Distance Misperception in VR and the Possibility of Individual Calibration

by

Alex Peer

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Computer Sciences)

at the

UNIVERSITY OF WISCONSIN–MADISON

2021

Date of final oral examination: 05/27/2021

The dissertation is approved by the following members of the Final Oral Committee:
Kevin Ponto, Department of Computer Sciences, UW–Madison
Mohit Gupta, Department of Computer Sciences, UW–Madison
John D. Lee, Department of Industrial and Systems Engineering, UW–Madison
Bilge Mutlu, Department of Computer Sciences, UW–Madison
Karen Schloss, Department of Psychology, UW–Madison

To the family, friends, and colleagues who have supported me along the way.

ACKNOWLEDGMENTS

To the members of the Virtual Environments Group, and to those who knew it in its earlier incarnation as the Living Environments Lab, I express my heartfelt gratitude. You gave my research a home, and gave me the support and camaraderie a budding researcher needs to flourish. Patti Brennan, Kevin Ponto, and Karen Schloss – you've built something to be proud of, and something I hope I can be a credit to in the future. Gail Casper, Erica Gill, Ross Treddinnick, Catherine Smith and Marcus Broeckner lent me support and advice both in their official roles in the group, and as occasional test subjects. Many members of the group helped make tea, board games, sharing VR experiences, and simple conversation around the lab into memories worth having. Thanks to Chris Racey, Madeline Parker, Jake Lundy, Simon Smith, Benny Wysong-Grass, Melissa Schoenlein, Karen Chen, Hyo-Jeong Kang, and Kuderat Alimi for making lab events places to have fun and also grow as a researcher, and to Bryce Sprecher for helping to co-organise weekly showcases to explore the possibilities of VR. Particular thanks to my labmates Hyo-jeong Kang and Chris Racey, who never hesitated to lend an ear, and provided invaluable feedback that helped to shape my research. Thank you to students Peter Ullrich and Jay Sovinec, who assisted in data gathering and implementation of early pilots of the tracking alignment and acuity projects.

The Human-Computer Interaction and Visual Computing labs gave me a temporary home at the start of my PhD. Bilge Mutlu and Mike Gleicher built labs with a studious but convivial atmosphere. They helped me to understand what research could be, and their students gave me a glimpse of what growth as a researcher looks like, and that I could become one. In particular, I'd like to thank Alper Sarikaya, Brandon Smith, Dan Szafir, Danielle Albers Szafir, Eric Alexander, Michael Correll, Nathan Mitchell, and Tomislav Pesja for always being sure I felt welcome at the weekly lab

lunch at Jordan's.

Thank you to the members of my committee for their invaluable feedback, and for helping me to find direction and plausible scope within the broad set of my many interests.

Before UW - Madison, I completed a Master's Degree at Michigan State University. Thank you to my Master's Joyce Chai and the members of her lab, who helped me learn that research was something I wanted in my future. Thank you to the Instructors and my colleagues teaching CSE 101, and to the other students who helped make my time at MSU something to cherish.

Thank you to the IEEE VR community for being a welcoming place. Doug Bowman served as my Doctoral Consortium advisor at a conference I attended early in my PhD work, and was sure to check in in subsequent years to provide conversation and advice, and overall make me feel welcome in the community. Conversations with researchers like Peter Willemsen, Scott Kuhl, and Betty Mohler helped me not just to better understand their work, but to put a face to some of the foundational work in my topic of study. Meeting Adam Jones and other researchers actively pursuing perception research further humanized the process, and showed me that this field might be one that I want to be a part of.

There are many friends that lent me support, even as my studies kept us physically distant. Adam Hogan is part of the impetus that set me on the path towards my graduate studies, and was was there to help me navigate along the way. Mo Kakwan was always willing to proof a paper draft or listen to whatever difficulties I'd encountered in my research. Kimi Shirasaki is a source of boundless optimism and dependable coding expertise. Jonathan Greca, Andrew Piskorowski, and Nick Testorelli: we may not see each other as often as we'd like, but in many ways you were here with me for every step.

To my family, Sandra, David, Samantha, Grace, and my extended family as well: you have helped shape so many aspects of who I am, being grateful is almost redundant, and certainly insufficient. The work in this document is made possible because you helped to make me who I am.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Virtual reality is poised to become an ubiquitous, accessible tool with applications in training, design, remote collaboration, and leisure. When viewing environments using the immersive displays that power virtual reality, people sometimes perceive things to be closer or further than intended - they experience a distance misperception. Decades of research have identified several influencing factors, but no definite cause or solution.

Early in my explorations of this effect, I encountered some evidence that individual differences between viewers may have a strong influence on distance misperception. This finding influences the bulk of the rest of the work described here. It was also unclear whether established manipulations of perceived distance could be used to mitigate distance misperception without influencing other aspects of scene rendering or other perceptual tasks. This ambiguity and the possible influence of individual differences motivated a novel method of manipulating perceived distance that attempts to mitigate distance misperceptions in a way that can be calibrated to an individual viewer, while leaving other aspects of scene rendering untouched. This technique is called *perceptual space warp*, and through a series of user studies it's efficacy as a mitigation strategy for distance misperception is explored, as well as its possible detrimental effects on other perceptual tasks.

Another thread of research attempts to observe individual differences that might lead to variance in distance misperception more directly, by establishing methods of measuring visual acuity and stereo accuracy as mediated by immersive displays. An effort is also made to establish whether comparisons can be made between different methods of measuring perceived distance, and to provide a standardized method of aligning virtual and real scenes for perceived distance experiments while correcting for a tracking artifact that may cause inaccurate measurements. Also included

is a discussion of previous literature and the difficulties of making comparisons between experiments in the literature.

This dissertation makes a number of practical, empirical, and methodological contributions. These include artifacts of software and method that make novel techniques accessible to future researchers, and user studies that establish their efficacy. It also provides a mitigation strategy for distance misperceptions that user studies suggest could be used by immersive experiences that demand veridical distance perception now, and that can adapt to new findings that may change our understanding of distance misperception in the future.

1 INTRODUCTION

Virtual Reality (VR), Augmented Reality (AR), and other immersive display technologies have developed to a point where they are more accessible than ever before, and may be poised to become ubiquitous tools for learning, work, and play. They provide powerful, immersive illusions of being present in a virtual space. These illusions are so powerful that they may mask subtle imperfections that viewers are not consciously aware of, imperfections that may lead to discomfort, fatigue, sensory confusion, error-prone interactions, or otherwise degrade the utility of VR. One such imperfection is that of distance misperception, which causes viewers to misjudge distances between themselves and objects in the virtual scene. How this impacts practical application of VR is largely unexplored, though it has been shown that performing a task while underestimating distances in VR can lead to overestimations when performing the same task in a real environment Waller and Richardson (2008); this has clear implications for the use of VR in training for real-world tasks, and possibly for AR-assisted tasks that mix virtual and real spaces.

Over two decades of research have established that distance misperception exists, and that it can be reliably demonstrated and measured using a variety of methods. Several factors influencing distance misperception have been identified, but none has been shown to entirely account for observed misperceptions; no certain cause or means of correcting these misperceptions has been found. The initial focus of the work described in this dissertation was to explore a novel method of manipulating perceived distance. Early explorations provided evidence that individuals may experience distance misperception differently, which in turn implied that methods of measuring and manipulating perceived distance might be modified to account for individuals' difference in misperception. This finding also suggests that there may be aspects of an individual's perceptual experience as mediate by immersive displays that might influence distance misperceptions, and that could be measured more directly. This dissertation argues that the difference between individual viewer's perceptual experience when viewing VR is a meaningful influence on distance misperception, provides an example of a correction that is improved through calibration to these individual differences, suggests changes to existing measures to better account for individual differences, and proposes that additional measurements of individual difference should become commonplace in future research.

1.1 Motivation and Overview

VR can be a powerful thing. You could learn to fly a plane, without the worry of hurting yourself in a crash. You could conduct micro-scale surgery while being embodied in a tiny surgical robot, with the robot's micro world mapped to your natural spatial sense in a way that allows you to guide the robot's movements with surgical precision. But if you can't trust that spatial sense, you might overshoot the runway when you fly your first real plane, or make too deep an incision with your robot scalpel. Distance misperceptions go unnoticed when interacting with VR, but there are certainly opportunities for them to lead to inconvenient or outright dangerous consequences. They are a threat to VR's ability to leverage our natural spatial sense and experience in interacting with a three dimensional world, arguably VR's chief strengths.

It is well established that distances perceived via virtual reality displays are not always perceived as intended. The mean difference in perceived and intended distances has been reported as an average of 26% over 30 articles (Renner et al., 2013), and sometimes 50% or worse (Knapp, 2001; Thompson et al., 2004). In the most frequently studied contexts these misperceptions manifest as underestimation, earning the effect the name *distance compression*; I prefer *distance misperception*, as overestimations are expected under some conditions.

One of the earliest works studying this effect found that it impacted the speed and distance judgements of pilots using a helicopter flight training simulator (Wright, 1995). The subsequent decades of research have explored some portion of the combinatorial space of factors possibly influencing distance misperception, and expanded it by establishing new factors of interest. (See Chapters 2 and 3 for discussion of prior work on the topic.) These efforts have not yet provided a clear cause for distance misperceptions or clear answers to the extent of their influence in practical VR use - these are difficult questions to answer. Past research has shown several manipulable factors that we might use to influence perceived distance to shift things back to their intended place – or at least, to induce minimal error within a specific perceived distance measure. This may not be equivalent to correcting distance misperceptions: it may introduce new perceptual artifacts, and may not transfer as desired to other perceptual contexts. However, if such a technique can induce a correct perception of distance, it might allow relatively direct exploration of the difficult questions of the cause and the practical implications of distance misperception through experiments that create conditions in which distance misperception is not in effect. Chapters 5 and 6 explore this idea by presenting a novel method for manipulating perceived distance that leaves other aspects of rendering untouched, and can be calibrated to an individual's degree of distance misperception.

The several decades of previous work on distance misperception span eras of fast-changing hardware and software, and see the use of a variety of research methodology. Chapter 3 discusses some of the trends seen across this body of work, and the difficulties of making meaningful comparisons between many of the experiments in the literature. Chapter 4 attempts to establish a baseline for distance misperception seen when using contemporary display devices, and explores several established methods of measuring perceived distance to establish whether they could be compared across experiments. In a related vein of facilitating accurate, repeatable methods that facilitate cross-experiment comparison, Chapter 8 explores a method of aligning virtual and real scenes, while correcting for a tracking artifact seen in a consumer-grade tracking system.

Evidence from user studies presented in Chapters 4 and 5 suggest that individuals experience different amounts of distance misperception. This motivates both the per-individual calibration procedure seen in Chapter 6, and also attempts to more directly measure differences between individuals that may influence perceived distance in VR. Chapter 7 describes tools to measure the visual acuity and stereo accuracy of individuals, as mediated by immersive display devices. Though these aspects of participants' perceptual experience are sometimes used to exclude potential participants from an experiment, these measurements or even measurement methods are rarely discussed in detail. Including these measurements alongside other experimental results could reveal correlations with degree of distance misperception as well as the degree to which different influences of perceived distance affect misperceptions; including measures of a participants' visual acuity and stereo accuracy as part of an experiment's results might provide a deeper understanding of distance misperception.

1.2 Contributions

In exploring measures and manipulations of distance misperception in VR, this work makes several contributions. These contributions are discussed in more detail in their respective chapters, and are given a more fine-grained summary divided into practical, empirical, and methodological components in Chapter 10. But broadly:

Individuals' Differences in Perceived Distance as an Influence of Distance Misperception: Early in my explorations of distance misperception, user studies provided evidence suggesting individuals experience different degrees of distance misperception. This finding informs much of this work, and practical and methodological tools are provided such that it might inform the future work of others in the field.

Perceptual Space Warp: A method for manipulating perceived distance is proposed, called here perceptual space warp. This method preserves known-good properties of scene presentation, can be calibrated to an individual's degree of perceived distance, and is theoretically flexible enough to be extended to nonlinear or piecewise misperception spaces. (Chapters 5, 6)

Measures for Visual and Stereo Acuity in Immersive Displays: Methods for measuring visual and stereo acuity as mediated by immersive displays are presented, as examples of measurable differences between individuals that should be expected to be mediated by immersive displays. (Chapter 7)

Scene Alignment and Tracking Correction: Through the course of the experiments described in this document, an artifact causing inaccurate tracking in our consumer VR system was discovered. A method to correct this source of inaccuracy while also aligning virtual and real stimuli was developed, and provided to the larger community in hopes of providing an accessible, accurate, and more standardized method of presenting aligned stimuli and taking measurements for distance perception research. (Chapter 8)

Suggestions for More Comparable Research: There have been decades of distance misperception research, but such a variety of technique and technology is used across this timespan that comparisons of results between experiments are often difficult to make with any confidence. I suggest several possible changes to experimental methodology that, if adopted, might make comparisons between future work more practical, allowing it to better guide further progress in the field. (Chapters 3, 4)

2 BACKGROUND

2.1 Perceiving Distance

> *The sheer number of information sources renders implausible any blindly systematic and thorough experimentation of the perception of layout. [...] Such combinatorics suggests that researchers must set aside global experimentation as being simply unfeasible.*
>
> — Cutting and Vishton, 1995

The process by which we establish the positions of objects in the world relative to ourselves is called egocentric depth perception, and it is one of the key processes by which we build an understanding of the physical world. Perceiving the location of objects in space relative to ourselves (egocentric) or to other objects (exocentric, also called allocentric) is a key process in making sense of our environment. The percept of egocentric distance is often called depth perception; Cutting and Vishton (1995) identify several depth cues and suggest their relative importance at different distances, as seen in Figure 2.1; see that work for a detailed discussion of how cues are constructed, and conditions under which cues are accurate. A brief summary of these cues follows here. See Section 2.3 for a discussion on existing research into how binocular cues may influence perceived distance.

Binocular Cues

Binocular cues are depth information obtained from both eyes working in concert; if a display or image is called 3D (e.g. 3D TV, 3D movie), this is usually meant to indicate that it presents binocular cues. Alongside robust motion tracking, binocular cues are often one of the key factors that differentiate immersive displays used in virtual reality systems from

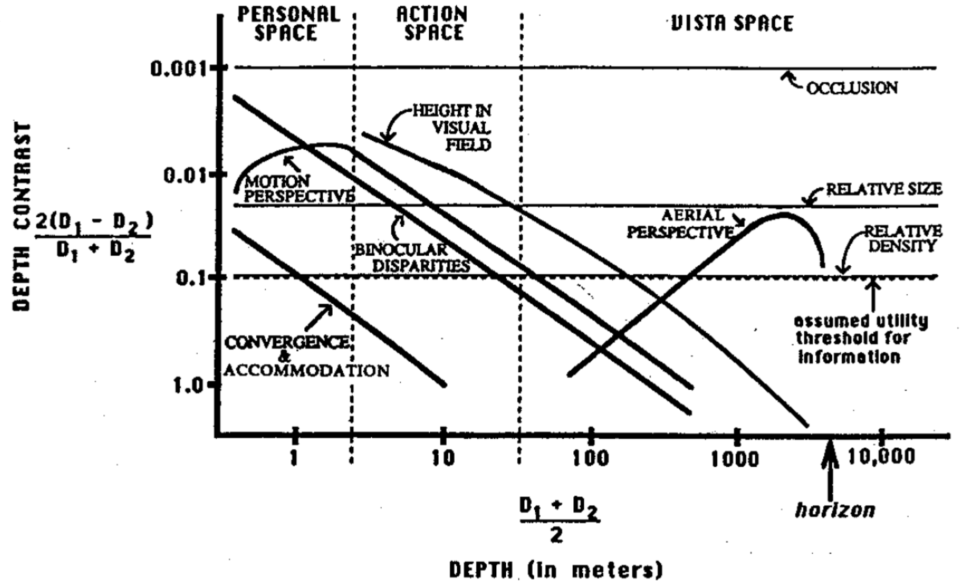

Figure 2.1: The efficacy of various depth cues at different distances. From Cutting and Vishton (1995).

conventional desktop displays. These cues depend on potentially variable human biology, and are facilitated and limited in virtual environments (VEs) by both hardware and software.

Stereopsis is the fusing of both eyes' visual stimulus into a single percept of depth. This depth is indicated by the relative position of objects on the two retinas, called binocular disparity. The absence of stereopsis results in either diplopia (double vision), or suppression (ignoring one eye's view in favor of the other).

Accommodation is the act of adjusting the eyes' lenses to bring an object into focus; **vergence** is the act of rotating the eyes to aim them at an object. These proprioceptive cues are the result of the manipulations of the eye necessary for stereopsis; they're generally regarded as less precise than the visual component of stereopsis (Patterson and Martin, 1992). Under normal stereoscopic vision, degrees of accommodation and vergence correspond to specific distances, and are known to be a strong distance cue out to a distance of at least 2 meters (Cutting and Vishton, 1995). Accommodation may also function as a monoscopic cue, but is closely associated with stereopsis. Capacity for accommodation decreases throughout life, beginning to decline in adolescence and can be expected to be functionally lost between 40 and 50 years of age (Atchison, 1995).

Stereoscopic acuity, or **stereoacuity**, is a measurement of the smallest difference in depth one can perceive through binocular vision, and is measured in degrees of arc subtended on the retina between the two noticeably different depths. It is commonly measured through the use of static images meant to depict specific depths at specific viewing distances. Stereoacuity has been seen to be sensitive to both spatial and temporal frequency of visual stimulus (Schor and Wood, 1983; Schor et al., 1984): in static (low temporal frequency) scenes that offer finer visual details (high spatial frequency), stereopsis can be expected to be reliable to about 80 arcmin; in static scenes with low spatial frequency, stereoacuity may be as poor as 8 arc degrees; stereoacuity is poorer when both spatial and temporal frequencies are high, and is at its best in the presence of either low temporal frequency paired with high spatial frequency, or low spatial frequency paired with moderate temporal frequency (Patterson, 2015).

Monocular Cues

Depth cues that don't depend on the cognitive and oculomotor systems of binocular vision are called monocular cues. Those that don't depend on motion, and so could create an illusion of depth in a still picture, are sometimes called pictorial cues. In the case of virtual reality, monocular cues are primarily limited by tracking hardware, fidelity achieved by the rendering software and display hardware, and the contents of the displayed scene.

Occlusion is one object hiding another by being in front of it, relative to the viewer. Assuming clear boundaries between objects, it provides a strong cue as to the relative depth order of objects in a scene, but does not directly provide absolute depth information.

Atmospheric perspective refers to the affects on light reflected from an object as it travels through the space between the object and an observer. For example, a mountain viewed at a distance would be expected to be of decreased contrast and saturation than closer objects, due to the influence of the air in the intervening space; there may also be some color bias, with air tending to make distant objects more blue (Cutting and Vishton, 1995). Atmospheric perspective can be expected to provide depth order cues.

Relative size is the change in retinal size of multiple objects of similar actual size, but at different distances (e.g. the change in size of near and distant automobiles). **Relative density** is the change in relative retinal position of objects or textures that are arranged in some regular pattern, and are viewed at multiple distances (e.g. the pattern of bricks in a wall that recedes into the distance). This provides both depth order and some degree of absolute depth via relative scale; both are dependent on the viewer's familiarity with, and degree of variation between, the objects' size or density (Cutting and Vishton, 1995).

Height in visual field refers to the vertical position of an object as projected onto the retina. Assuming objects can be grounded to the same plane of reference (e.g. they are all on the floor, or some are on a table that is itself is on the floor), a vertical position closer to the horizon indicates greater distance. As distances grow, the relative differences of height in the visual field grow smaller, rendering the cue progressively less effective (Cutting and Vishton, 1995).

Motion Parallax is the movement of a single stationary object relative to the movement of an observer; **motion perspective** is this relative movement, given a field of observed objects (as in most realistic scenes). Motion parallax in real-world viewing conditions has been shown to inform depth perception in the absence of other cues (Ferris, 1972; Gogel and Tietz, 1979), but to have less effect in the presence of binocular disparity (Watt and Bradshaw, 2003). Motion parallax can be expected to provide absolute depth cues (Cutting and Vishton, 1995).

Partitioning Space

Cutting and Vishton (Cutting and Vishton, 1995) divide space into three partitions that roughly correspond to the distances at which various cues have the most influence.

