatial effects on temporal categorization. *Perception*, 38(5):748–762, 2009.

[RH94] W. Robinett and R. Holloway. The visual display transformation for virtual reality. Technical report, University of North Carolina at Chapel Hill, 1994.

[Rie10] B. E. Riecke. *Virtual Reality*, chapter Compelling Self-Motion Through Virtual Environments Without Actual Self-Motion - Using Self-Motion Illusions ("Vection") to Improve User Experience in VR, pages 149–176. InTech, 2010.

[RKW01] S. Razzaque, Z. Kohn, and M. C. Whitton. Redirected Walking. In *Proceedings of Eurographics*, pages 289–294, 2001.

[RL09] R. A. Ruddle and S. Lessels. The benefits of using a walking interface to navigate virtual environments. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 16(1):5:1–5:18, 2009.

[ROC97] R. A. Rensink, J. K. O'Regan, and J. J. Clark. To See or Not to See: The Need for Attention to Perceive Changes in Scenes. *Psychological Science*, 8(5):368–373, 1997.

[RPAG95] J. J. Rieser, H. L. Pick, D. H. Ashmead, and A. E. Garing. Calibration of human locomotion and models of perceptual-motor organization. *Journal Experimental Psychology Human Perceptual Performance*, 21:480–497, 1995.

[RSA+06] B. E. Riecke, J. Schulte-Pelkum, M. Avraamides, M. von der Heyde, and H. H. Bülthoff. Cognitive factors can influence self-motion perception (vection) in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 3(3):194–216, 2006.

[RVB10] R. A. Ruddle, E. P. Volkova, and H. H. Bülthoff. Walking improves your cognitive map in environments that are large-scale and large in extent. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(2):10:1–10:22, 2010.

[RVH13] R. S. Renner, B. B. Velichkovsky, and J. R. Helmert. The perception of egocentric distances in virtual environments - a review. *ACM Computing Surveys*, 46(2):1–40, 2013.

[SBH+09] F. Steinicke, G. Bruder, K. Hinrichs, J. Jerald, H. Frenz, and M. Lappe. Real walking through virtual environments by redirection techniques. *Journal of Virtual Reality and Broadcasting (JVRB)*, 6(2):1–16, 2009.

[SBHW10] F. Steinicke, G. Bruder, K. H. Hinrichs, and P. Willemsen. Change blindness phenomena for stereoscopic projection systems. In *Proceedings of IEEE Virtual Reality (VR)*, pages 187–194, 2010.

[SBJ+10] F. Steinicke, G. Bruder, J. Jerald, H. Frenz, and M. Lappe. Estimation of detection thresholds for redirected walking techniques. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 16(1):17–27, 2010.

[SBK11] F. Steinicke, G. Bruder, and S. Kuhl. Realistic Perspective Projections for Virtual Objects and Environments. *ACM Transactions on Graphics (TOG)*, 30(5):112:1–112:10, 2011.

[SBL+11] F. Steinicke, G. Bruder, M. Lappe, S. Kuhl, P. Willemsen, and K. H. Hinrichs. Natural perspective projections for head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 17(7):888–899, 2011.

[SBRH08] F. Steinicke, G. Bruder, T. Ropinski, and K. H. Hinrichs. Moving Towards Generally Applicable Redirected Walking. In *Proceedings of Virtual Reality International Conference (VRIC)*, pages 15–24, 2008.

[SBS+12] E. A. Suma, G. Bruder, F. Steinicke, D. M. Krum, and M. Bolas. A taxonomy for deploying redirection techniques in immersive virtual environments. In *Proceedings of IEEE Virtual Reality (VR)*, pages 43–46, 2012.

[SCRT+15] J. K. Stefanucci, S. H. Creem-Regehr, W. B. Thompson, D. A. Lessard, and M. N. Geuss. Evaluating the accuracy of size perception on screen-based displays: Displayed objects appear smaller than real objects. *Journal of Experimental Psychology: Applied*, pages 215–223, 2015.

[SDO06] W. Stuerzlinger, D. Dadgari, and J.-Y. Oh. Reality-based object movement techniques for 3D. In *Proceedings of ACM CHI Workshop: What is the Next Generation of Human-Computer Interaction?*, pages 1–4, 2006.

[SES99] D. Schmalstieg, L. M. Encarnação, and Z. Szalavári. Using transparent props for interaction with the virtual table. In *Proceedings of ACM Symposium on Interactive 3D Graphics (i3D)*, pages 147–153, 1999.