Personal space is closest, extending from ourselves out to 2m, though this boundary is "somewhat arbitrary" (Cutting and Vishton, 1995); this space is meant to be that which is within arms reach, and so subject to direct physical manipulation. Occlusion, stereopsis, relative size, convergence, and accommodation can be expected to be dominant cues at this range.

Action space is next closest, extending from the end of arm's reach to 30m. This is the range in which we can walk to, talk to, throw to, or otherwise meaningfully take immediate action. Dominant distance cues within this range include occlusion, height in the visual field, stereopsis, motion perspective, and relative size. The 30m boundary is chosen as it is seen (by Cutting and Vishton) as the distance at which stereopsis and motion perspective lose their utility.

Vista space is all space more distant than action space. Objects in vista space are too distant for the motion of eyes or body to provide immediate depth information. Occlusion, height in visual field, relative size, and aerial perspective can be expected to provide the strongest depth cues in this range.

2.2 Comparing Methods of Measuring Perceived Distance

Measuring perceived distance is nontrivial: cognitive state is not directly observable, and the process by which we synthesize depth cues into a perceived depth is not well understood. All available measures are subjective to some degree; the most commonly used class of measure in distance misperception research, directed action measures, achieve some objectivity by asking participants to perform an action that translates to a distance estimate in some objectively measurable way. Most of the measures described here are directed action measures; lesser used measures of verbal report, perceptual matching, and affordance judgements are discussed towards the end of the section. Also note that most directed action measures ask participants to perform their judgement *blind* – eyes closed, wearing a blindfold, or while viewing a blank screen.

Blind Walking

Blind walking is the most commonly used directed action measure; it asks participants to view a target, don a blindfold, and walk to where they remember the target to have been. The location where participants stop walking is taken to be the perceived location of the target. People have been shown to be able to accurately perform this task up to 20m (Loomis and Knapp, 2003).

When performing a blind walking task, one concern is that participants may not walk in a straight line, or that participants may be hesitant to walk without sight. Different strategies have been adopted to avoid this, including simply instructing participants to walk "confidently and without hesitation" (Waller and Richardson, 2008), providing tactile feedback in the form of a physical pathway (Waller and Richardson, 2008), providing audio feedback as participants deviate from the path (Grechkin et al., 2010; Nguyen et al., 2011), having an experimenter walk next to the participant (Creem-Regehr et al., 2005), or through practice (Sahm et al., 2005; Willemsen and Gooch, 2002b).

Blind walking requires a large space – at least as much space in a straight line as the largest distance measured, and preferably significantly more to allow for a variety of starting and target locations so as to avoid learning effects. When using head mounted displays that are connected to additional hardware for power, computation, or tracking, the available walking distance is limited by cable length. This necessitates long bundles of cables, a mobile cart to hold the additional hardware, or some combination of both; carts and cables are sometimes guided by an experimenter (Jones et al., 2008; Interrante et al., 2006, 2008; Willemsen et al., 2009), though often this problem of cables, and whether any solution was needed, is not reported (Sahm et al., 2005; Nguyen et al., 2011; Grechkin et al., 2010; Steinicke et al., 2009b); I am unaware of any research investigating potential biases introduced by cables or cable wranglers. Treadmills have also been used to adapt blind walking to smaller spaces (Bergmann et al., 2011; Witmer and Sadowski, 1998). Many other directed action measures employed in distance misperception research were introduced as space-constrained replacements to blind walking.

Participants have been seen to improve blind walking estimates over time, perhaps due to proprioceptive feedback (Philbeck et al., 2008; Gibson and Gibson, 1955), though Jones et al. (2011) suggest that in HMDs, accidental peripheral visual stimulation due to poor fit may contribute.

Timed Imagined Walking

Timed imagined walking asks participants to judge the amount of time they imagine it would take them to walk to the target, and uses their previously measured average walking time to convert this into a distance. This measure has been found to be comparable to blind walking (Plumert et al., 2005). The time estimate is generally performed with a stopwatch (Grechkin et al., 2010; Plumert et al., 2005; Ziemer et al., 2009).

Triangulated Walking, Pointing

Blind triangulated walking involves walking a short distance in some direction askew of the target, then indicating the perceived angle to the target; it is assumed that this angle points to the perceived location of the target, on the line between the viewer's starting position and the target. People have been shown to be able to accurately perform this task in a real environment at distances up to 15 to 20 meters (Fukusima et al., 1997; Loomis and Knapp, 2003). This perceived angle-to-target has been indicated in a variety of ways: turning and taking just a few more steps (Rébillat et al., 2011; Richardson and Waller, 2007; Thompson et al., 2004; Willemsen et al., 2009), turning and dropping a beanbag (Klein et al., 2009), or, in a variant termed triangulated pointing, pointing a tracked wand (Bruder et al., 2015). The angle at which the initial walk is made askew of the target varies from 40 degrees (Richardson and Waller, 2007) to 90 degrees (Bruder et al., 2015). Distances walked during the initial leg range from a carefully enforced constant (Klein et al., 2009), a random distance participants are unaware of until asked to turn (Thompson et al., 2004), or an informal two steps to the left (Bruder et al., 2015). Triangulated walking estimates may be influenced by participant's knowledge of room geometry (Klein et al., 2009). Triangulated walking or pointing can be adapted to a variety of spaces, requiring only as much space as is walked; however, errors when indicating the perceived angle-to-target can be expected to be amplified when calculating the perceived location of the target as distances walked grow smaller, as angles askew of target become greater, or as distances to targets increase.

Blind Throwing

Blind throwing involves throwing beanbags such that they land at the location of targets. It was first used in virtual environments by Sahm et al. (2005). In my work I've found blind throwing to be a consistent measure with minimal explanation or practice, across broad age ranges, and for distances from 2 to 6 meters (Peer and Ponto, 2016a, 2017, 2019). Sighting and recording a beanbag's initial point of impact are potential sources of error (Sahm et al. (2005) requires agreement between two spotters), and unintended visual, aural, or vibrational feedback is possible as beanbags land, and as experimenters walk through the space to record and retrieve them. Blind throwing is a measure that can be used without increasing the range of tethered HMDs, but requires a clear space as large as the distances of interest, and so is impractical for use in most CAVEs.

Indirect Walking

Indirect walking is similar to blind walking, but differs in that the measure is executed in a different location than that where the stimulus is viewed. More explicitly, one might view the distance to the target, be taken to another room, and then walk the remembered distance. This has been shown to yield estimates comparable to blind walking (Lin et al., 2011), and is argued to emphasize the distance to the target, rather than the positions of viewer and target (Geuss et al., 2012).

Blind Reaching

Blind reaching is a common measure of perceived distances within personal space, involving placing one's hand or a tracked instrument at the perceived location of a target. It has been found that people can perform it with about 8% error when viewing a real target under binocular viewing conditions and when performing the task blind (Bingham et al., 2001);

that same experiment showed significant overestimation of 15% when viewing a virtual target, and another experiment showed significant underestimation in both virtual and real viewing conditions (Napieralski et al., 2011). Using an AR display and a task variant that has participants move a slider along a track, rather than freely position their hand in space, real-world targets at distances within 50cm were found to exhibit 5.5cm of underestimation, on average (Singh et al., 2010).

Verbal Estimate

Verbal report of estimates, often called verbal report or verbal estimate, involves simply asking participants to say how far away they think the target is. This measure has been shown to be influenced by manipulations differently when compared to other directed action tasks in the same experiment (Andre and Rogers, 2006; Kunz et al., 2009); there is some question as to whether it operates using a different cognitive representation than the other measures (Bridgeman et al., 2000; Parks, 2012; Goodale and Milner, 1992; Creem-Regehr and Kunz, 2010; Loomis and Philbeck, 2008). Some researchers have found verbal estimates to be too imprecise for meaningful analysis (Swan et al., 2007; Napieralski et al., 2011), while others have found them to be comparable to distance judgements made via blind walking (Mohler et al., 2006).

Perceptual Matching

Perceptual matching tasks usually involve participants performing some manipulation to match two stimuli. Perceptual matching tasks are able to measure perceived distance directly, but are also able to test other spacial perceptions, such as shape constancy, and are sometimes used alongside distance measures to establish the relationship between different aspects of perception. Sinai et al. (1998) found a perceptual matching task to be equivalent to blind walking when viewing targets in the real world; their task asked participants to view a target, then turn and instruct an experimenter to move another target to the same distance as the first. Repeating this task with two matching virtual world targets presented in an HMD, negligible error is found (average 7%), though it is unclear if this were a test of just noticeable differences rather than distance estimation (Sinai et al., 1999); tests mixing real and virtual stimulus directly might be more revealing, but this is difficult to achieve in an HMD. Jones et al. (2013) overcomes this limitation by first displaying a virtual environment in an HMD, then as a static image on a projection screen; participants were asked to position themselves relative to the projected image such that the static environment matched the size of that remembered from the HMD. Kenyon et al. (2007) use a CAVE to present a reference object (a two-liter Coke bottle) and ask participants to manipulate a randomly rescaled virtual bottle to match; the virtual bottle was presented at different distances, and error in perceived size, rather than distance, was analyzed. AR also makes direct comparisons possible between VE and RE stimuli, though these tasks often differ from the "open-loop" directed action tasks done blind, and instead present "closed-loop" viewing conditions where both targets are in full view during alignment (Singh et al., 2010; Swan et al., 2007). Ponto et al. (2013) use a variety of perceptual matching tasks to establish and evaluate a corrective calibration for spatial misperceptions when using a CAVE. Jones et al. (2016) use a stimulus of two non-overlapping single-pixel vertical lines viewed through pinhole filters, moving the real line to match the virtual line, to provide strong controls on available cues while comparing variations of their anatomically inspired camera placement model.

Other perceptual tasks have involved, instead of direct matching: a two-alternative forced-choice task asking which of two objects is closer or larger, each presented with some change in the visual properties of the

VE (Bruder et al., 2012); adjusting an object to be placed at the midpoint between themselves and a target (Bodenheimer et al., 2007).

Affordance Judgement

Affordance judgements ask participants to judge whether some aspect of the viewed environment affords some action. Participants have been asked if they can pass through a space without rotating their shoulders (Geuss et al., 2010; Walker et al., 2012). Like perceptual matching tasks, affordance judgements are often used to test if other spacial perceptions are affected by some manipulation, rather than measuring perceived distance directly.

Comparisons Between Measures

A few studies have directly compared perceived distance measures within the same experiment, using the same display device and visual stimulus. Sahm et al. (2005) establish that blind walking and blind throwing are comparable in REs and VEs, and Lin et al. (2011) establish that indirect blind walking is comparable to direct blind walking in REs and VEs. Plumert et al. (2005) and Grechkin et al. (2010) show that timed imagined walking performs similarly to blind walking using a virtual environment, but shows underestimation relative to blind walking with a real environment. Lin et al. (2011) find direct blind walking estimates to be more accurate than those of indirect blind walking in VEs and REs. Kunz et al. (2009) manipulate quality of graphics in VEs and find an effect on blind walking estimates, but not verbal estimates. Napieralski et al. (2011) find blind reaching estimates to be more consistent than verbal estimates in REs and VEs. Klein et al. (2009) compare verbal estimates, timed imagined walking, and triangulated walking, and find verbal and timed estimates to be comparable in both RE and VE, while triangulated walking was underestimated relative to the other two measures, particularly in the VE.

Andre and Rogers (2006) test blind walking and verbal estimates in an RE, and find blind walking to be more accurate than verbal estimates, that a change in environment (indoor and outdoor) influenced only verbal reports, and that a vertical displacement via prism lenses influenced only blind walking.

2.3 Factors Influencing Distance Misperception

Many factors of possible influence to distance misperception have been explored, though results are often difficult to synthesize due to methodological differences, and are sometimes outright contradictory. A summary of the work exploring factors influencing distance misperception follows; it is highly likely that all of the factors mentioned here have some effect on distance misperception, though to what degree and how they might interact is largely unknown.

Head-Mounted Display Hardware

There are some clear differences inherent to wearing HMDs compared to unencumbered viewing of the real world. Two have been investigated: weight of worn equipment, and restriction of field of view (FOV). In the real world, restricted FOV caused underestimation only if head movement was also restricted (Creem-Regehr et al., 2005), and a headband reproducing the weight of an HMD had no effect (Willemsen et al., 2009). A mock HMD replicating both restricted FOV and weight showed significant underestimation, though less than the full HMD (Willemsen et al., 2004); a study using a dual AR-VR HMD to view the real world saw either no underestimation, or significant underestimation when two overestimating participants were removed from the analysis (Grechkin et al., 2010); similar studies using AR HMDs to view only a real environment found no significant underestimation, though this seems to have been due to their

HMDs not fully sealing at the periphery when used as blindfold during execution of a blind walking task (Jones et al., 2013, 2011). Some recent work comparing the Oculus DK1 or DK2 HMDs to older NVIS nVisor SX60s showed reduced underestimation (Creem-Regehr et al., 2015; Li et al., 2015), no underestimation (Li et al., 2014), or even slight overestimation (Young et al., 2014); the authors are generally surprised by this result, and tentatively suggest the increased FOV of newer hardware as a potential cause; reduced distance misperception is not seen in my own work with newer HMDs (Peer and Ponto, 2016a, 2017, 2019). Finally, it is worth noting that distance misperception effects have been seen in CAVE-like systems, despite their unrestricted FOV (Bruder et al., 2015; Klein et al., 2009).

Augmented reality HMDs have also seen some study. Swan et al. (2007) found 14% underestimation using an AR HMD that was mounted to a rack, rather than worn, using a blind walking task. Jones et al. (2008) replicates this experiment with a worn HMD and sees negligible underestimation in VR and AR; they attribute this to their use of an HMD capable of both VR and AR, and so able to reuse calibrations performed in AR directly in VR, though later studies suggest peripheral stimulation due to an incomplete seal around the HMD may have contributed (Jones et al., 2011). Swan et al. (2015) used perceptual matching and closed-loop reaching tasks in AR and found overestimation via matching, possibly due to lens properties causing an accommodation-vergence mismatch; they also found that reaching exhibited an initial fixed offset underestimate of about 4cm that was corrected over time, possibly due to an interplay between HMD lenses and closed-loop feedback.

CAVEs and Large-Screen Immersive Display Hardware

Some work has been done investigating distance misperception using large-screen immersive displays, such as CAVEs. Murgia et al. (2009) find distance underestimation in a CAVE, using size estimation as a proxy. Ryu et al. (2005) use a CAVE-like system and joystick navigation as a proxy for walking, and report distance underestimation of 20-40%. Klein et al. (2006) see distance underestimation in a CAVE using timed imagined walking, verbal estimates, and triangulated walking, though the first two perform similarly in both RE and VE (about 25% average underestimation), and the latter yields double the underestimation in VE than RE (about 50%, on average). Bruder et al. (2016) examine the effect of participants' distance from CAVE walls, which effectively changes accommodative distance; they find increased underestimation for virtual objects positioned behind the physical CAVE wall, and a lesser degree of overestimation for objects in front.

Accommodation and Vergence

Most modern virtual and augmented reality display systems achieve stereo vision by displaying 2D images on a surface, presenting different images to either eye; these images can be manipulated to induce any vergence, while viewers' accommodation is bound to the physical display surface. In a CAVE, the accommodative distance is the viewer's distance from the viewed CAVE wall. In HMDs, the displays are generally too close to the eye for comfortable accommodation, and lenses are used to provide a comfortable accommodative distance (e.g. 1.5m to favor near-world interactions, or an "infinite" distance to minimize physical focusing effort). Both kinds of display hardware map accommodative distance cues to distances that may compete with the vergence cues they manipulate to create a sense of depth. This is sometimes called an accommodation-vergence mismatch, or vergence-accommodation conflict (VAC). In current consumer display hardware, VAC is unavoidable; see Kramida (2016) for a discussion of potential hardware-based solutions, such as light-field displays, that may be more practical in the future. Bingham et al. (2001) found that reducing focal distance with -2 diopter lenses decreased distance estimates, suggesting accommodation as an influencing factor. Kenyon et al. (2008) found no evidence to support accommodation as an influencing factor, by attempting to inhibit accommodation using pinhole apertures; other researchers suggest use of pinholes may not remove focus as a cue, but rather offer a focus cue indicating flatness (Hoffman et al., 2008; Frisby et al., 1996; Watt et al., 2005). It is also suspected that disrupting normal accommodation-vergence coupling, as is done in VR displays, could cause fatigue and discomfort (Wann and Mon-Williams, 1997; Ukai, 2007).

Environment: Quality and Cues

The study of differences in environment presented in a VE can be phrased as studying the effects of presenting different monocular depth cues; however, existing work rarely explicitly controls the presentation of specific depth cues, and rather favors the broad and varyingly defined categories of rich and sparse, or high- and low-quality, environments. It is also unclear if distance cues can be meaningfully isolated: the presence or absence of one cue might significantly change the perceptual role of others; and, as more cues are made absent, the resulting environment less directly represents the immersive virtual reality that we hope to study. Construction of environment seems to be an open methodological problem in distance misperception research.

Surdick et al. (1997) perform one of the few studies to explicitly isolate specific depth cues, and may somewhat run afoul of this problem. They isolate cues by removing them one at a time in a virtual environment, but achieve this degree of control by using an unusual display (a Wheatstone stereoscope modified to display images from a greyscale monitor 1-2 meters away) and a greatly impoverished virtual environment (consisting of a two dimensional black square floating in a white room implied by varying amounts of black lines). Cues involving the lines (called here linear perspective, foreshortening, and texture gradient) were found to influence depth discrimination more than those involving the square (relative height, size, and brightness). Extreme control leaves this experiment far removed from cue-rich environments, displays, and tracking systems, though perhaps further experiments in progressively less controlled environments and more conventional displays could bridge that gap.

Other researchers have varied properties of environments more broadly, rather than isolating specific cues. In one case floor and wall textures of varying density were found to have no influence on verbal estimates or blind walking (Witmer and Kline, 1998), and in another a perceptual matching task was found to be more accurate using a brick patterned floor rather than a grass pattern (Sinai et al., 1999). In one experiment, no difference was found between abstract, outdoor, and indoor virtual environments (Armbrüster et al., 2008); in another experiment, differences were found between outdoor and indoor VEs (Bodenheimer et al., 2007). Kenyon et al. (2007, 2008) found a "rich" (target on a textured table, itself on an infinite checkerboard floor plane) rather than "poor" (target in a grey void) environment improved perceptual matching accuracy. Kunz et al. (2009) found that high quality and low quality environments had no influence on blind walking estimates, but did influence verbal estimates; Phillips et al. (2009) found that blind walking is influenced by similar levels of quality if the VE is a replica of the RE the participant is located in. Thompson et al. (2004) found no difference between triangulated walking estimates made when viewing stereoscopic panoramic photos, a matching environment with abstract textures, and a similarly matching wireframe environment.

Interpupillary Distance, Ocular Depth

Interpupillary distance (IPD) is the distance between one’s eyes; the IPD for which software or hardware is designed is sometimes called the

stereo base, or, in the case of software, the geometric IPD (**gIPD**), and is a key measure used to align display hardware and software so as to present correct binocular disparity and minimize lens-induced distortions in HMDs. That said, many studies in the distance misperception literature do not report the IPD for which their display system is designed, or any adjustments made to match participant's IPD. Small deviations from a participant's IPD should be expected to induce measurable distortions in perceived binocular disparity; with the range of human IPD estimated at 45 to 75 millimeters, any hardware or software built to a fixed IPD should be expected to induce distortions in some viewer's distance perception (Drascic and Milgram, 1996). Displays that use mirrors to simulate a larger IPD when viewing real environments have been shown to have an effect on distance perception (CuQlock-Knopp et al., 2001). Kellner et al. (2012b) find no effect on perceived distance between viewing a VE with a stereo base of 65mm and one calibrated to an individual viewer's IPD; Willemsen et al. (2008) replicate this effect for distances form 5 to 10 meters, but find a significant difference for estimates of distances greater than 15 meters – a distance at which binocular disparity should be less informative (Cutting and Vishton, 1995). Bruder et al. (2012) show that larger stereo bases result in closer perceived distances.

Ocular depth is discussed by Jones et al. (2016) as the position of one's eyes relative to some point forward of the eyes (in their case, the front surface of their display apparatus), and is used to establish software camera parameters that they find improve distance perception. Manipulating simulated ocular depth relative to a display screen in software causes changes in the rendered image that are equivalent to those of manipulating a software camera's field of view, also called geometric field of view (**gFOV**), which maps the scene rendered in software to the physical space of the display hardware; manipulating gFOV results in magnifying or minifying the viewed scene. A geometrically correct gFOV exists, given hardware characteristics such as screen size and lens characteristics, and human characteristics such as the nodal points of the viewer's eyes; unfortunately, the actual parameters of display hardware often differ from manufacturer's specifications (Kuhl et al., 2009; Stanney et al., 1998; Steinicke et al., 2011a), and eye motion necessitates either low-latency, high-accuracy eye tracking (possibly impractically so), or choosing a static nodal point that will induce the least error (Jones et al., 2016). Manipulations of gFOV have been seen to influence distance estimations made via blind walking (Kuhl et al., 2006; Kellner et al., 2012a), to not influence verbal report despite influencing blind walking in the same experiment (Zhang et al., 2012b), and to have no effect on affordance judgements (Walker et al., 2012).

Some experiments have studied the influence of manipulating gIPD and gFOV simultaneously. Bruder et al. (2012) vary both gIPD and gFOV as participants perform a two-choice forced-action task in a stationary HMD, finding that both have an influence on perceived distance, with gFOV being a stronger influence. Ponto et al. (2013) manipulate both gIPD and gFOV to find a corrective calibration, and see an influence on a variety of perceptual matching tasks, with some indication that motion is required for an influence on perceived shape constancy.

Adaptation and Transfer

> *Of course, our participants likely recalibrated very quickly and accurately back to the physical world after the experiment was over. None, for example, walked into the door on the way out of the lab!*
>
> — WALLER AND RICHARDSON (2008)

Users of VR seem able adapt to distance misperception, as improved distance estimations have been seen when they receive feedback through any of several methods. This effect is referred to as **adaptation**. Richardson and Waller studied this effect in a series of experiments, and found:

feedback adapting to underestimations of egocentric distances may cause overestimation of allocentric distances (Richardson and Waller, 2005); feedback given for one measure (blind walking) may not have the same effect for other measures (a variant of blind triangulated walking where the full distance is walked) (Richardson and Waller, 2005); adaptations last for at least 1 week (Richardson and Waller, 2005); explicit feedback given via schematic depictions of error are effective, but explicit feedback may tune task execution rather than the underlying percept (Richardson and Waller, 2005); implicit feedback from 18 trials of executing a walking task not blind but fully sighted (closed-loop) significantly corrects for distance misperception seen with both direct and triangulated blind walking measures (Richardson and Waller, 2007); blind walking with an audible tone indicating the correct distance caused a similar correction, while simulated implicit feedback using only optical flow did not (Waller and Richardson, 2008). Mohler et al. (2006) find improved blind walking and verbal estimates using three feedback methods: fully sighted walking, viewing targets after blind walking, and blind walking until verbally told they had reached the target. (Witmer and Sadowski, 1998) found no effect of feedback on distance judgements made by blind walking on a treadmill, given 15 trials of explicit feedback through viewing the target again after performing their measure. Altenhoff et al. (2012) show visual and haptic feedback to improve blind reaching, but not verbal estimates. Ebrahimi et al. (2015) provide feedback via closed-loop blind reaching in an AR HMD, and see improved, but not veridical, distance estimations.

Once adaptation has occurred, it may persist when transitioning from VE to RE; this is referred to as **transfer**. Transfer is desirable for some adaptations, such as learning the skills to use a surgical instrument in a VE; transfer may be undesirable in the case of distance misperception, if it leads to overestimation in the real world. Work has been done establishing the presence of transfer effects in perceived distance. Waller and

Richardson (2008) test blind walking in an RE before and after walking in a VE with full vision, and see consistent, lasting overestimation – within the blind walking task, but not when performing more mundane tasks such as exiting the lab. Witmer and Sadowski (1998) found participants to underestimate distance in the RE after receiving feedback in the VE using a treadmill; Proffitt et al. (2003) saw similar underestimation using optical flow as feedback on treadmills. Interrante et al. (2006) show overestimation in an RE after the implicit VE feedback of passive haptics from interacting with a physical table with a matching representation in the VE. Some studies presenting no deliberate feedback see underestimation in the RE after exposure to a VE (Plumert et al., 2005; Ziemer et al., 2009), or overestimation in an RE, after viewing a VE that is an enlarged or shrunken replica of the RE (Interrante et al., 2008).

2.4 Calibration

Kellner et al. (2012b) create a geometric calibration of the software view frustum by registering multiple points via an AR reaching task, then test the effects of the calibration via blind walking; the calibration improves distance estimation error from an average of 24% to 20%. Ponto et al. (2013) establish a calibration by adjusting software camera IPD (stereo base) and FOV (gFOV) to minimize perceived error in two closed-loop perceptual matching tasks tailored to a CAVE, then evaluate the calibration using two mixed reality tasks aligning position and angle of declination, respectively, and one shape constancy task exclusively in the VE; all measures show some improvement under the established calibration, except for shape constancy when participants' movements were restricted.

Cues Through Motion

Intending to isolate motion parallax as a potential depth cue, Creem-Regehr et al. (2005) found that restricting head movement and FOV when viewing a real-world environment leads to underestimation. They note, however, that restricting head motion may restrict cues other than parallax, such as near-to-far ground scanning (Wu et al., 2004).

Stereo Acuity

Armbrüster et al. (2008) measured stereoacuity with subtest 4 of the TITMUS Vision Tester, and found that greater stereoacuity corresponded with more underestimations.

Presence

Several authors have suggested that participants' sense of presence may influence their distance estimations (Interrante et al., 2006; Mohler et al., 2008; Phillips et al., 2010; Ries et al., 2009; Steinicke et al., 2009b). No direct support for this theory exists as of yet, though Renner et al. (2013) suggest two avenues of research for preliminary support: avatars and transitional environments. Adding avatars to virtual environments has been shown to enhance presence (Bruder et al., 2009; Slater and Usoh, 1994), and has been shown to improve distance estimates (Mohler et al., 2008; Phillips et al., 2010; Ries et al., 2009; Leyrer et al., 2011); however, none of these studies reports measures for both presence and distance perception. Steinicke et al. (2009b) find transitional environments to enhance both presence and distance estimation in the same publication, but don't measure both simultaneously in the same experiment. Phillips et al. (2012) use both a questionnaire and physiological measures to measure presence alongside perceived distance, but found no correlation between the two.

I suspect presence is more effect than cause of lessened distance misperception – if some part of presence is the feeling of being able to interact with the world naturally, then aligning virtual spacial sense with its natural real-world counterpart would be expected to enhance presence.

2.5 Cross-Device Studies

Only a handful of studies have presented the same virtual environment across different classes of devices. Naceri and Chellali (2012) found distances between two virtual objects to be more accurately compared when displayed on a stereoscopic widescreen, rather than an HMD; Klein et al. (2009) show significantly greater distance misperception with a tile display than with a CAVE; Grechkin et al. (2010) shows similar distance misperception across large-screen stereoscopic projection displays, a VR HMD, and an AR HMD. Riecke et al. (2009) showed participants monoscopic photos while carefully controlling FOV, eye height, head movement, and worn equipment weight to provide uniform stimulus across an LCD monitor, a large-screen display, and an HMD, and found no distance misperception across devices. Combe et al. (2008) find a difference in size estimation between an HMD and cylindrical large screen display, but report no tests of statistical significance. Jones et al. (2011) compare blind walking using the same HMD to present VR and AR, and see significantly more distance misperception in VR than AR; they also test the effect of the presence of peripheral stimulus, which is seen to improve estimates in VR but not to affect those in AR.

3 CONSIDERING 25 YEARS OF RESEARCH

Decades of research have seen a plurality of technique and technology brought to bear in the study of distance misperception. This variety often foils attempts to compare results across experiments. It is also sometimes suggested that as technology advances, distance misperception may eventually go away. This chapter makes an attempt to observe the behavior of distance misperception over time, as seen in the results of different experiments in the literature. Factors limiting constructive comparisons between experiments are discussed. Several aspects of head mounted displays that have changed over time are not found to have a strong effect on distance misperception. Recommendations are offered, in hopes that future research adopt methods that make for easier comparisons between experiments. [1]

3.1 Overview

It is sometimes suggested that we have seen an improvement in distance misperception over time (Feldstein et al., 2020). This isn't an unreasonable assumption: aspects of VR such as graphical quality, tracking, and interactivity should be expected to lead to perceptual artifacts (Renner et al., 2013), and it might be assumed that the technology providing these have improved with time, thereby improving distance misperceptions. To investigate whether this assumption holds given available data, Figure 3.1 collects results from publications from the last 25 years. While a regression line shows a slight trend for lower error over time, the relatively low R^2 suggests that this relationship does a poor job of capturing the variance in

---

[1] This chapter is adapted from Peer and Ponto (2021) [Manuscript submitted for publication].

the dataset. The range and variance of error appears relatively constant over time.

We note that publication year is a gross approximation for technological advancement, and that these papers utilize different measures, devices and populations for their studies; this method of comparison is a relatively blunt instrument. However, even across these methodological differences, if a shift in technology had had a large effect it should be apparent.

In this paper we aim to explore hypotheses for the misperceptions seen in virtual reality through an analysis of published results over the past 25 years. All told, the analyzed work consists of 40 different papers, with over 700 participants sampled, which have explored issues of perceived egocentric distance in virtual environments over the years. Through this data, we examine various potential factors such a accommodation, weight, field of view, and resolution and clarity. Finally, we discuss the limitations of our approach and offer guidance for future research.

3.2 Related Work

Feldstein et al. (2020) provide several figures and tables providing perceived distance estimates collected from across the literature. In particular, their Figure 2 is, like our Figure 3.1, an attempt to capture any change in distance misperception over the years. Their tentative conclusion is that distance misperception has strongly improved over time and may now be gone. This is a compelling conclusion that feels intuitive: the past twenty-five years have seen rapid advancement in many technologies, so why wouldn't VR devices have followed suit, and why wouldn't improved technology reduce perceptual artifacts? It may be so compelling a conclusion that there is some danger of allowing out intuition to supplant scientific rigor. This - again, tentative! - conclusion, and the construction of this figure, wasn't the focus of Feldstein, and detailed analysis and

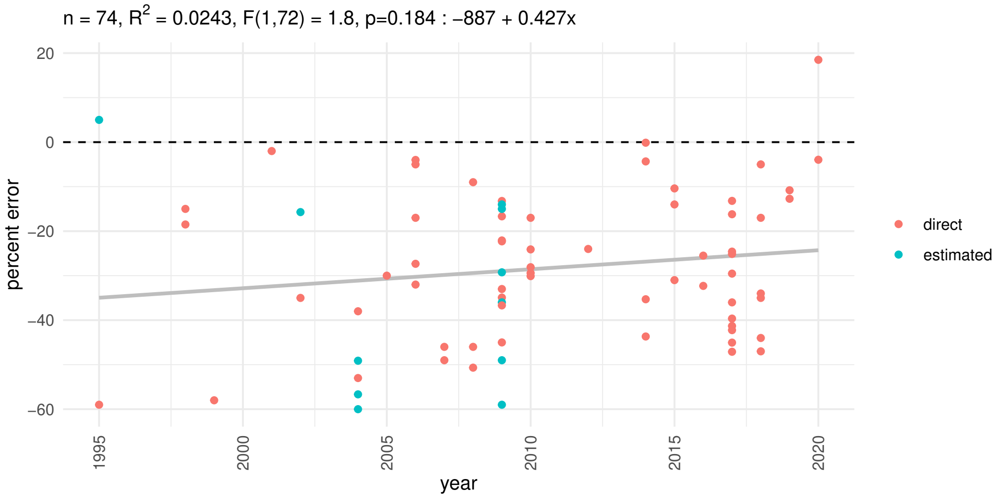

Figure 3.1: The percent error in perceived distance, as measured in 74 different conditions of 40 different experiments performed between the years of 1995 and 2020. A linear fit shows some trend towards less error, but a similar range of results largely holds across all 25 years of research. The dotted line represents 0 error; estimates above show overestimation, and below, underestimation.

discussion of how this figure was constructed was outside of the scope of that work; this chapter seeks to investigate this effect with an expanded dataset, and to describe the criteria for inclusion and exclusion of data points in detail enough that any potential bias introduced can be detected and corrected by any interested future researchers. This chapter may not be an exhaustive exploration of the available literature, nor a definitive example of how to establish rigor in similar explorations, but I hope it provides a foundation from which to build towards further understanding of the effects of distance misperception.

In their survey, Renner et al. (2013) provide several tables collecting the results of many experiments, but focus on broad results such as whether a manipulation leads to a significant difference, or to over- or under-estimation. Here we focus on collecting numeric data in a format that allows comparisons across experiments.

Jones et al. (2008) collect perceived distance error, measurement method, and target distances tested from several published works testing perceived distance in augmented reality, virtual reality, and real-world viewing conditions. They collect these details in order to enhance the discussion of the results of the main focus of their paper; this is similar to the goal of Feldstein et al. (2020). There may be some merit in a similar effort to simply gather contemporary results in one place, in an easily comparable format.

3.3 Methods

The dataset visualized in Figure 3.1 was constructed first by collecting papers relevant to the topic, then searching for those who report some metric that can be compared across experiments. Before advancing to analysis, we further filtered the dataset according to criteria described in Section 3.3.

Data Gathering

An initial set of papers were collected from those referenced in the survey by Renner et al. (2013), and by following citation trees forward from there - papers citing Renner, papers citing those, and so forth. Further papers were found using Google Scholar searches using keywords such as "perceived distance", "distance compression", "distance misperception" combined with "in VR", "in AR", "in XR", "in immersive displays", and "in virtual environments".

Papers were then restricted to those that reported their measures in a way that lets us compare them. The most frequently used and most directly comparable metric reported is error in perceived distance as a ratio of intended distance, which we'll call *percent error*. Figure 3.1 contains data only from papers that provided some numeric metric from which percent error could be derived directly (Buck et al., 2018; Li et al., 2016, 2015; Creem-Regehr et al., 2015; Kuhl et al., 2006; Siegel and Kelly, 2017; Kelly et al., 2014, 2017; Alexandrova et al., 2010; Jones et al., 2008; Wright, 1995; Bingham et al., 2001; Sahm et al., 2005; Mohler et al., 2006; Richardson and Waller, 2007; Steinicke et al., 2009b; Feldstein et al., 2020; Solini et al., 2019; Li et al., 2014; Zhang et al., 2012a; Yang et al., 2020; Peer and Ponto, 2016b, 2019, 2017; Interrante et al., 2006; Durgin et al., 2002; Knapp, 2001; Willemsen et al., 2004; Witmer and Kline, 1998; Witmer and Sadowski, 1998; Young et al., 2014; Ziemer et al., 2009; Grechkin et al., 2010; Kunz et al., 2009), or provided a graph from which percent error could be reasonably estimated (Rolland et al., 1995; Willemsen and Gooch, 2002a; Thompson et al., 2004; Phillips et al., 2009; Willemsen et al., 2009; Klein et al., 2009). The values that were estimated are marked in blue in Figure 3.1.

In general, the papers we surveyed are not interested only in measuring perceived distance, but rather the effects of some manipulation on perceived distance, or the correlation of perceived distance in XR with some other measure. In the case of manipulations, we recorded only the results of the pre-manipulation conditions.

Very few papers report per-user results, but several papers provided more than one viable measure, either through multiple experiments or multiple conditions within the same experiment; in this case, each of the multiple measurements is a separate datapoint – that is, each datapoint represents the mean percent error for a single condition. Infrequently, measurements were reported separately for different distances or individual participants; in this case these measurements were collapsed to a group mean, to facilitate comparison to the majority of the dataset which already aggregated on these factors.

Given these restrictions, an initial dataset of 74 mean percent errors

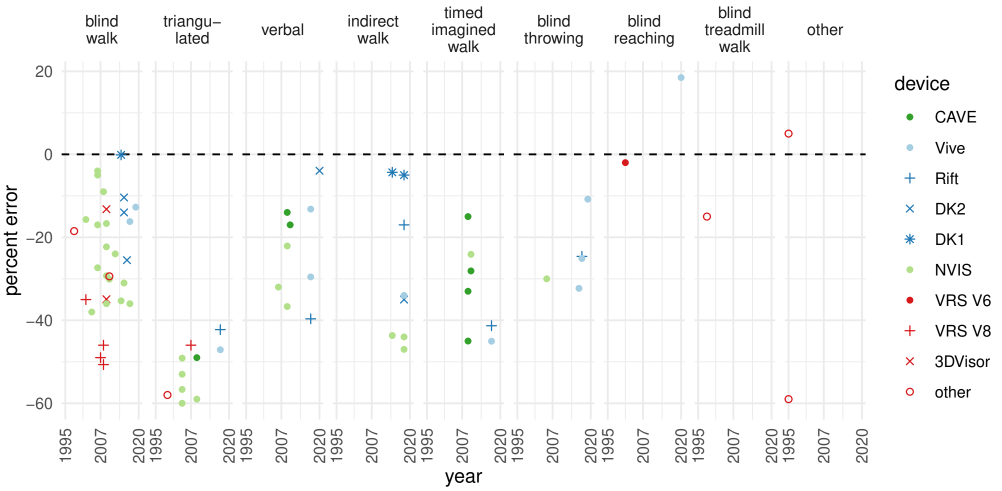

Figure 3.2: Percent error found in the full set of experiments, divided by measure. The dotted line represents 0 error.

from 40 papers was constructed. The results are presented in Figure 3.1.

Data Exclusion

There are several aspects of distance perception research that make it difficult to compare results between studies, which may merit exclusion of results from our dataset. Those that are reported widely enough to be used as exclusion metrics include: method used to measure perceived distance, the visual stimulus used to elicit distance judgements, and the device used to display the visual stimulus.

Perceived Distance Measure

Perceived distance is a cognitive state that cannot be directly observed, and so must be measured indirectly. Various measurement strategies have been employed by the studies that comprise our sample, as summarized

Table 3.1: Summary of percent error by measure used.

	measure	count	mean	sd
1	blind walk	30	-0.24	0.13
2	triangulated	10	-0.52	0.06
3	verbal	9	-0.23	0.12
4	indirect walk	8	-0.29	0.18
5	imagined	7	-0.33	0.11
6	throwing	5	-0.25	0.08
7	reaching	2	0.08	0.14
8	other	2	-0.27	0.45
9	treadmill	1	-0.15	

in Table 3.1. How these methods are performed is discussed in more detail in see Chapters 2 and 4, and for further discussion see Renner et al. (2013). Some prior work has evaluated whether these measures yield comparable distance estimates: blind throwing has been shown to be comparable to blind walking (Sahm et al., 2005); indirect blind walking has been shown to be comparable to blind walking when vertical FOV is at least 21.1°(Lin et al., 2011); timed imagined walking has been shown to perform similarly to blind walking when using a virtual environment, but to elicit underestimation relative to blind walking with a real environment (Plumert et al., 2005; Grechkin et al., 2010); verbal report and timed imagined walking were shown to be comparable (Klein et al., 2009); triangulated walking showed underestimation relative to both verbal report and timed imagined walking with a virtual stimulus, but not with a real stimulus (Klein et al., 2009); verbal report was seen to elicit greater underestimation relative to blind walking (Andre and Rogers, 2006) and blind reaching (Napieralski et al., 2011); a change of graphical fidelity in presented virtual environments was seen to influence blind walking estimates, but not verbal report (Kunz et al., 2009); a change from indoor to outdoor environment in the real world was seen to influence verbal estimates, but not blind walking (Andre and Rogers, 2006); blind throwing, timed imagined walking, triangulated pointing, and verbal report are shown to be more comparable when the difference between real and virtual estimates is taken into account and results are aggregated over distance, but still significantly different at individual intervals of a 2-4m range (Peer and Ponto, 2017). Overall, it is unclear from prior work if these different perceived distance measures should be expected to produce the same estimate for a given stimulus, which would be required for the sort of comparisons we would like to make.