[SFR+10] E. Suma, S. Finkelstein, M. Reid, S. Babu, A. Ulinski, and L. Hodges. Evaluation of the cognitive effects of travel technique in complex real and virtual environments. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 16(4):690–702, 2010.

[SGPB04] J.-C. Sarrazin, M.-D. Giraudo, J. Pailhous, and R. J. Bootsma. Dynamics of balancing space and time in memory: Tau and Kappa effects revisited. *Journal of Experimental Psychology: Human Perception & Performance*, 30:411–430, 2004.

[SJK+07] J. E. Swan II, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman. Egocentric depth judgments in optical, see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 13(3):429–442, 2007.

[SKFB11] E. Suma, D. Krum, S. Finkelstein, and M. Bolas. Effects of redirection on spatial orientation in real and virtual environments. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pages 35–38, 2011.

[Sla09] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society (B)*, 364(1535):3549–3557, 2009.

[SLF+12] E. A. Suma, Z. Lipps, S. Finkelstein, D. M. Krum, and M. Bolas. Impossible Spaces: Maximizing Natural Walking in Virtual Environments with Self-Overlapping Architecture. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 18(4):555–564, 2012.

[SM90] B. Sivak and C. L. MacKenzie. Integration of visual information and motor output in reaching and grasping: The contributions of peripheral and central vision. *Neuropsychologia*, 28(10):1095–1116, 1990.

[SRS+11] J. L. Souman, P. Robuffo Giordano, M. Schwaiger, I. Frissen, T. Thümmel, H. Ulbrich, A. De Luca, H. H. Bülthoff, and M. Ernst. CyberWalk: Enabling unconstrained omnidirectional walking through virtual environments. *ACM Transactions on Applied Perception (TAP)*, 8(4):1–22, 2011.

[SSG00] S. Steinman, B. Steinman, and R. Garzia. *Foundations of Binocular Vision: A Clinical Perspective*. McGraw-Hill Medical, 1st edition, 2000.

[SSR98] R. J. Snowden, N. Stimpson, and R. A. Ruddle. Speed perception fogs up as visibility drops. *Nature*, 392(6675):392–450, 1998.

[SSV+09] J. Schöning, F. Steinicke, D. Valkov, A. Krüger, and K. H. Hinrichs. Bimanual interaction with interscopic multi-touch surfaces. In *Proceedings of International Conference on Human-Computer Interaction (INTERACT)*, pages 40–53, 2009.

[STU07] M. Schwaiger, T. Thümmel, and H. Ulbrich. Cyberwalk: Implementation of a Ball Bearing Platform for Humans. In *Proceedings of International Conference on Human-Computer Interaction: Interaction Platforms and Techniques (HCI)*, pages 926–935, 2007.

[SUS95] M. Slater, M. Usoh, and A. Steed. Taking steps: The influence of a walking metaphor on presence in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 2(3):201–219, 1995.

[SVCL13] F. Steinicke, Y. Visell, J. Campos, and A. Lecuyer. *Human Walking in Virtual Environments: Perception, Technology, and Applications*. Springer, 2013.

[SVH11] S. Strothoff, D. Valkov, and K. H. Hinrichs. Triangle Cursor: Interactions with objects above the tabletop. In *Proceedings of ACM International Conference on Interactive Tabletops and Surfaces (ITS)*, pages 111–119, 2011.

[SWF+15] W. Song, D. Weng, D. Feng, Y. Li, Y. Liu, and Y. Wang. Evaluating visual discomfort in stereoscopic projection-based CAVE system with a close viewing distance. In *Proceedings of SPIE 9495, Three-Dimensional Imaging, Visualization, and Display*, pages 1–7, 2015.

[SZM10] Z. Shi, H. Zou, and H. J. Müller. *Advances in Haptics*, chapter Temporal Perception of Visual-Haptic Events in Multimodal Telepresence System, pages 437–449. InTech, 2010.

[TFCRS11] W. B. Thompson, R. W. Fleming, S. H. Creem-Regehr, and J. K. Stefanucci. *Visual perception from a computer graphics perspective*. CRC Press, 2011.

[TS11] R. J. Teather and W. Stuerzlinger. Pointing at 3D targets in a stereo head-tracked virtual environment. In *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pages 87–94, 2011.

[TS13] R. J. Teather and W. Stuerzlinger. Pointing at 3D target projections with one-eyed and stereo cursors. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, pages 159–168, 2013.

[TWG+04] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall. Does the quality of the computer graphics matter when

judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments*, 13(5):560–571, 2004.

[UAW+99] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks, Jr. Walking > Walking-in-Place > Flying, in Virtual Environments. In *Proceedings of ACM Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 359–364, 1999.