Figure 3.2 shows the same dataset as Figure 3.1, split by measure. Blind treadmill walking, blind reaching, perceptual matching, and the one perceived distance measure that defies these categories have particularly few representatives in this sample (< 2), and so were removed from further analysis. Employing Tukey's HSD to compare the remaining measurement methods showed triangulation measures to be significantly different from timed imagined walking ($p = .036$) as well as all other measures ($p < .001$). As such, conditions employing triangulation measures were excluded from further analysis (Richardson and Waller, 2007; Thompson et al., 2004; Willemsen et al., 2004; Klein et al., 2009; Knapp, 2001; Peer and Ponto, 2017; Willemsen et al., 2009). We remove these datapoints not because they represent invalid data, but because they appear to represent a different enough class of response that they may merit exploration on their own, an exploration that is outside of the scope of the current work. However, a brief discussion of why they may be different seems merited.

Blind reaching (Bingham et al., 2001; Yang et al., 2020) elicited some of the the only instances of overestimation. Reaching necessitates the use of target distances closer than most other measurement tasks are used to explore, which may be expected to cause overestimation due the influence of an accommodation-vergence conflict. See Section 3.3 for more discussion related to this effect. See further in this section for a description of Rolland et al. (1995), the other instance of overestimation.

Triangulation measures have been used more broadly (Thompson et al., 2004; Richardson and Waller, 2007; Peer and Ponto, 2017; Knapp, 2001; Willemsen et al., 2004, 2009; Klein et al., 2009). Klein et al. (2009) did find triangulated walking to show greater underestimation relative to both verbal report and timed imagined walking given a virtual stimulus, and no significant difference in estimates made using the three measures given a real stimulus. They suggest this may be due to the smaller physical space used to administer the virtual stimulus providing additional cues or difficulty when executing the measure; as triangulation measures are often chosen to facilitate target distances outside of the available physical space, this problem may be common. The nature of triangulation may also serve to exaggerate error, as the same degree of angular error has greater impact on distance judgements as target distances grow - though it is not clear that this source of error would manifest as a bias towards underestimation. It is also worth noting that the difference between triangulated pointing and the three other distance measures used in Peer and Ponto (2017) shrank considerably when the error in virtual estimates was defined as relative to real estimates, rather than to the target distance. There may be some merit to that technique, but the majority of experiments in this dataset do not report a real condition to compare to.

We've also removed the two oldest studies on distance perception in the set, Rolland et al. (1995) and Wright (1995). These early contemporaries in the field are interesting in that they roughly span the range of misperception seen through to today ($+5\%$ and -59% respectively). They also both use unique perceived distance measurement methods that are hard to compare against those that have seen more use as the field has developed.

Rolland et al. (1995) uses a unique measure of perceived distance. This study asked 3 participants to perform a two-alternative forced choice perceptual matching task, targeted distances of 0.8 and 1.2 meters, and used a custom AR haploscope to display the stimulus. Their measurement task asks participants to choose whether two objects are at the same or different distances from the viewer, either or both of which can be real or virtual; multiple trials with one object fixed and another at a distance selected using the method of constant stimulus allowed the researchers to determine the participant's point of subjective equality, here the distance from the viewer at which the distance between the objects could not be determined any better than chance. This is a very different method of measuring perceived distance than is used in the other experiments in the dataset. In particular, participants may have based their distance estimates on the distance between the two objects, an allocentric distance, rather than the egocentric distance between themselves and the objects. Allocentric distance judgements have been seen to show less error than egocentric in VR (Waller, 1999; Richardson and Waller, 2005). However, Rolland et al. do show a difference between real and virtual percieved distances, though small (5%) relative to the rest of our dataset; they suggest several possible causes, including image distortion due to lenses, change in interpupillary distance (IPD) due to eye movement, the luminance of the virtual stimulus, and difference in the accommodation of the lens of the eye between real and collimated virtual images. Rolland et al. (1995) is a poor fit for comparing to the rest of our dataset due to it's different measurement technique, the possibility it isn't measuring egocentric distance judgments, and because of its relatively close range of tested distances. However, it should be noted that the issues it was exploring have either been shown to be an influence and have well documented mitigation strategies (lens distortion), or are still being explored (IPD (Jones et al., 2016), luminance and collimation (Singh et al., 2018)).

Wright (1995) asked 6 participants to make distance judgements using the head-mounted display of an ARI STRATA flight training research simulator. Two displays were provided for each eye: a higher resolution

25x18 field-of-view center display inset into a lower-resolution 125x65 field-of-view outer display. The resolutions of these displays are described as 1.5 and 5 "arc min TV line resolution", which may mean 0.2 pixels per arcmin for the low resolution portion of the display, and 0.67 for the high resolution inset: see Section 3.4 for further discussion of how these might compare to spatial resolutions seen across the dataset, but both the shift in how this aspect of technology is reported, and being the sole example of a dual-resolution display, makes it hard to compare this work to others with confidence. Relatively detailed measurements of the various virtual objects and their locations are included, but unfortunately no images of the virtual stimuli are provided, rendering comparisons of visual quality impossible. The perceived distance measurement method used was also unique: a reference distance or height was shown using target objects, then a joystick was used to pilot the virtual helicopter along a fixed path to match the distance. Speed estimates are also performed, in which participants were provided with a target value and asked to modify the virtual helicopter's speed to match. A median (notice, not mean) −59% error is seen for forward distances (the task most like the egocentric estimates of the rest of our dataset) and speed estimates, −50% for lateral distances, and −28% for height estimates. To the best of my knowledge, height, speed, and lateral estimates have not seen further exploration in the literature. The relatively large amount of error for forward distances is somewhat surprising given two results in subsequent literature: the target stimulus may have been an allocentric distance between two objects, which has been shown to elicit less error than egocentric distance estimates (Waller, 1999; Richardson and Waller, 2005); it appears that participants had full visual feedback while moving the virtual helicopter, which might be expected to elicit an adaptation to the virtual environment, and thus reduced error in distance judgements (Richardson and Waller, 2005, 2007; Waller and Richardson, 2008; Mohler et al., 2006; Altenhoff et al., 2012; Witmer and

Sadowski, 1998; Ebrahimi et al., 2015). Most adaptation experiments provide feedback during or in response to walking; possibly, joystick-driven simulated movement does not elicit strong adaptation. Overall, Wright (1995) provides compelling evidence that distance misperceptions may influence practical applications of VR such as flight training, but differences in the methods used, the unique display device, and the lack of examples of the visual stimuli make it difficult to make comparisons between this work and the rest of our dataset with any confidence.

Visual Stimulus

Immersive visual displays provide an image or series of images – a visual stimulus. The goal of most immersive displays is to manipulate distance cues so as to create the illusion of space and place – the viewer's perceived distance. If changing the distance cues provided by the visual stimulus can change perceived distance, it might be expected that changes to the quality of the visuals would influence their ability to convey distance. Several prior works have investigated this by dividing visual stimuli into categories roughly corresponding to high and low quality, or possibly cue rich and cue poor: no difference was found between distance estimates made when viewing the same virtual environment depicted by photorealistic stereo panoramas, abstract textures, or a wireframe (Thompson et al., 2004); in a conflicting result, a difference was found between photorealistic and wireframe-like environments (Phillips et al., 2009);high and low quality environments were found to elicit different estimates through verbal report, but not blind walking (Kunz et al., 2009); perceptual matching accuracy was found to be greater in a "rich" environment with a target on a textured table than in a "poor" environment, with a target in a grey void (Kenyon et al., 2007, 2008).

The apparent conflict between Phillips et al. (2009), who saw a difference between estimates made when viewing a wireframe and photorealistic environments, and Thompson et al. (2004), who did not, is particularly interesting. Phillips et al. suggest that the wireframe environments used in the two experiments may have differed in geometric complexity, and so provided distance cues of some different quantity or quality - that there is some aspect of scene construction not captured in the broad binary of high and low quality, and that the in many ways similar cue poor environments used in the two experiments were different in some key way, as yet unmeasured. In Kenyon et al. (2007, 2008), it is no surprise that an infinite checkerboard plane might offer more distance cues than blank void, but how different is this checkboard from the wireframes used by Thompson or Phillips? When Bodenheimer et al. (2007) find a difference between distance estimates made when viewing a 3D rendered indoor environment and a photorealistic outdoor environment, how should we compare this outdoor environment to environments used in other experiments? Is it most similar to the void of Kenyon, due to having no ceiling? Is it most similar to the photorealistic panoramas of Thompson, despite their depicting a relatively small space? Is a large room more like a hallway, or a field? It may be that through quantifying the available depth cues, a measure of scene similarity could be established. The development of such a metric is outside of the scope of the current work; in its absence, we hesitate to classify the environments in our dataset, or exclude datapoints based on them.

Interrante et al. (2006) found that a virtual environment matched to the real environment in which the experiment takes place elicited low error in distance estimates (-5%), regardless of which environment was seen first. They further investigate this result in several following experiments, and begin to suspect the effect is due to virtual environments evoking more feelings of presence (Interrante et al., 2008; Steinicke et al., 2009b, 2010; Phillips et al., 2010, 2009). It may be that rather than geometric complexity or available depth cues, presence is most predictive of the visual stimuli's effect on perceived distance, and a simple presence questionnaire (Witmer and Singer, 1998; Usoh et al., 2000) is the comparison metric we need; however, few of the studies in this dataset report a measure of presence. It is also difficult to establish the degree to which environments are "matched" in our dataset; contrary to the results of Interrante et al. (2006) which suggest we should see reduced error, several experiments using seemingly matched environments show significant error: -30% (Sahm et al., 2005), -36% (Willemsen et al., 2009), and -38% (Willemsen et al., 2004). It is possible that, as with the conflicting results of Phillips et al. (2009) and Thompson et al. (2004), these environments vary in some way that is not immediately apparent when viewing the few still images of the environments that saw publication. In might benefit the field to establish methods to better record and distribute the real and virtual environments that experiments are situated within.

Our hypothetical scene similarity metric might help quantify how matched an environment is to the real world, or to the environments of other experiments. Developing it could make for a fascinating course of research and source of discussion within the field. It could also take no small amount of time. In it's absence, we suggest that at least virtual visual stimuli be published alongside their textual counterparts; there's some question of whether the 25-year-old programs used in Rolland et al. (1995) or Wright (1995) would run faithfully, or at all, on today's hardware, but if they did they could answer many questions that text and images cannot. Establishing the infrastructure to preserve and distribute these stimuli alongside publications might, again, take some time. Another approach might instead be to establish some common set of stimuli; this would obviate the need for comparison metrics, or at least provide a baseline when changes are made; it would also simplify any preservation efforts, as the effort to run one 25 year old stimuli is likely more practical than one for each of our collection of forty papers.

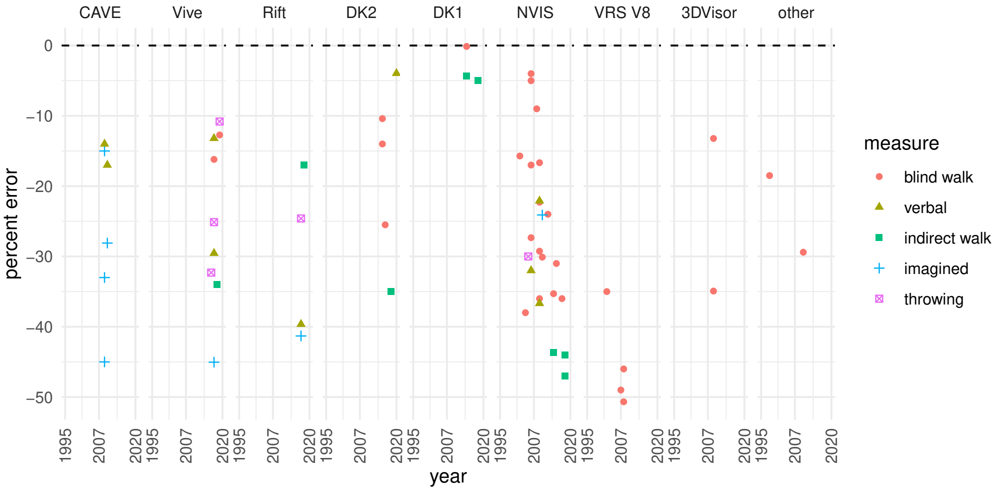

Figure 3.3: The data from Figure 3.1, filtered to the measures selected in Section 3.3, split by device used.

As regards to the current work, visual stimulus has no impact on an experiment's inclusion in our dataset.

Device

The studies in our dataset used many different devices to display their visual stimulus, as can be seen in Figure 3.3. These include CAVEs of varying configurations, a variety of head-mounted displays (HMDs), and two other systems, from Rolland et al. (1995) and Wright (1995), that are unique in the dataset (and already to be excluded, due to their unique measurement methods). Comparisons across these devices might capture differences in factors that are otherwise difficult to produce, such as varying resolution or weight, but there are some challenges to cross-device comparisons that led us to further filter our dataset before analysis.

CAVE CAVEs present an interesting environment in which to execute distance judgements, in that many aspects of the display are dependent on the distance of the viewer from the projection surfaces - angular resolution, field-of-view, and the distance between viewer and projection surface should be expected to vary moment-to-moment during normal viewing, and certainly between experiments. This could have an unpredictable effect on perceived distance. Bruder et al. (2016) positioned participants at varying distances to a CAVE wall, and a target virtual stimulus at varying distances either closer or further away than the CAVE wall, and asked participants to estimate the distance between themselves and the target. They find that position of the target relative to the wall has a strong effect: when the target is closer than the wall participants overestimate; when further, they underestimate. Bruder et al. conclude this bias of distance judgements towards the CAVE wall is likely due a conflict between accommodation (the flexing of the eye's lens to focus at a specific distance) and vergence (how far one's eyes rotate to fuse stereo images); they suspect this because the distance estimates seemed biased towards the location of the projection surface of the wall, which is what, in a CAVE, determines the accommodative distance of the eye. They also suggest changes in angular resolution as a possible influence.

Given the dynamic nature of many properties of the viewing experience in a CAVE, we omit datapoints that used these devices from our further analysis (Grechkin et al., 2010; Ziemer et al., 2009; Alexandrova et al., 2010; Klein et al., 2009). However, it should be noted that CAVEs exhibit a similar range of underestimation as the rest of the dataset, as seen in Figure 3.3.

Collimation If the distance the eye accommodates to influences perceived distance in a CAVE, should we also expect it's influence in our comparisons across HMDs? To determine this, we would need to know the HMD's equivalent of the distance between the viewer and the CAVE wall - the distance at which an HMD's image is presented. Let's call this distance, the distance to which the viewing eye accommodates, *accommodative distance*. Unfortunately, accommodative distance is rarely described in most papers, and often omitted from manufacturer device specifications. In contemporary HMDs, the focal length of their lenses would be expected to be the most direct influence. It is largely believed that the HTC Vive and Oculus Rift CV1 and DK2 have accommodative distances of between 0.5m and 2m; The Oculus DK1 is believed to present a collimated image, with a focus roughly at infinity. The Oculus DK1 and DK2 had interchangeable lenses, and so may have been customized to provided different accommodative distances. The authors were able to contact NVIS and confirm that their HMDs present a collimated image in their default configuration, though clients could request a customized accommodative distance.

Singh et al. (2018) provide perhaps the most relevant result regarding collimation. They investigate the influence of accommodative distance on distance estimation in AR using a perceptual matching task to estimate distances between 0.34 and 0.5 meters. They use a custom haploscope that can be configured to present variable accommodative distance; when presented with a collimated image, participants showed consistent overestimation, increasing with greater distance.

Figures 3.5 and 3.4 compare results from 27 conditions that used collimating displays (Buck et al., 2018; Creem-Regehr et al., 2015; Kuhl et al., 2006; Siegel and Kelly, 2017; Kelly et al., 2014; Jones et al., 2008; Willemsen and Gooch, 2002a; Sahm et al., 2005; Mohler et al., 2006; Phillips et al., 2009; Li et al., 2014; Zhang et al., 2012a; Interrante et al., 2006; Willemsen et al., 2004; Young et al., 2014; Grechkin et al., 2010; Willemsen et al., 2009; Kunz et al., 2009), and 18 that used non-collimating (Buck et al., 2018; Li et al., 2016, 2015; Creem-Regehr et al., 2015; Kelly et al., 2017; Feldstein et al., 2020; Solini et al., 2019; Peer and Ponto, 2016b, 2019, 2017). Both

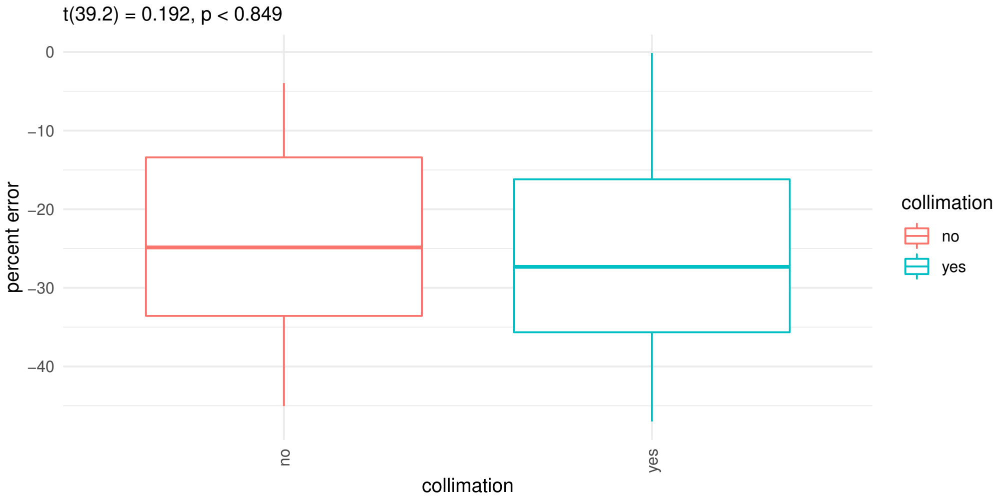

Figure 3.4: 19 studies that used collimating displays show similar degrees of distance misperception than 12 that did not.

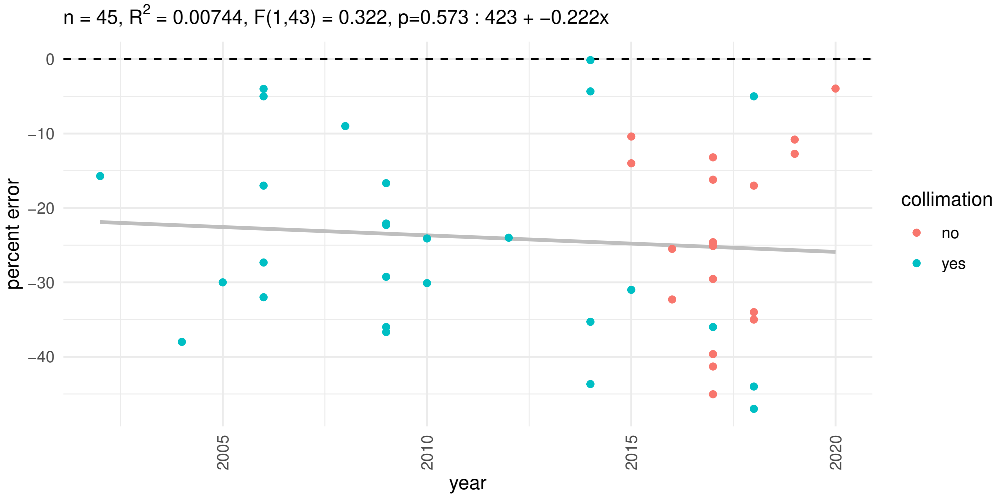

Figure 3.5: Collimation data formatted to better compare to Figure 3.1.

collimating (M=-0.25, SD=0.14) and non-collimating (M=-0.24, SD=0.12) show similar amounts of underestimation, with a t-test showing no significant difference, t(39.2) = 0.192, p = 0.849. This is surprising, as the findings of Singh et al. (2018) and less directly Bruder et al. (2016) imply that collimating displays should show overestimation; this surprise may suggest that accommodation-vergence mismatch is not alone the chief contributor to distance misperception in the majority of these studies. Note that we assume that Oculus DK1, DK2, and NVIS displays are in their default configurations, an assumption made somewhat blindly. Consistent reporting of the accommodative distance of immersive displays by manufacturers and researchers, or standardized methods for researchers to measure and report accommodative distance of displays used in their experiments, would be needed to better answer questions of the influence of accommodative distance on distance perception, and ideally with more granularity than the binary of collimating or not.

The two groups are similar enough that both remain a part of the dataset for our further analysis.

Further Missing Data The filtering criteria described thus far in this section yields a dataset of 57 mean perceived error measures. Our further analysis explores the influence of display device weight, field of view, and resolution on distance misperceptions; these properties were only available for some of the devices seen in our dataset, further reducing our dataset to 46 datapoints for the sake of this analysis.

3.4 Analysis

Given our final dataset of 46 mean perceived error measures taken from 32 papers, we're able to investigate the influence of several aspects of VR displays on perceived distance.

Weight

Previous work has directly explored the effect of headset weight on perceived distance, with somewhat conflicting results. One factor which has been thought to contribute to distance misperceptions is the weight of the device. Willemsen et al. (2004) produced a mock HMD using the shell of an NVIS nVisor SX, reproducing the weight and FOV restriction of that HMD while allowing participants to view the real world. They saw underestimation with this mock HMD, though not as much as as when viewing virtual stimuli; Willemsen et al. (2009) produced an intertial headband that reproduced only the weight of the HMD, down to correct moments of inertia, and found slight but not statistically significant underestimation. Buck et al. (2018) explore two conditions in which the a DK1 is modified to approximate that of an NVIS SX60. They find that the increased weight elicits greater underestimation, but not to the degree seen with the NVIS HMD.

Figure 3.7 shows a linear regression of HMD weight over time. There has been a significant ($p < .001$) trend towards use of lighter headsets over time, though the overall range remains similar. Figure 3.6 shows a linear regression of percent error in perceived distance over HMD weight. There is a trend of reduced error with reduced weight, but it is not significant ($p = .0874$). We cannot say that the change in weight over time has had an effect on perceived distance.

Horizontal Field of View

Previous work has directly explored the effect of restricted field of view on perceived distance, with somewhat conflicting results. Willemsen et al. (2004) produced a mock HMD using the shell of an NVIS nVisor SX, reproducing the weight and FOV restriction of that HMD while allowing participants to view the real world. They saw underestimation with this

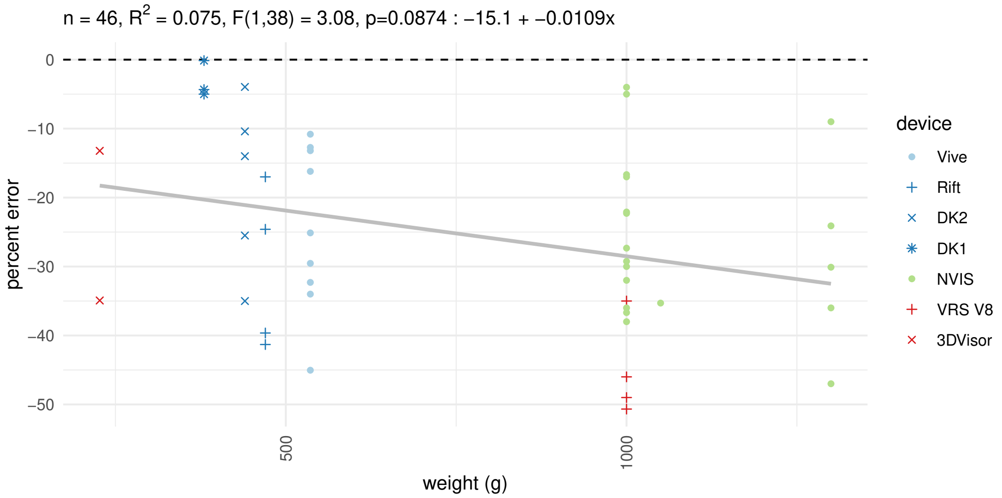

Figure 3.6: Percent error by weight, for the subset of experiments using blind walking and hardware with documented weight.

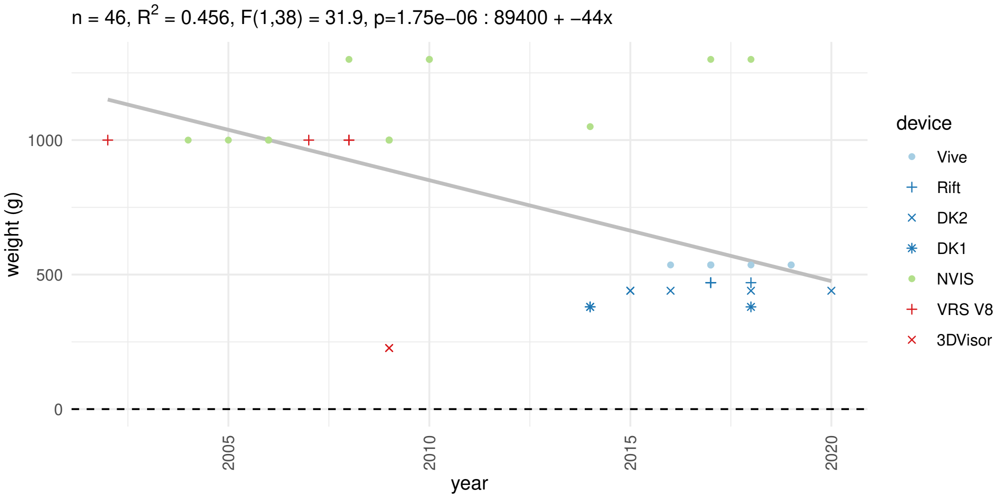

Figure 3.7: Headset weight over time.

mock HMD, though not as much as as when viewing virtual stimuli in the matching HMD. Creem-Regehr et al. (2005) build custom occluding goggles to restrict FOV when viewing the real world, and find some but not significant levels of underestimation; when they also use a neck brace to restrict head movement, they find significant underestimation. Buck et al. (2018) modify a DK1 to approximate the FOV of an NVIS SX60. They find that the restricted FOV elicits a similar amount of underestimation as seen with the NVIS HMD.

Horizontal field of view was gathered from documented hardware specifications; for the VRS V8 and 3DVisor, the horizontal field of view are estimated from documented diagonal field of view. Figure 3.9 shows a linear regression on horizontal field of view over time. Horizontal field of view has increased significantly over time ($p < .001$). Figure 3.8 shows a linear regression of percent error in perceived distance over horizontal FOV. No significant effect is found ($p = .968$). We cannot say that the change in FOV over time has had an effect on perceived distance.

Resolution and Clarity

Display resolution may be where we would expect the most progress over time - cell phone, television, and monitor resolutions have improved significantly since 1995, HMDs may have as well. Greater resolutions would be expected to facilitate a clearer display. When viewing a flat screen straight on, the apparent resolution received by the viewer's eye is determined by the distance between the eye and the screen. For HMDs, this calculation is complicated by their lenses, which would be expected to change pixel density across field of view differently in each HMD, possibly further influenced by eye pose, eye-lens alignment, software predistorts, and other aspects of the display. Unfortunately these complicating factors for determining resolution are largely ignored by manufacturers and the literature, and developing methods to quantify them is beyond the scope

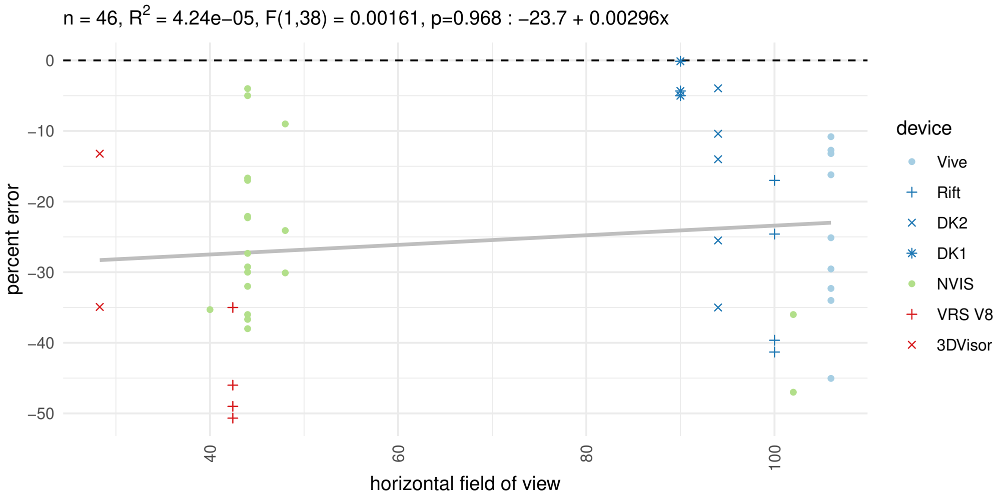

Figure 3.8: Percent error by horizontal field of view, for the subset of experiments using blind walking and hardware with documented field of view; VRS V6, V8, and 3DVisor headset's horizontal field of view are estimated from documented diagonal field of view.

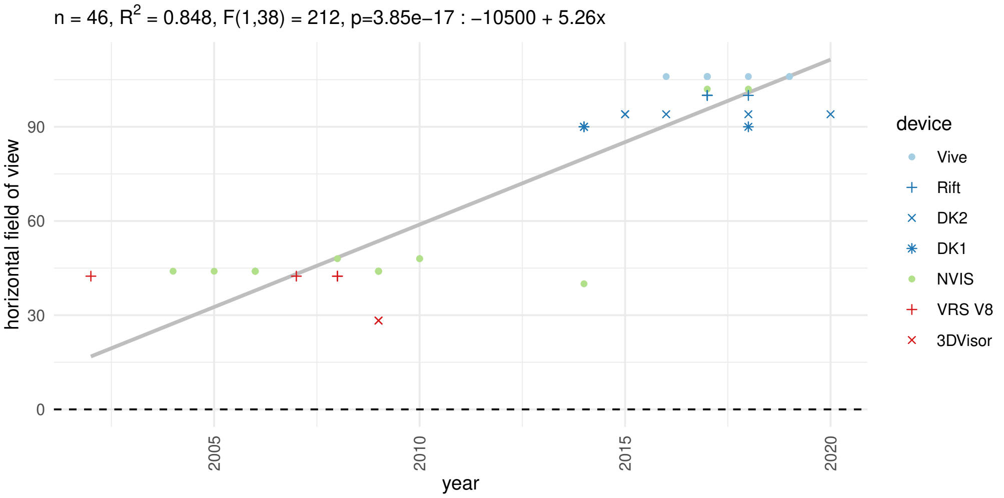

Figure 3.9: Horizontal field of view over time.

of the current work.

Perhaps the best approximation current data can provide is to calculate the spatial resolution as a function of display resolution and FOV:

$$\text{spatial resolution} = \frac{\text{horizontal resolution}}{\text{horizontal FOV in degrees} * 60} \tag{3.1}$$

This provides the pixels per arcminute a given display would be expected to subtend on the retina. This metric corresponds to visual acuity (Bach, 2006; Fidopiastis et al., 2005; Peer and Ponto, 2020), with a value of 1.0 corresponding roughly to the clarity necessary to facilitate 20/20 vision, and lower numbers indicating less visual clarity. We emphasize again that this is expected to be a poor approximation for the actual visual stimuli received when using one of these HMDs, but it may be the best approximation given current data.

Figure 3.10 shows a linear regression on the changes in display spatial resolution over time. The trend is significant ($p < .001$), and its direction may be surprising as it shows that spatial resolution has been reduced over time. This is due to increasing FOVs outpacing the growth of pixel density.

Figure 3.11 shows a linear regression of percent error in perceived distance over spatial resolution. The effect of spatial resolution was not found to be a significant influence on distance misperception ($p = .398$).

3.5 Limitations and Recommendations

Distance misperception research is a relatively young field, involving all the complexity of human perception, and all the fast-moving evolution of computer graphics and immersive displays. In that light, it is no surprise that there might be many complicating factors limiting attempts to compare data from decades of experiments. Some limitations of this

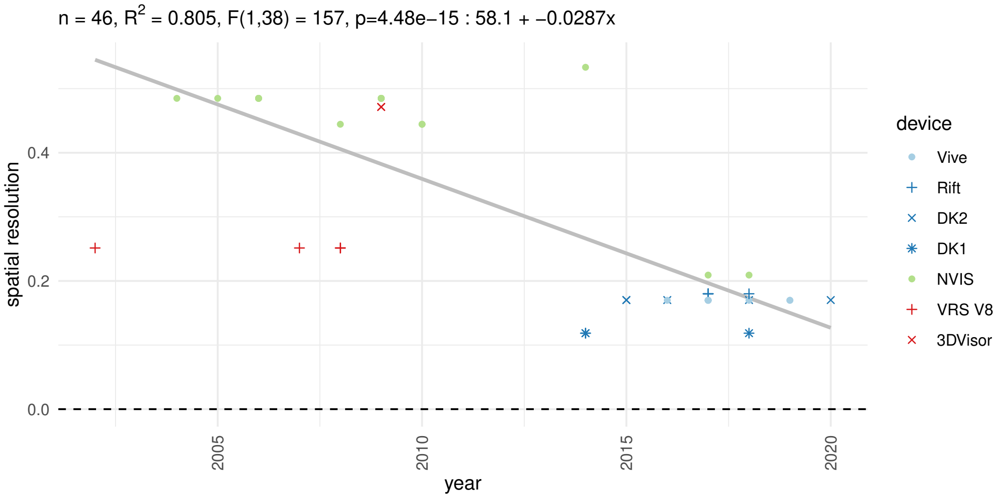

Figure 3.10: Spatial resolution over time. Older science-grade headsets have greater or equivalent spatial resolution, due to distributing their pixels across a smaller FOV.

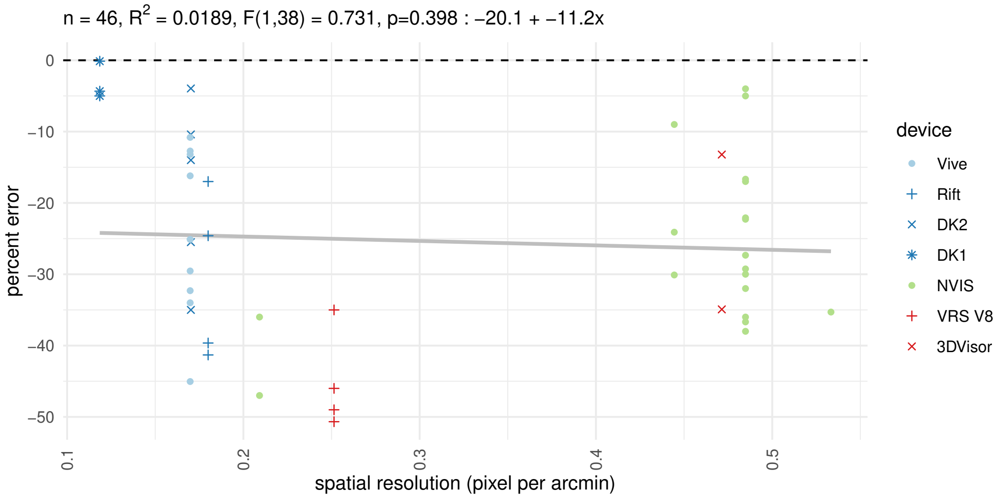

Figure 3.11: Percent error by spatial resolution, for the subset of experiments using blind walking and hardware with documented or estimable pixels per arcminute.

work are due to the technique and assumptions used to make comparisons practical, or simply the scope practical within a single work; these might be addressed by different approaches or focuses in future work. Some limitations are due to the factors that complicate comparisons of literature in the field, and these might be addressed by future researchers adopting changes in method and reporting in future experiments.

This sample of the distance misperception research literature is explicitly not exhaustive, as many works were omitted due to differences in reporting, method, and devices used. A more complete sample might lead to different conclusions. This is a limitation. It is also somewhat the point: it is my hope that this work inspire further efforts to report future work in (and perhaps transform more past work into) some more standardized form that affords more comparisons between the experiments performed by different researchers.

It may be possible that the factors we observe in our analyses interact in complex ways with perceived distance and each other, and different methods of analysis and more complex models are needed to observe any potential effects. While this is possible and could make for productive future work, this work still shows that several aspects of head mounted displays have changed over time without having a strong effect on perceived distance. Perhaps more importantly, it describes several complicating factors that make it difficult to establish comparisons between larger groups of studies that might make more complex models more practical and more informative.

Over two hundred papers were surveyed to find the set of forty that reported error in perceived distance in a form that could be compared with some confidence. This set was further limited by concerns that different methods of measuring perceived distance may yield very different perceived estimates, even by the same person viewing the same stimuli (Peer and Ponto, 2017). In general, the field has tended to express error in perceived distance as a percentage of the target distance, under the assumption that this yields an error uniform enough across different distances that most analysis can be done ignoring the influence of distance. That this percent error is used so frequently makes it the best way to report distance misperceptions for cross-experiment comparisons. When assumptions of uniformity over distance are violated, another transformation of the data may be more appropriate for some analysis, but it still might be worth reporting percent error for the sake of cross-experiment comparison. Further, (Peer and Ponto, 2017) suggest expressing error in perceived distance as the difference between distance estimates made when viewing real and virtual stimuli, yielding a measure called relative percent error; adopting relative percent error should afford more accurate comparisons between studies using the same method of measurement, and might afford better comparisons between methods. If the measurement methods and reporting of error in perceived distance were more standardized, more papers could have informed not just this work, but future research in the field in general. Or better still, if all data gathered in an experiment could be distributed in some standardized format alongside publications, future researchers could make whatever transformations best serve their future needs.

The visual stimuli presented was not consistent between different studies: some studies present photorealistic environments, while some present abstract or otherwise cue-sparse stimuli. There are no established objective metrics to quantify these differences, though establishing one could make for impactful future work. Further, the visual stimuli used in experiments are often poorly described or simply omitted in publications; in part this is a limitation of mapping immersive stimuli to the static two-dimensional format of academic publishing. It may be possible to share richer representations of experimental stimuli. Virtual stimuli can be shared as an executable or as source code, though this depends on associated operating systems and software libraries remaining viable. Both real and virtual spaces might be converted to a volumetric representation (e.g. point-cloud), using photogrammetry or LiDAR-based capture methods for real spaces, and something like simple voxelization for virtual. Sufficiently detailed video explorations of the spaces might be a viable approximation for volumetric representations. Any method that allows future researchers to view a spatial visual stimuli first hand could facilitate detailed analyses that an image on a page simply never will.

Another possible solution is to establish some common or at least baseline visual stimuli for researchers to use; using the same visual stimuli would remove it from consideration as an influencing factor in cross-experiment comparisons. An exact match across all experiments may be impractical for experiments involving real environments, which may be more difficult to standardized; standard virtual environments might be given some set of adjustable parameters so as to adapt to provide spatial stimuli comparable to their real counterparts, with some attention to not accidentally inducing matched environment effects seen by Interrante et al. (2006, 2008). Even in the case that using identical virtual stimuli is impractical, using a common stimuli as a baseline might make easier the cataloguing or quantifying of changes from baseline between experiments. The idea of establishing a common set of stimuli isn't entirely novel: see the The Virtual Environment Performance Assessment Battery (VEPAB) of Lampton et al. (1994), the related augmented-reality-focused battery (ARPAB) of Kirkley Jr (2003), or the several tasks in the testbed of Bowman et al. (2001). The methods described in these works don't see use through most of the distance perception literature, but they were developed in an era where code distribution may have been more difficult. With the general accessibility of somewhat standardized interactive software development platforms (e.g. Unity and Unreal Engine), alongside the ubiquity of code distribution and collaboration platforms (e.g. Github, Bitbucket), efforts to establish a standardized set of research tools may have a better chance to see wider and longer-term adoption.

There may be differences between the populations these studies draw participants from, possibly having different biases in factors such as age, gender, or experience with virtual reality. It is also unclear how the overall population may have changed over time - is the average participant in 2020 more accepting of being immersed than those of 1995? Is their concept of "immersed" the same? Some demographic data may have been reported in detail enough to afford some analysis; this was not explored. Aside from aggregate demographics, there is some evidence that individuals experience different degrees of distance misperception (Peer and Ponto, 2017, 2016a, 2019); reporting every participants' data in a publication may be infeasible due to space considerations, but this further argues the case for sharing all experimental data alongside a publication. Measuring and reporting more of the differences between individuals, and more aspects of their perceptual experience, might also help determine the sources of the difference in their distance misperceptions. Peer and Ponto (2020) suggest a method of measuring individual's visual acuity and stereo accuracy as mediated by immersive displays; one could imagine a battery of similar measures of individuals' perceptual differences being performed alongside each experiment to better lend the specifics of their perceptual experience to the interpretation of their misperceptions.