[UCAS99] M. Usoh, E. Catena, S. Arman, and M. Slater. Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9(5):497–503, 1999.

[VB04] D. Vogel and R. Balakrishnan. Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*, pages 137–146, 2004.

[VFML04] A. Viau, A. G. Feldman, B. J McFadyen, and M. F. Levin. Reaching in reality and virtual reality: A comparison of movement kinematics in healthy subjects and in adults with hemiparesis. *Journal of NeuroEngineering and Rehabilitation*, 1(11):1–7, 2004.

[VGH12] D. Valkov, A. Giesler, and K. H. Hinrichs. Evaluation of depth perception for touch interaction with stereoscopic rendered objects. In *Proceedings of ACM International Conference on Interactive Tabletops and Surfaces (ITS)*, pages 21–30, 2012.

[vR01] M. von der Heyde and B. E. Riecke. How to cheat in motion simulation - comparing the engineering and fun ride approach to motion cueing. Technical Report 89, Max Planck Institute for Biological Cybernetics, Tübingen, Germany, 2001.

[VSB+10] D. Valkov, F. Steinicke, G. Bruder, K. H. Hinrichs, J. Schöning, F. Daiber, and A. Krüger. Touching floating objects in projection-based virtual reality environments. In *Proceedings of the Joint Virtual Reality Conference of EuroVR - EGVE - VEC*, pages 17–24, 2010.

[VSBH11] D. Valkov, F. Steinicke, G. Bruder, and K. H. Hinrichs. 2D touching of 3D stereoscopic objects. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1353–1362, 2011.

[War98] W. H. Warren, Jr. Visually Controlled Locomotion: 40 Years Later. *Ecological Psychology*, 10(3-4):177–219, 1998.

[WC11] A. Wedel and D. Cremers. *Stereo Scene Flow for 3D Motion Analysis*, chapter Optical Flow Estimation, pages 5–34. Springer London, 2011.

[WCCRT09] P. Willemsen, M. B. Colton, S. H. Creem-Regehr, and W. B. Thompson. The Effects of Head-Mounted Display Mechanical Properties and Field-of-View on Distance Judgments in Virtual Environments. *ACM Transactions on Applied Perception (TAP)*, 2(6):1–14, 2009.

[WCF+05] M. Whitton, J. Cohn, P. Feasel, S. Zimmons, S. Razzaque, B. Poulton, and B. McLeod und F. Brooks. Comparing VE locomotion interfaces. In *Proceedings of IEEE Virtual Reality (VR)*, pages 123–130, 2005.

[Wer94] A. H. Wertheim. Motion perception during self-motion, the direct versus inferential controversy revisited. *Behavioral and Brain Science*, 17(2):293–355, 1994.

[WGTCR08] P. Willemsen, A. A. Gooch, W. B. Thompson, and S. H. Creem-Regehr. Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoperators and Virtual Environments*, 17(1):91–101, 2008.

[WIH+08] A. D. Wilson, S. Izadi, O. Hilliges, A. Garcia-Mendoza, and D. Kirk. Bringing physics to the surface. In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*, pages 67–76, 2008.

[Woo81] W. E. Woodson. *Human factors design handbook*. McGraw-Hill, 1981.

[WS02] M. Woollacott and A. Shumway-Cook. Attention and the control of posture and gait: A review of an emerging area of research. *Gait Posture*, 16(1):1–14, 2002.

[WWG03] D. Whitney, D. A. Westwood, and M. A. Goodale. The influence of visualmotion on fast reaching movements to a stationary object. *Letters to Nature*, 423:869–873, 2003.

[YB85] L. R. Young and S. R. Bussolari. An experimental evaluation of the use of vestibular models in the design of flight simulator motion washout systems. In *Proceedings of AIAA Simulation Technologies Conference*, 1985.

[YJN+96] C. Youngblut, R. Johnson, S. Nash, R. Wienclaw, and C. Will. Review of virtual environment interface technology. Technical Report IDA Paper P-3186, Institute for Defense Analyses (IDA), 1996.

[YOO06] H. Yamanoue, M. Okui, and F. Okano. Geometrical analysis of puppet-theater and cardboard effects in stereoscopic HDTV images. *IEEE Transactions on Circuits and Systems for Video Technology (CSVT)*, 16(6):744–752, 2006.

[ZLB+05] C. Zanbaka, B. Lok, S. Babu, A. Ulinski, and L. Hodges. Comparison of path visualizations and cognitive measures relative to travel techniques in a virtual environment. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 11:694–705, 2005.