We perform some analysis on the characteristics of different head mounted displays, but other characteristics of these devices that were not explored may be of interest. The fidelity of the tracking system may contribute to a participant's understanding of the space, and this might be expected to have changed over time; this was not explored, but may be reported widely enough to afford some analysis. Devices may also have adjustable components: lenses that adjust to accommodate the interpupillary distances of different viewers, which when misaligned may influence participants' sense of depth; adjustable lens depth to accommodate different eyeglasses or different face shapes, which would be expected to change FOV; additional batteries, counterweights, or cushioning that increase weight or otherwise change fit. An argument might be made to standardize the method of making some adjustments, such as for IPD, but many others are difficult to measure, and more are difficult to predict the utility of when allocating space in publications. At minimum, it might be enough to catalogue any changes made to display systems that deviate from factory defaults, and report the value selected for any adjustable aspects the system that are easily quantifiable, such as the current IPD adjustment selected. Standardized methods for calibrating systems to individuals would also help to establish that something like IPD had been selected in a comparable way.

Even given the limitations of our ability to make comparisons between experiments, we can still say with some confidence that researchers have, across these methodological differences, found a similar range of misperceptions in virtual reality across the last two-and-a-half decades.

3.6 Chapter Conclusions

In this work we analyze the results of publications from the past 25 years which have studied distance perception in virtual reality. While there seems to be a slight trend towards a reduction of misperceptions over time, the range of these results remains similar throughout the history of the topic. Analysis of several factors that might have changed over time and might be expected to influence distance misperceptions were similarly not found to have a significant influence. This broad survey approach has limitations, but if the factors subject to our analyses were to have had a large effect, one would expect our analyses to have observed it.

Experiments studying distance misperception in VR can investigate many influencing factors, leading to a combinatorial explosion of possible conditions for investigation. Better facilitating comparisons between publications means more of this space can be studied more conclusively, with fewer experiments. It might make the next 25 years of research more productive; it might mean we don't have another 25 years to wait before distance misperception is more fully understood. In particular we feel that it is critical to:

- report data in ways that afford cross-experiment comparisons, such as by using relative percent error
- report adjustments made to hardware, and the current state of adjustable components (eg IPD, lens depth)
- develop a standard visual stimuli, and/or a method to quantify differences between stimuli
- develop methods of preserving spatial stimuli
- when at all possible, make data and stimuli publicly available

This not an exhaustive list, but adopting these changes would make comparisons as were attempted here more feasible and more informative.

4 METHODS OF MEASURING PERCEIVED DISTANCE

This study was meant to establish a baseline for distance misperceptions when viewing contemporary immersive stimuli, across perceived distance measures and contemporary devices. There was some concern that different results seen in the literature might have been due to different measurement methods being used, and that measurements made using different methods could not be meaningfully compared to each other. This study provides evidence of individuals experiencing different degrees of distance misperception, and proposes that this degree of misperception be defined as the difference between an individual's percent error in perceived distance when viewing real and virtual stimuli. This *relative percent error* makes measurements more similar across measurement methods, and could allow better comparisons between experiments that adopt it in the future. [1]

4.1 Overview

In this chapter, a user study is used to evaluate measures of percieved distance that can be performed within the relatively small (4m x 4m) space tracked by a consumer-grade VR device. One could imagine these sorts of measures being used by consumers to measure, and possibly correct for, the degree of distance misperception they are experiencing, and the ability to use consumer-grade hardware would make distance misperception research to more accessible more researchers.

There is also the perennial question of whether contemporary VR displays have improved so as to eliminated distance misperception: some studies suggest so (Creem-Regehr et al., 2015; Li et al., 2015, 2014; Young et al., 2014), though other work suggests that the most obvious improve-

---

[1] This chapter is adapted from Peer and Ponto (2017)

ments, such as larger field of view, should not have so strong an effect (Willemsen et al., 2008; Grechkin et al., 2010). Two different contemporary head-mounted displays (HMDs) are employed, to establish if either greatly reduces or eliminates distance misperceptions.

Not many studies have directly compared different perceived distance measures using the same visual stimulus and presentation context. Of those that have, they generally establish performance relative to blind walking (Sahm et al., 2005; Plumert et al., 2005; Grechkin et al., 2010; Napieralski et al., 2011), which is impractical in small spaces. In addition to blind walking, Napieralski et al. (2011) compare two measures practical in small spaces, finding blind reaching to be more accurate and precise than verbal estimates. Klein et al. (2009) contribute a rare work directly comparing three measures suited to small spaces (verbal estimates, timed imagined walking, and triangulated walking), and find one (triangulated walking) to exhibit significantly more underestimation. The lack of work directly comparing measures using the same visual stimulus and presentation context makes comparisons between works using different measures difficult - and the work that does make these comparisons provides evidence that measures made using different methods should not be expected to be equivalent. The work described in this chapter seeks to provide a baseline across measures using contemporary displays and stimuli, and a method of reporting results that better facilitates comparisons across experiments. A fully-within participant design is employed, observing each participant's performance with all measurement techniques under all viewing conditions - both headsets, real and virtual environments. This unified perceptual context provides for stronger comparisons of the distance misperception observed between measures and devices.

4.2 Experiment

Virtual scenes were displayed using two consumer-grade head-mounted displays (HMDs): an Oculus Rift CV1, and an HTC Vive. Manufacturer specifications suggest these HMDs have similar display characteristics: OLED screens with 2160x1200 resolutions and 90Hz refresh rates, and a 100 degree field of view. Their weight varies slightly, with the Rift (470g) weighing slightly less than the Vive (555g). Both HMDs were adjusted for an interpupillary distance of 63.5mm for all participants. Rendering for both HMDs was provided through the Unity game engine, using the SteamVR. Positional tracking was provided by each headset's respective tracking solution; the Vive provided tracking for two handheld controllers, used by participants and experimenters to register participants' distance judgements. Each headset had a separate dedicated computer for rendering and tracking, both located in the same room and synced over a network to show the desired stimulus; a Vive controller was used by participants when wearing both headsets.

The virtual environment (VE) displayed in the HMDs was made to be a rough, non-photorealistic approximation of the dimensions, color, and visual texture of the real environment (RE). We choose not to use a photorealistic match to avoid the previously studied effects on perceived distance of presentation order (Ziemer et al., 2009) and transitional environments (Steinicke et al., 2009b). During trials, the virtual environment displayed a yellow triangle, used as the target of the participant's distance judgments. A matching physical target was constructed using cardboard, tape, and colored paper. Examples of the target stimuli can be seen in Figure 4.1. The physical target was placed before real environment trials, using a Vive controller to guide the experimenter to the next target location using a flashing LED and haptic feedback. Participants were asked to close their eyes while the target was placed.

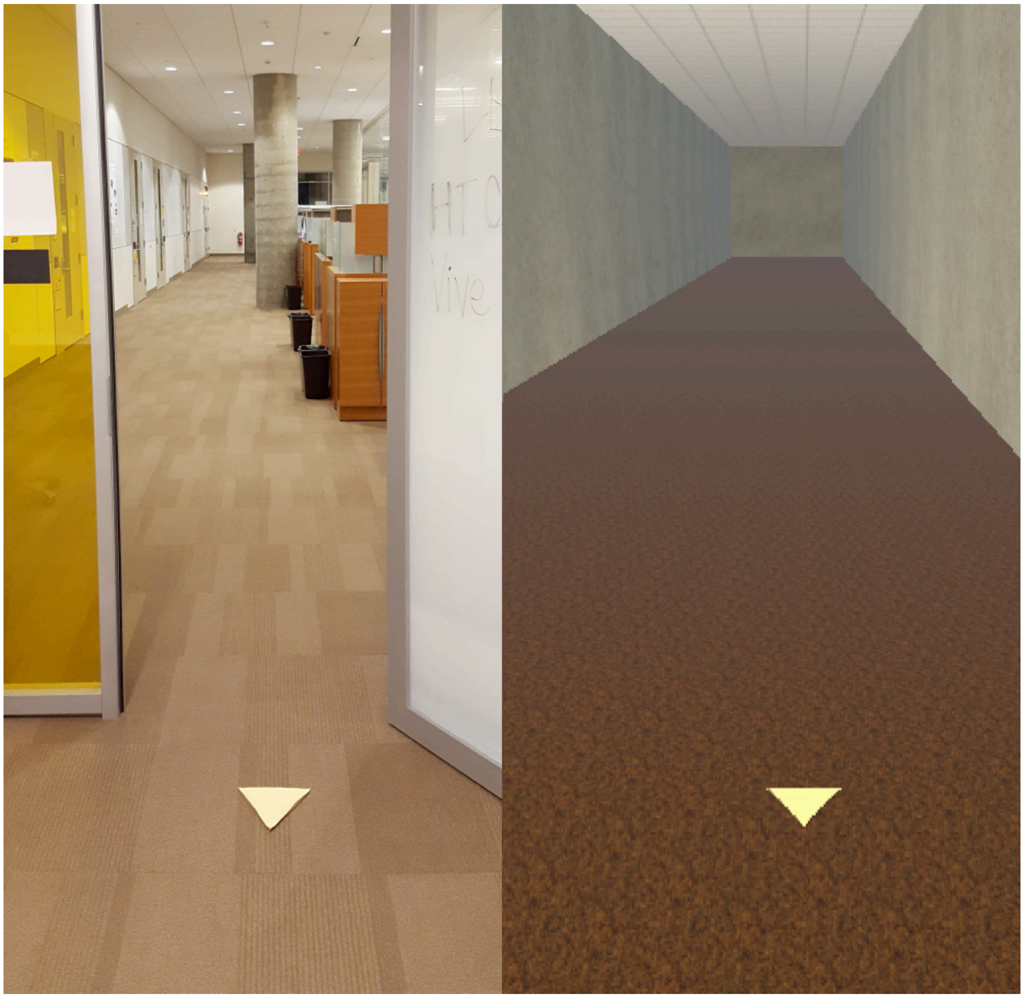

Figure 4.1: The physical (left) and virtual (right) environments used in the experiment, as viewed by the participant.

Figure 4.2: The experimental environment. A confederate demonstrates the blind throwing protocol while wearing the HTC Vive.

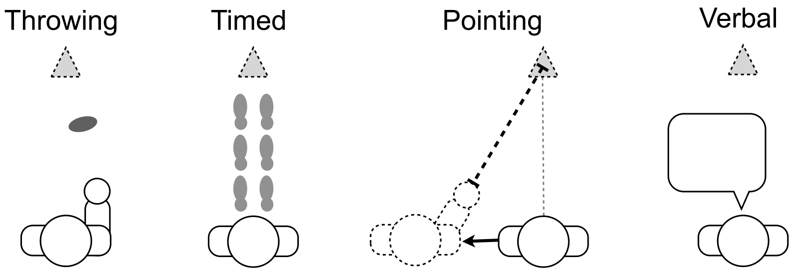

Figure 4.3: The four measures of distance perception used in this experiment: blind throwing, time-imagined walking, blind triangulated pointing, and verbal report.

Participants

While 19 participants were recruited for the study, only 17 participants are included in the evaluated data: one participant's data was removed, as they were unable to complete the full experiment due to time constraints; the other participant was removed due to recording extreme distances of 100 to 400 meters during the real environment condition of the triangulated pointing task, which is suspected to have been caused by a tracking or task performance error. Our final sample of 17 participants consists of 11 men and 6 women ranging in age from 20 to 64 $(M = 27.6, SD = 10.1)$. Participants were recruited from a local university campus. 5 participants had substantive experience with virtual reality, computer graphics programming, or had participated in previous studies affiliated with our lab, but had no knowledge of the current experiment; the other 12 had no previous experience with virtual reality.

Perceived Distance Measures

A variety of perceived distance measures have been established in the psychometrics and VR literature, but not all are practical within a small space. Blind Walking, in which a participant views a target, closes their eyes, and walks to where they remember the target to have been is one of the most used methods of measuring perceived distance (Renner et al., 2013). However, its use presents some challenges: the walker needs to be guided to avoid obstacles in an uncontrolled environment, or often to walk in a straight line even in a controlled one; the act of walking may also provide information that influences distance estimations (Philbeck et al., 2008), particularly if the real world is visible through an incomplete seal between the HMD and the wearer's face (Jones et al., 2013, 2011). For small spaces and uncontrolled environments, other measures may be more appropriate. In this experiment, we explore four other established measures which can be executed in small spaces, which are depicted in Figure 4.3 and further described in this section.

Blind Throwing As implemented here, blind throwing follows the methods proposed by Peer and Ponto (2016a) and Sahm et al. (2005). After viewing the target, participants close their eyes and throw a beanbag at the target. Participants were instructed to aim the center of the beanbag at the center of the target, and that the beanbag's initial point of impact would be recorded. Perceived distances were recorded by placing a tracked controller above the beanbag's point of impact and pushing a button to record its position, projected onto the floor plane and on the axis running between the participant and the target. Pilot trials suggested that participants would be hesitant to throw beanbags blind without practice, so they were allowed training throws at up to three target distances displayed in the real environment; training was terminated when participants felt comfortable with the task. Training distances were at intervals not seen in evaluated trials. It should be noted that most participants threw at only one target, or waived practice altogether.

Timed Imagined Walking Similar to the implementations used by Grechkin et al. (2010), Plumert et al. (2005), and Ziemer et al. (2009), timed imagined walking asks participants to judge the amount of time they imagine it would take to walk to the target, using their separately measured average walking speed to convert this into a distance. In our experiment's implementation, the participant was first taken to a nearby hallway and asked to walk between two lines of tape placed 8 meters apart while the experimenter timed them with a stopwatch. This was repeated twice, and their average time to walk 8 meters was used to calculate their average walking speed. Participants then returned to the main experimental space and resumed trials. When performing the measurement task, participants held a tracked controller, viewed the target, then closed their eyes and imagined walking to the target's location. Participants pressed a button when they began their imagined walk, and again when they imagined they had reached the target's location; the first press began a timer, and the second stopped the timer and recorded the result. Perceived distance to the target was calculated by multiplying the participant's average walking speed, in meters per second, and their imagined walk duration, in seconds.

Blind Triangulated Pointing This measure saw use by Bruder et al. (2015), and is a space-bound adaption of the triangulated blind walking technique used by many (Klein et al., 2009; Richardson and Waller, 2007; Thompson et al., 2004; Willemsen et al., 2009), and the several techniques presented in Fukusima et al. (1997). When performing this measurement task, participants hold a controller whose position and orientation are tracked. After viewing the target participants close their eyes, take two steps to their left, then point a tracked controller toward the target and push a button to record the measurement. Before beginning the pointing phase of the experiment, participants viewed a demonstration by the experimenter and were asked to demonstrate themselves. During this demonstration, participants viewed a real-world target at a single distance interval different from those seen in the evaluated portion of the experiment. After recording their measurement, participants were prompted to return to their original position. Perceived distance was calculated by casting a ray from the position of the held controller in the direction the controller was pointed; the intersection between this ray projected onto the ground plane, and the axis running between the participant and target was taken to be the perceived position of the target. This implementation ignores any error in the vertical angle of the participant's pointing.

Verbal Report This method of perceived distance measurement has seen several variations in the literature (Klein et al., 2009; Kunz et al., 2009; Proffitt et al., 2003; Mohler et al., 2006). Our implementation is most similar to Klein et al. (2009), as we simply ask participants to close their eyes and tell us how far away the target seemed, using whatever unit of measure they are most comfortable with (12 used feet, 5 used meters, and 1 used centimeters).

Procedure

Participants were asked to stand above a tape mark on the floor of the physical world that corresponded to the origin of target distances in the tracked coordinate system and virtual environment.

Virtual environment trials progressed as follows: Participants begin with their eyes closed. By pushing a button they trigger a chime prompting them to open their eyes, as well as show the scene containing the target on the display device's screen; after three seconds the scene disappears, an audio cue plays to prompt participants to close their eyes, and they then perform the measurement task.

Real environment trials progressed similarly, but, as the experimenter manually placed the target between trials, participants waited to view the scene until prompted. In these real environement trials, the physical target is placed at a location matched to the virtual targets' positions in the tracked coordinate system by using the haptic feedback and flashing LEDs of a tracked controller to indicate a distance within $\pm 0.25\text{m}$ of the base target distance for the trial – the larger region being faster for the experimenter to find, and allowing for an approximation of the random jittering used in VE trials (see §4.2 : Methods). After the experimenter registered the position of the target by positioning the controller above it and pressing a button, they then moved away from the target and prompted the participant to pull the trigger and open their eyes. Three seconds later an audio cue prompted the participant to close their eyes, and they performed the measurement task.

Participants triggered trial progression themselves in all conditions but the blind throwing measure, where they held beanbags rather than a controller; in this case, the experimenter triggered the next trial, after verbally warning the participant.

Before the experiment began, participants read and signed a consent form. They then were introduced to the audio cues meant as prompts for opening and closing their eyes. Next, they were shown how to put on and adjust the fit of both headsets, and instructed to adjust them for clear viewing and comfortable wear. They were then introduced to the controller, and shown how to operate the trigger. The general form of the experiment was then described – pull trigger, view a scene, perform an action. The first measurement task was then introduced. When switching to a new measurement task, an explanation of the task was given (see: §4.2 : Perceived Distance Measures), and it was confirmed that participants understood the task. When switching between HMDs, participants were allowed to re-adjust the headset for comfort and clear viewing.

If participants made a clear mistake in executing a measurement task, the scene was reset and participants were able to try again; 3 participants had two trials reset, and 4 participants had one trial reset.

Methods

The experiment follows a within subjects, repeated-measures design, with all participants experiencing all conditions.

The independent variables are the perceived distance measurement task (*measure*), display device (*display*), and target distance (*distance*). *Measure* is a factor of four levels: (*throwing*, *pointing*, *walking*, *verbal*). *Display* is a factor of three levels: the RE (*real*), and the VE in both HMDs (*rift*, *vive*). Distance is a factor of three levels, corresponding to the three base distances used in our experiment: (*2m*, *3m*, *4m*).

The dependent variable is *percent error* (PE) of measured perceived distance relative to actual target position, with negative values indicating underestimation.

Percent error in VE conditions relative to that observed in the RE is also investigated; I call this transformation of the dependent variable *relative percent error* (RPE). This transformation is intended to better capture error in perceived distance, independent of individual participants' difference in task performance. Relative percent error was derived by first calculating mean RE percent error for each participant, task, distance, and device combination; this value was then subtracted from the percent error observed for each VR trial under the same conditions. Note that with regards to relative percent error, *device* is a factor of only the two VE condition levels (*vive*, *rift*).

The experiment progressed through measures in a random order; for each measure, the displays were presented in a random order; for each measure and display combination, a series of 9 distances (three repetitions of 2m, 3m and 4m), were presented in a random order. That is, each combination of measure, display, and distance was presented 3 times, for a total of 108 trials per participant. Measure was chosen as the factor at the top of the randomization hierarchy, as we were concerned that participants might forget how to perform a measure over time. Displayed target distances were randomly positioned within $\pm 0.25\mathrm{m}$ of the selected distance, so that the repetition might be less obvious: adjusted distances are called jittered target distances; unadjusted distances are called base target distances. After the experiment, a brief demographic survey and post-experiment interview was conducted.

We aimed to investigate the following questions:

Q1 Are distances misperceived in the HMDs, relative to the physical world?

Q2 Does perceived distance change between measurement methods?

Q3 Does perceived distance change between HMDs?

Q4 Is distance misperception affected by distance?

4.3 Analysis

Our results were analyzed using repeated measures ANOVA and Tukey's multiple comparisons test with a 5% significance level. When Mauchly's test indicated that the assumption of sphericity had been violated, Green-Geisser estimates of sphericity were used to correct degrees of freedom. The results of these analyses are summarized in Table 4.2; descriptive statistics are summarized in Table 4.1.

A 3(device) x 4(measure) x 3(distance) repeated measures ANOVA was performed investigating effects on percent error of perceived distance (percent error). A significant main effect of display was found ($F(2, 32) = 49.06, p = 0.001$). Post-hoc tests show that distances are perceived as significantly different ($p < 0.001$) in the real condition compared to either the Vive or Rift conditions; as seen in Figure 4.5 and the left portion of Table 4.1, distances in the two virtual environment conditions are perceived to be closer than in the real condition (**Q1**).

No significant effect was found between percent error measured in the Rift and Vive (**Q3**). Significant effects were found for the interaction between task and device ($F(6, 96) = 4.92, p = 0.002$), task and distance ($F(6, 96) = 6.44, p < 0.001$), and the three-way interaction between task, device, and distance ($F(6, 96) = 2.86, p = 0.013$).

Due to the presence of underestimation in real environment conditions, we constructed a new measure of percent perceived distance relative to mean estimates in the real environment (relative percent error). Due to the interaction effects seen between task, distance, and device, means for real condition estimates were calculated by pooling by participant, task, distance, and device. Relative percent error was then calculated by subtracting these real condition means from virtual condition percent error values, similarly matched by participant, task, distance, and device. This yields a new data set with the additional measure of relative percent error, and where *device* is a factor of only the two virtual condition levels (*vive*, *rift*). Relative percent error is depicted in Figure 4.4, and is summarized in the right portion of Table 4.1.

A 2(device) x 4(measure) x 3(distance) repeated measures ANOVA was preformed investigating effects on relative percent error. The results can be found in the RPE section of Table 4.2.A significant main effect of task was found ($F(3, 48) = 3.96, p = 0.013$), with significant interactions between task and device ($F(3, 48) = 6.53, p = 0.001$) and task and distance ($F(6, 96) = 2.86, p = 0.013$). Post-hoc tests on device show significant differences between all combinations of measures ($p < 0.001$), except verbal and throwing, and verbal and pointing. Post-hoc tests on the interaction between task and distance show: for throwing, a significant difference between 4m and 2m ($p = 0.008$), showing about 4% more mean underestimation at 4m; for timed, no significant difference between distances; for pointing, a significant difference between 2m and both 3m ($p = 0.002$) and 4m ($p = 0.002$), showing about 6% more mean underestimation at 2m;

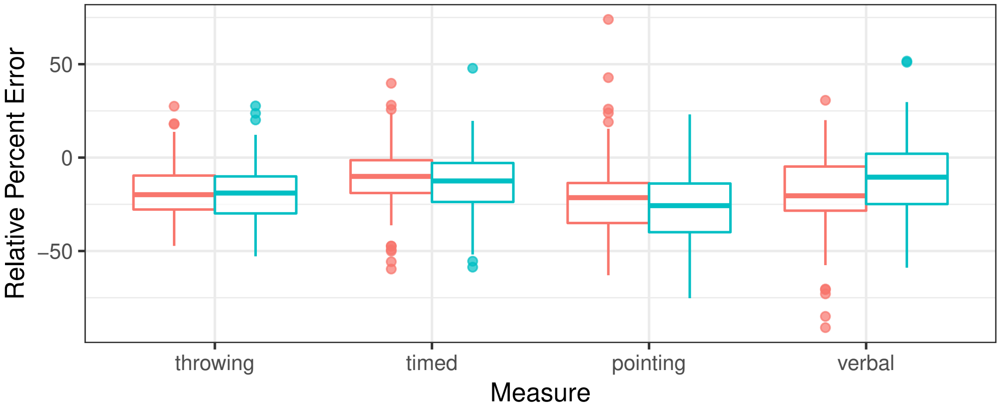

Figure 4.4: Percent error relative to mean estimates made in the real environment, divided by task and colored by display device: red is rift, vive is blue.

for verbal, a significant difference between 4m and both 2m ($p = 0.008$) and 3m ($p = 0.01$), showing about 5% more mean underestimation at 4m. Post-hoc tests on the interaction between task and device show: in the Rift, a significant difference between timed and all other measures ($p < 0.001$), yielding 7-12% less mean underestimation, and no other significant differences; in the Vive, pointing yields 9-17% more mean underestimation than all other measures ($p < 0.001$), and throwing additionally yields 3-5% more mean underestimation than timed ($p = 0.002$) and verbal ($p < 0.001$).

The interaction effects involving device complicate conclusions regarding **Q3**. The main effect and interaction effects involving task complicate conclusions regarding **Q2**. The interaction effects involving the factor distance complicate conclusions regarding **Q4**.

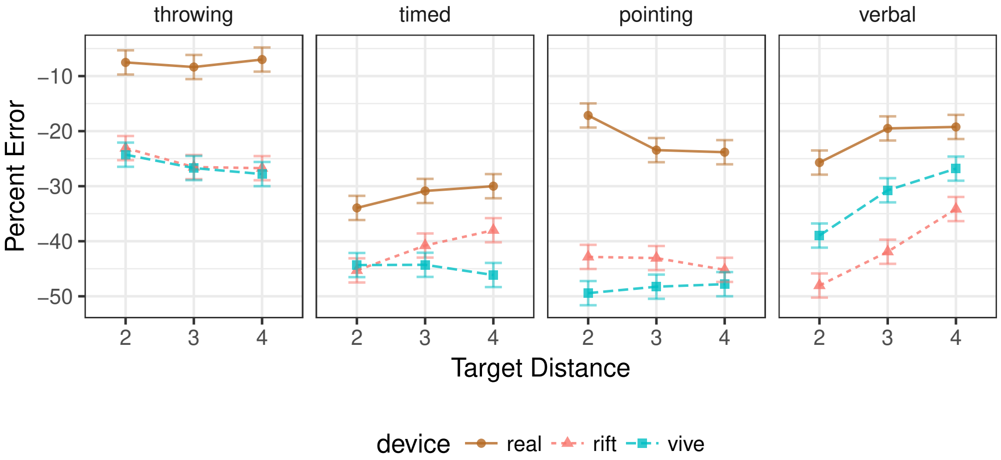

Figure 4.5: Percent error of perceived distance, by device, measure, and target distance. Target distances are unjittered base target distances. Error bars show 95% confidence intervals.

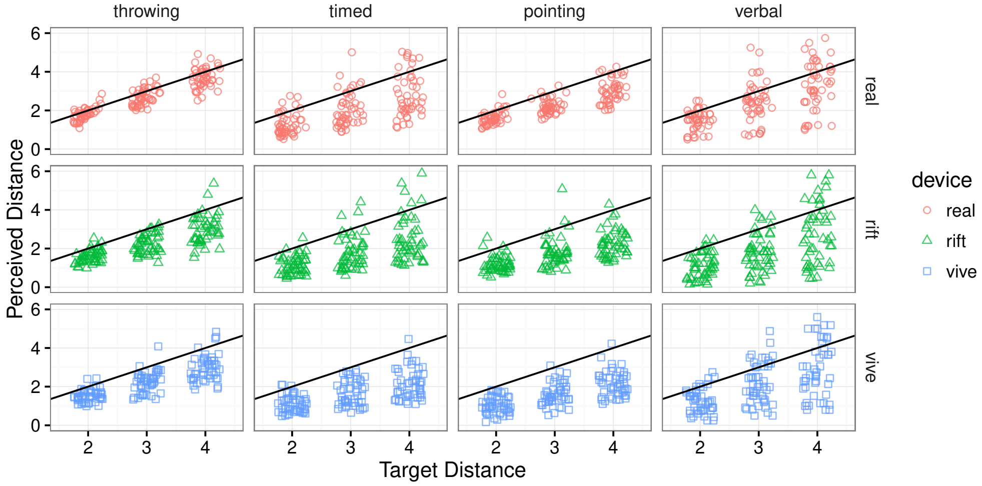

Figure 4.6: Scatterplots of perceived distance relative to jittered target distance, both in meters. Black lines indicate veridical perception; points beneath the black line indicate perceived distance measured to be closer than intended.

Task	Display	N	Error (%) Mean	Error (%) SD	Relative Error (%) Mean	Relative Error (%) SD
throwing	real	153	-7.96	11.07	–	–
	rift	153	-24.60	14.13	-17.82	14.37
	vive	153	-25.12	14.75	-18.64	15.19
timed	real	153	-31.16	26.04	–	–
	rift	153	-41.31	24.69	-9.75	16.08
	vive	153	-45.04	19.35	-13.31	16.59
pointing	real	153	-22.01	9.64	–	–
	rift	153	-42.24	18.43	-22.23	20.67
	vive	153	-47.10	16.11	-27.02	18.50
verbal	real	153	-19.90	30.20	–	–
	rift	153	-39.64	31.68	-19.89	20.35
	vive	153	-29.54	32.67	-10.69	19.41

Table 4.1: Summary of percent error in measured perceived distance per device and task; unadjusted error is described on the left, and the two rightmost columns describe error relative to mean error in real environment conditions, calculated separately for each combination of participant, task, device, and distance.

4.4 Discussion

We see clear distance misperception in the underestimation in VE conditions. VE conditions show 37.97% ($SD = 25.75$) mean underestimation overall, 17% more than RE conditions. This difference is clearly visible in Figure 4.5. PE for all devices and measures ranges from 22% to 47% mean underestimation. These results are in keeping with previous literature (Renner et al., 2013), suggesting that distance misperception should be expected in contemporary consumer hardware.

We also see significant underestimation in the real environment conditions, with a mean of -20.54% ($SD = 24.32$), vary different amounts seen between measures. This error suggested our implementation of rela-

Effect	On	Df_n	Df_d	F	p	Sig.	η^2	η_P^2
Display (D)	PE	2	32	49.06	$< .001$	***	.12	.75
Measure (M)	RPE	3	48	3.96	.013	*	.09	.20
Display (D)	RPE	1	16	0.00	.996		$< .01$	$< .01$
Distance (Dst)	RPE	2	32	1.62	.214		$< .01$	$< .01$
M x D	RPE	3	48	6.53	.001	***	.03	.29
M x Dst	RPE	6	96	2.86	.013	*	.02	.15
D x Dst	RPE	2	32	0.74	.447		$< .01$	.04
M x D x Dst	RPE	6	96	1.48	.192		$< .01$	.08

Table 4.2: ANOVA results of interest. Df_n and Df_d are the numerator and denominator degrees of freedom, F is the F-value, p is conditional probability of the F-test, η^2 is eta-squared, η_P^2 is partial eta-squared. All effects on RPE are shown; several effects on PE are omitted for space.

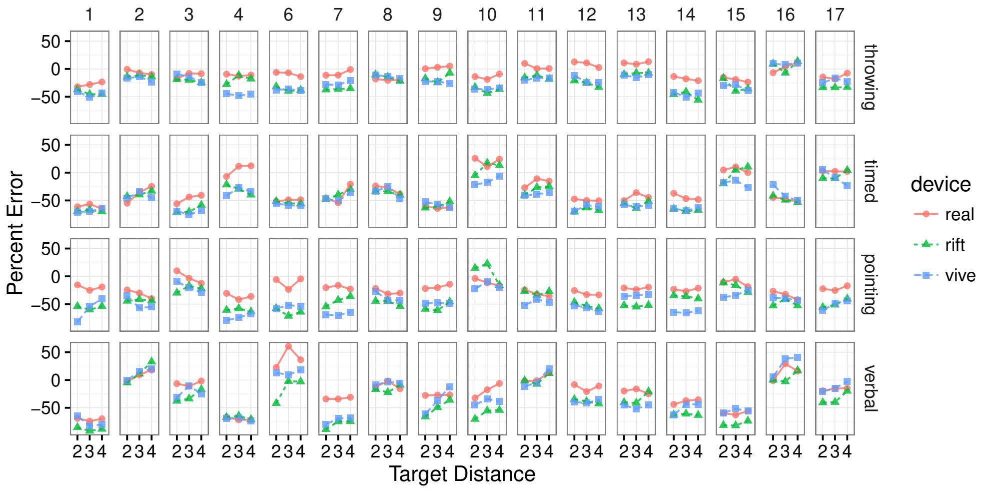

Figure 4.7: Mean perceived distance per participant, by measure and distance, colored by device.

tive percent error; using RPE, we see a mean underestimation of -17.42% ($SD = 18.58$), and, as seen in Table 4.1, this has drastic effects. Mean underestimation across task and device in virtual conditions is reduced significantly, from a range of 19-47%, to 9-27%; underestimation measured by timed imagined walking is reduced by nearly a factor of 4, transforming it from the measure exhibiting the most distance misperception using unadjusted error ($M = -39.29, SD = 25.53$), to the measure exhibiting the least ($M = -11.53, SD = 16.41$). This suggests that establishing a real environment baseline may be important to interpreting distance misperception results, particularly when comparing results across studies.

Distance estimates for all measures are more similar using RPE (compare Figures 4.5 and 4.4), but still significantly different ($F(3, 48) = 3.96, p = 0.013$). This suggests that reporting RPE as described here may better facilitate comparing error in perceived distance across studies using the same measure, but that there are some additional, unaccounted for sources of variation to consider when comparing between measures. As RPE seems the more consistent measure of an individual's experienced misperception across measures, the rest of our analysis considers only RPE and VE conditions.

The interaction effect between measure and distance ($F(6, 96) = 2.86, p = 0.013$) suggests that all measures but timed behave differently across the range of distances we explore. Throwing seems to exhibit more underestimation at 4m than 2m ($p = 0.008$), pointing exhibits more underestimation at the nearest distance (2m) ($p = 0.002$), and verbal shows more underestimation at the outer edge (4m) (2m:$p = 0.008$, 3m:$p = 0.011$). This may be due to shifts in technique while performing measures (eg. changes in the physical demands of throwing to 2m versus 4m), differing mental models of the space influenced by cues presented by the real and virtual environments unique to this experiment, or per-measure effects that occur at the outer edges of a learned range. Further study using different environments and distances may help elucidate. Mean difference caused by these effects is relatively small, ranging from 4-6%.

Interaction effects between HMDs and measure ($F(3, 48) = 6.53, p = 0.001$) are harder to interpret. The published specifications of both the Oculus Rift and HTC Vive are all but identical, yet the timed estimates performed using the Rift show less underestimation than those of all other measurement tasks by 7-12% ($p < 0.001$), while when using the Vive pointing shows more underestimation than all other measures by 9-17% ($p < 0.001$), and throwing differs from all measures but pointing by the more modest 3-5% (verbal: $p < 0.001$, throwing: $p = 0.002$). This may be due to the published difference in weight of 85g, though previous work suggests that HMD weight should not have an effect (Willemsen et al., 2009); it be may due to unpublished specifications, such as the accommodative distance of the HMDs' lenses, as accommodative distance has been shown to influence distance perception in CAVEs (Bruder et al., 2016); it may be due to unknown deviations of our specific HMDs from manufacturer specifications. Issues in fit may also be an influence, as participants repeatedly put on and took off both HMDs, with only subjective confirmation of comfort and clarity of vision; one headset may have trended to some misalignment of participant's eyes to HMD lenses, or some gap between the HMD and face facilitating peripheral stimulus, as in Jones et al. (2013). Some more rigorous measure of correct fit would help eliminate these possibilities.

One possible source of error, and a limitation inherent to our within-subjects design, are order effects. Previous work has shown order of presentation between real and virtual environments to influence distance estimations, in the specific contexts of single changes between environments, when VEs that are a photorealistic match to the RE are used (Witmer and Sadowski, 1998; Ziemer et al., 2009; Interrante et al., 2008; Steinicke et al., 2009b). Other learning or fatigue effects may also have been present. A simple linear regression on relative percent error over all VE trials yields a slope of 0.04, which would lead to a reduction of underestimation of 4.3% over a participant's 108 trials; RE trials yield a slope of 0.08, suggesting an 8.64% improvement. This suggests a slight overall learning effect twice as strong for RE as for VE, but no convergence between estimates in the two environments as might be suggested by environment presentation order effects (Witmer and Sadowski, 1998; Ziemer et al., 2009), and no strong overall bias suggesting interference from other order effects. This, along with our randomized presentation order and non-photorealistic VE, suggests that our results are not the product of an order effect. Further experiments using a counterbalanced or between-subjects design would further eliminate the possibility.

4.5 Chapter Conclusions

This chapter evaluates four different measures of determining distance on two consumer grade HMD devices, within the tracked space provided by consumer-grade hardware. As a baseline for future research, it provides evidence that distance misperceptions can be expected to occur when using contemporary head-mounted displays and visual stimuli. Results of the user study provide evidence that individuals experience different degrees of distance misperception, and that distance estimates made using different measures cannot be confidently compared. However, using a method that defines distance misperception as the difference between an individual's estimates when viewing real and virtual stimuli, distance estimates made using different measurement methods are much more similar.

5 PERCEPTUAL SPACE WARP

Many influences of distance misperception have been identified, but all of those that we can easily manipulate are thought to be correct already. Changing aspects of VR presentation in a way that violates assumptions of how to construct a valid immersive stimulus would be expected to introduce visual distortions or other perceptual artifacts – to produce a perceptually "wrong" stimulus. The possibility that these assumptions might be wrong is explored elsewhere in the literature; this chapter introduces *perceptual space warp* - or simply *warp* - a method of manipulating perceived distance that avoids manipulating these known-good aspects of rendering by changing the misperceived objects' positions directly. A user study provides evidence that individuals experience different degrees of distance misperception, and suggests that perceptual space warp might benefit by being tuned to mitigate an individual's misperception. Further, it provides evidence that distortions introduced by these warps are not a disruption to the immersive experience. [1]

5.1 Overview

While many researchers have noted the presence of distance misperception, few methods have been described to mitigate the effect.

One method is to simply let viewers get used to their misperceptions - that is, to provide participants feedback that allows them to adapt. A variety of feedback methods have been explored (Richardson and Waller, 2005, 2007; Waller et al., 1998; Kuhl et al., 2009; Kelly et al., 2013, 2014; Altenhoff et al., 2012). Perhaps most simply, Waller and Richardson (2008) suggests that walking through a virtual space for 45m of fully-sighted undirected travel effectively eliminates misperceptions in subsequent blind walking

---

[1] This chapter is adapted from Peer and Ponto (2016a)

egocentric distance estimation trials. However, these sorts of adaptations to one perceptual task have been shown to transfer to other perceptual tasks in undesirable ways: receiving feedback for virtual estimates may cause overestimation of real estimates (Waller and Richardson, 2008), and feedback correcting underestimation in egocentric estimates showed an equivalent overestimation in subsequent exocentric judgements (Richardson and Waller, 2005). There may also be limits to the sorts of movement and visual feedback that induce adaptations: Kuhl et al. (2009) asks participants to freely observe a virtual scene with head rotation but no translation or other body movement for five minutes, and no adaptation is seen under normal viewing conditions, or under a variety of viewing conditions expected to cause geometric distortions.

Virtual environments that are a photorealistic match to the physical environment participants are situated in have also been shown to elicit reduced distance misperceptions (Interrante et al., 2006; Phillips et al., 2009; Interrante et al., 2008), even in subsequently viewed unmatched environments (Steinicke et al., 2009b).

Other methods have attempted to manipulate or calibrate parameters influencing the presentation of the virtual scene. Changing geometric field of view has been shown to influence perceived distance, effectively magnifing or minifying the virtual scene (Kuhl et al., 2006). Methods have been proposed that depend on participants' ability to establish some correspondence between real and virtual stimuli, to calibrate parameters such as geometric field of view (Kuhl et al., 2009; Steinicke et al., 2009a), the full view frustum (Kellner et al., 2012a), or virtual eye depth (Ponto et al., 2013). Kellner et al. (2012a) evaluate perceived distance under their method using a blind walking task, and find that their calibrations do not fully eliminate distance misperceptions; the authors suggest that incorrect view frustum is not the main cause of distance misperception. Further, these mixed reality solutions depend either on display hardware that affords viewing real and virtual stimuli simultaneously, or on the viewer's ability to maintain their spatial sense of both virtual and real stimuli while switching between them, possibly while repeatedly donning and removing an enitre head mounted display. These approaches are not always practical. They also begin from the assumption that the otherwise best known configuration for these parameters, from manufacturer provided information or factory defaults, are flawed. It is hard to determine if calibrations made using psychophysical methods are more accurate than those made at the factory, though there is some evidence that they will be different: Steinicke et al. (2009a) find that a perceptual matching task overall leads participants to select larger a geometric field of view for the HMD they use than manufacturer specifications would suggest.

This chapter presents *Perceptual Space Warping* (PSW), which aims to mitigate distance misperception by directly altering the spatial information presented by the virtual environment, rather than aspects of rendering that we otherwise think to be correct. This is accomplished in a vertex shader, by (1) finding the direction the camera is facing (*look axis*), (2) finding the distance from the camera to the vertex along the look axis, and (3) applying a *warp multiplier* to the distance to move the vertex in the direction of the look axis. In this way, with a warp multiplier of 1.5, a point that would have been at 1m would be warped to 1.5m, a point at 2m would be warped to 3m, and so on. More simply: if we expect objects will appear to be too close, PSW attempts to move them further away such that too close is now at the intended position.

Employing PSW results in a similar final image to that of altering the geometric field of view as in Kuhl et al. (2009), though PSW provides several potential advantages. PSW does not make global changes to scene rendering properties, but rather depends on a vertex shader that can be applied globally or selectively to the geometry most affected by distance misperceptions, perhaps leaving text or other user interface unmodified.

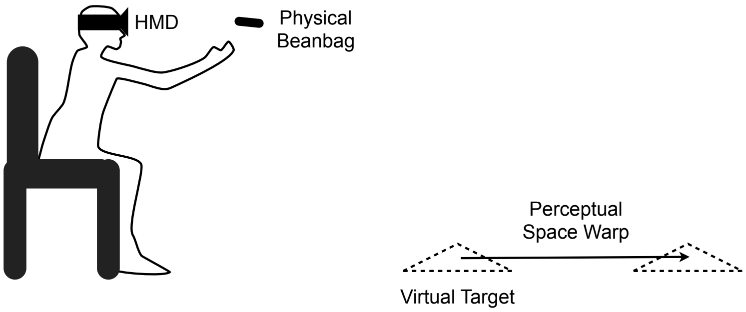

Figure 5.1: In the experiments described in this chapter, Perceptual Space Warp is used so that distances are extended from the user's perspective. The experiments measure the effects of this warping using a seated blind throwing task.

PSW also provides a fairly direct mapping from the error in perceived distance observed to a correcting warp multiplier. Although it is not explored here, PSW could also easily be adapted to piecewise or non-linear warps.

This chapter presents a series of user studies that aim to understand the effects of Perceptual Space Warping. To study the a participant's understanding of space, we use a blind throwing task based on Sahm et al. (2005). It explores a number of methods for determining how participants perceive the warped space. The procedures and results of these experiments are described and discussed in the sections below.

5.2 Method

Perceptual Space Warping is simple in both concept and implementation. The concept is to mitigate distance misperception by warping the virtual environment by the inverse of the misperception. This process is implemented using a vertex shader that transforms vertices by increasing their distance from the viewer by some *warp multiplier*. A multiplier of 1 makes no changes, < 1 moves things towards the viewer, and > 1 moves things further away. Formally, we define this transformation as:

$$V_{\mathrm{out}} = M^{-1} \times ((M \times V_{\mathrm{in}}) \times (1, 1, w_{\mathrm{m}}, 1)) \tag{5.1}$$

where V_{in} and V_{out} are the untransformed and transformed vertices, respectively, M is the modelview matrix, and w_{m} is the warp multiplier. This shader was implemented in the Unity game engine[2] and was applied to all objects in the scene. The Oculus Rift DK2 was used to display the virtual environments to the participants and the Oculus V0.6.0.0-beta SDK was used to integrate the virtual environment with the HMD.

To evaluate a participant's perception of distance, a blind throwing task was used that was similar to that described by Sahm et al. (2005). Rather than the more common measure of blind walking, blind throwing was selected because it could be performed while remaining comfortably within the Oculus Rift DK2's tracking volume. Throwing also allowed for more trials to be run more quickly.

The presented set of experiments differs slightly from Sahm et al. (2005) in that we have participants throw while seated, as shown in Figure 5.1. We measure distance from a point beneath the participant's chair to the initial landing point of the beanbag, ignoring any bounce. We discard any angular error; that is, we measure distance on the axis between the participant and the target, discarding travel on the perpendicular axis. Participants were instructed to throw such that the beanbags' center lands at the center of the target and told that any extra travel from bouncing or rolling would be discarded. All participants provided consent before beginning the experiment and were given $5 at the conclusion of their participation.

---

[2] https://unity3d.com/

5.3 Experiment 1

Experiment 1 explores whether PSW has an effect on distance estimations in VEs. Based on results from a earlier pilot experiment (n=5), a warp multiplier of 1.4 was selected. A total of 12 participants took part in this experiment. Participants were recruited during an outreach event held in concert between our lab and a local non-profit. The experiment was designed to investigate three questions:

Q1 Will participants will be able to accurately throw bean bags to a target in a seated position while viewing the real environment?

Q2 Will PSW have a significant and positive effect in the mitigation of distance misperception?

Q3 Will a cue-rich virtual reality environment produce better distance estimates than a cue-poor environment?

Design

We examine three factors of interest: (*warp*) presence or absence of our Perceptual Space Warp adjustment; (*environment*) the environment in which the target is presented; (*distance*) distance of target from the participant.

In VE conditions warp and distance are within-subjects variables, and all combinations of levels of the two are presented in random order; environment is a between-subjects variable with participants evenly distributed between levels. Two levels of warp are present (*no warp adustment*, *1.4 warp multiplier*). Eleven levels of distance (0.1m intervals between 3m and 4m, inclusive). Two levels of environment are present: a depth-cue-sparse hallway made of untextured walls (*sparse*), and a depth cue rich mirrored room with textures and furniture (*rich*), as shown in Figure 5.2.

In the real world condition, three levels of distance (3m, 3.5m, and 4m) are presented. Participants perform a total of 11 throws: 4 each at

Figure 5.2: The two environments used in Experiment 1. The left image shows a depth-cue-sparse hallway made of untextured walls while the right image shows a depth cue rich room with textures and furniture.

3m and 4m, and 3 at 3.5m. The order in which distances are presented is randomized, but all throws for each distance are done together in sequence before moving on to the next distance.

Procedure

Participants moved through three phases in the experiment: (1) practice throwing, (2) real environment (RE) throwing, and (3) virtual environment (VE) throwing. All participants proceeded through the phases in the same order.

During the practice phase, participants sat in a chair and were instructed to throw at a target made of tape in the shape of an "X" on the floor in front of them. Participants were instructed to throw until they felt comfortable with the task. The target was placed within the 3m-4m range used during experimental trials, but it did not match any of the three intervals used during real environment throwing. Participants elected for

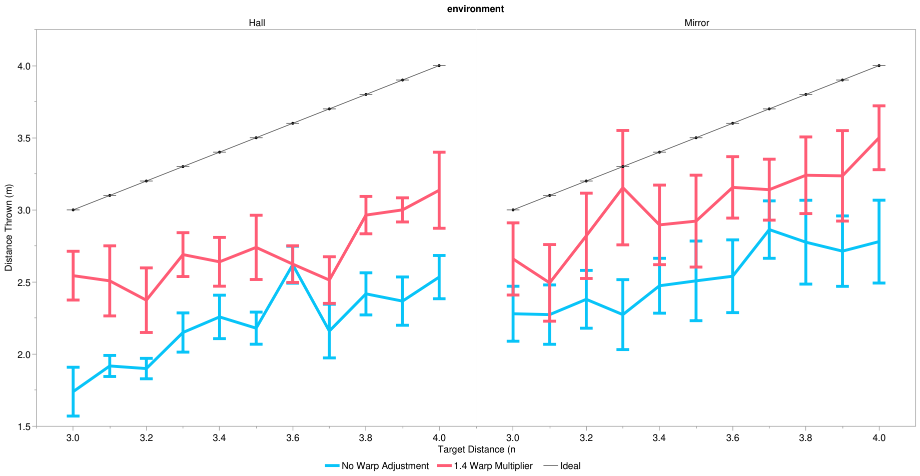

Figure 5.3: Distance misperception seen with presence and absence of warp, across both virtual environments. Error bars represent 1 standard error of the mean.

brief practice phases. Only one participant threw more than 8 beanbags, and most opted for 3 or 4.

In the real environment throwing phases, participants were led to another chair and presented with the new target of a cardboard triangle. The target was placed at the appropriate distance, and the participant was asked to throw at the target. Between throws, the landing point of the beanbag was measured; after 3 or 4 throws, the target was moved.

In the virtual environment phase, participants remained in the same chair and were asked to don the HMD and adjust it for comfort and visibility. After confirming that they could see the target clearly, participants were asked to throw. The landing point of the beanbag was then measured, and the next scene presented to the participant. Between scenes, the participant was shown a black screen for one second. Participants were instructed that after a black flash they were free to throw again once they felt they had a feel for the location of the target.

Results

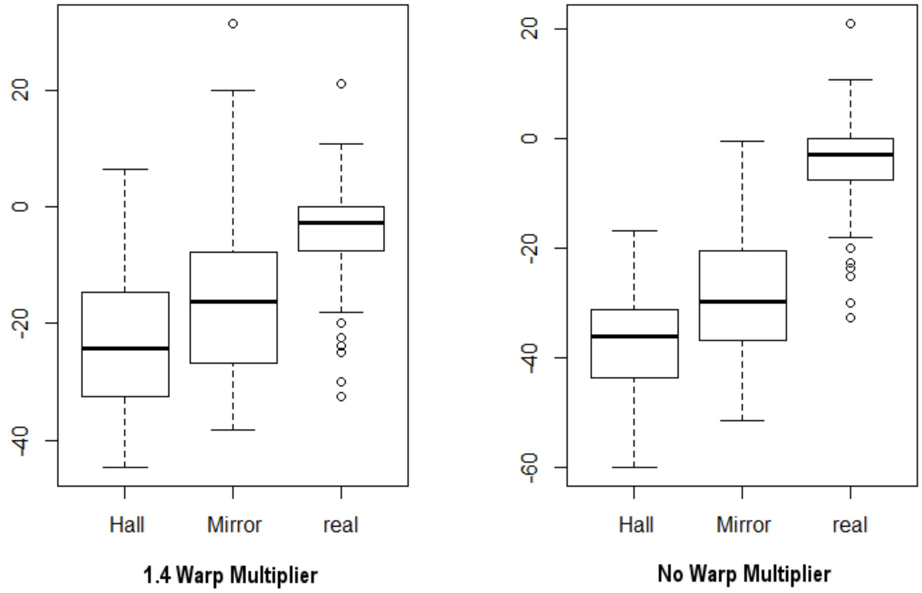

Figure 5.4: Throwing results from Experiment 1.

With regards to **Q1**, our results show that participants throw accurately when viewing a real stimulus, with mean percent error of -3.6%, and standard deviation of 8.1%. A 2(warp)x2(environment) ANOVA showed there was a significant increase in throw accuracy in the warp adjusted case, $F(1, 216) = 63.6, p < 0.001$. However, participants significantly underthrew in both virtual conditions, $F(1,213)=113.7$, $p < 0.001$, which does not suggest a strong positive effect of PSW (**Q2**). The effect of environment was not significant and there was no interaction between environment and warp, suggesting no strong effect between the two environments used (**Q3**). There was no significant effect of distance on percent error, F(10,209)=0.23, $p < 0.99$.

The amount of distance underestimation seen in conditions with no warp correction is very similar to results seen elsewhere in the literature (Renner et al., 2013). While pre-tests suggested a warp of 1.4 should correct for distance misperception on average, the underestimation seen in this first experiment is not fully corrected at that multiplier; while this experiment shows PSW to have an effect, the effect is much smaller than anticipated.

5.4 Secondary Experiments

The results of Experiment 1 motivated a second set of experiments meant to better understand the discrepancies between the predicted outcomes of PSW and the observed results. We explored whether distance estimations under warp might be further influenced by: view dependent artifacts, per user calibrations, incorrect warp multipliers, or the physical aspects of wearing an HMD. These experiments shared the same participants, and several common procedures; these commonalities are further described below.

Participants

A total of 9 participants (5 male, 4 female) took part in these experiments. Participants were recruited from the local university campus. Average participant age was 24, with a range from 19-35. All participants reported a high level of familiarity with computers in general; 3 participants had experienced an HMD before, one of whom had actively developed VEs.

General Procedure

Participants were brought into a controlled room, where they were provided with a brief description of the experiments and asked to sign a consent form. They were then seated and asked to practice throwing beanbags at a target placed at a distance of 3.5m. They were instructed to practice until they felt that they had mastered the task, and practice continued until they informed the experimenter they were done. No practice sessions lasted longer than 2 minutes.

The full battery of experiments took roughly 40 minutes. As pilot experiments indicated that longer VE viewing sessions may sometimes cause a fatigue effect, short breaks were built into the experiment during which participants removed the HMD and performed other tasks, including filling out brief surveys or being interviewed by the experimenter. All participants participated in all four areas of inquiry.

Throwing Protocol

In all experiments, the following general throwing protocol was followed: Participants were asked to put on our HMD and adjust it to provide comfort, a clear view, and minimal shifting during head movement. During this process, they viewed a test scene and confirmed that the virtual throwing target was clearly visible. They were then informed that they would view a series of virtual scenes containing similar virtual targets. Each scene would be displayed for a short time, then be replaced by a blank black screen. While the scene was visible, they were to fix the target's position in their mind. When the screen went blank, they were to throw a beanbag toward the target. Each scene was displayed for 5 seconds, and the next scene was displayed after the experimenter had finished measuring the last throw. The sparse hall from Experiment 1 was used as the test scene, and the rich room was used for all throwing scenes.

5.5 Experiment 2: View Dependent Artifacts

One possible reason for the discrepancies seen in Experiment 1 may come from *swim*, a distortion defined as "occurring whenever a static object appears to change position as the participant's viewpoint changes" (Ponto et al., 2013). PSW is dependant on view direction, so a swim effect might be expected with head movement. Therefore, the question of interest for this experiment:

Q4 Will eliminating the view dependent aspect of warp improve blind throwing accuracy?

To test this, we created a second vertex shader that performs the same task as our original PSW shader, but was not sensitive to changes in user head pose; that is, our viewpoint-independent shader performs a warp valid for a single fixed head rotation. This eliminates the swim distortion for the controlled scenes in the current experiment where user attention is directed in a single direction, but not in the general case.

Design

We investigate one factor at two levels: presence of swim distortion (swim) and absence of swim distortion (no swim). Participants threw beanbags at 11 distances in both conditions, for a total of 22 throws. Distances ranged from 3m to 4m at 0.1m intervals, and each distance was presented once per condition. Each participant saw all combinations of distance and the two levels of swim and no swim in random order.

Results

No effect was found between users throwing under no swim (M=-29.3, SD=15.8) and swim (M=-28.5, SD=18.2) conditions (t(192)=-0.33, p =0.7)

(Figure 5.5). This suggests that swim distortion does not contribute to distance underestimation within our experiments, or that the view direction dependency of PSW is not what produces the distortion. This result is discussed in greater detail in Section 5.9.

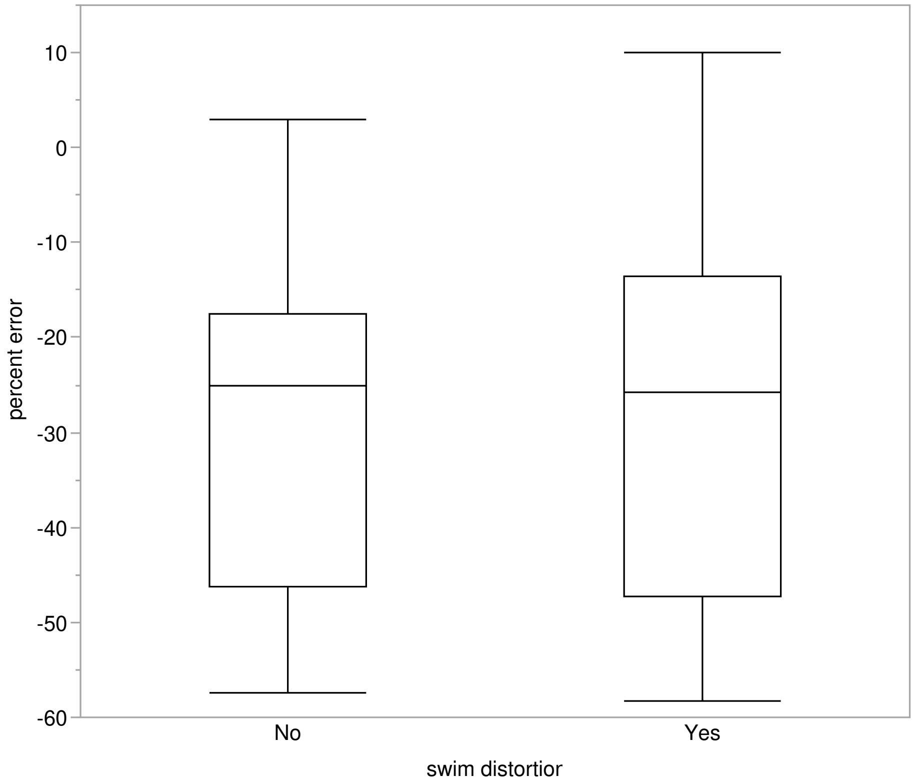

Figure 5.5: Swim effects caused by view dependent PSW are not seen to influence distance misperception.

5.6 Experiment 3: Per-User Calibration to the Real World

It's possible that no single warp multiplier will solve distance misperception for all participants, and that we may need a per-user calibration. This is suggested by the general increase in standard error we saw in warp-corrected conditions in Experiment 1 (as can be seen in Figure 5.3); it has also been suggested elsewhere in the literature (Ponto et al., 2013). As Kuhl et al. (2009) saw success calibrating geometric field of view by matching virtual objects to real environment counterparts, we explore the feasibility of adapting this technique to establish a user-calibrated warp multiplier to correct distance misperception on a per-user basis. In this regard, we have two questions for this experiment:

Q5 Can users accurately perform the calibration task of matching a virtual and real world target?

Q6 Will user-calibrated warp multipliers improve blind throwing accuracy?

Design

Two within-subject conditions were run: a user-calibrated warp multiplier (calibrated) and no warp multiplier (no adjustment). During calibration testing, participants threw beanbags at virtual targets displayed using either a warp multiplier that was an average of the multipliers selected during their four calibration trials, or no warp multiplier.

Procedure

This experiment began with a calibration phase. In this phase, participants were instructed to hold the HMD up to their eyes without fully fixing it to their head, such that they could quickly switch between viewing the real world and virtual scene. Targets were placed 4m from the participant in both the real and virtual environments, and they were given a joystick that adjusted the warp multiplier. Participants were asked to quickly switch between viewing the real and virtual targets and to use the joystick to move the virtual target such that the two targets were at the same distance with both aligned in the center of their vision. When the virtual and real targets matched, they pushed a button to confirm the alignment. Participants performed this task 4 times, starting from a 0.5 or 1.0 multiplier 2 times each. Starting multipliers were presented in a random order. An average of the four selected warp multipliers was then established as the user-calibrated warp multiplier. The rich room scene was used for this task.

Once a calibrated warp multiplier was established, participants fully donned the HMD and followed the throwing protocol as described earlier. While calibrating, scenes were presented using the view independent warp shader; while throwing, the view dependent shader was presented. Participants threw at targets positioned at each of the 11 0.1m intervals between 3m and 4m twice, for a total of 22 throws.

Results

No effect was found between calibrated (M=-28.5, SD=18.2) and no adjustment (M=-28.0, SD=19.2) conditions (t=-0.11, p = 0.9). This may suggest that per-user calibration is not an effective way to correct for distance misperception using PSW (**Q6**). However, as seen in Figure 5.6, most users calibrated to a warp multiplier near or below 1, suggesting participants did not select multipliers that would mitigate the underestimation the participants demonstrate in other experiments. This implies that this calibration method is not well suited for most users (**Q5**).

5.7 Experiment 4: Per-User Warp Space Exploration

This experiment aimed to sample a larger warp multiplier space in order to better describe the effect of PSW on participants' distance estimations. The question that is the focus of this experiment:

Q7 Is the optimal warp multiplier different for individual participants?

Design

Each participant threw over 8 distinct warp multipliers, ranging from 0.6 to 2.0, at intervals of 0.2. They threw at each interval 6 times for a total of 48 throws per participant. For each throw, the virtual target was placed at a random 0.1m interval between 3m and 4m.

Procedure

Participants followed the throwing protocol as described earlier, with the slight change of a 2 minute break halfway through, during which participants removed the HMD and were interviewed by the experimenter on their previous experience with virtual reality and computers in general.

Results

As shown in Figure 5.7, warp multipliers determined by the zero percent error crossing were varied among the participants, suggesting that **Q7** be answered in the affirmative. The mean warp multiplier was 1.65 (SD 1.56). It should be noted that three of the participant's warp multipliers are projected to lie outside of the range over which we sampled, making them particularly speculative. While the calibration procedure in Experiment

3 was unsuccessful, the results of this experiment indicate personalized calibration may be viable.

5.8 Experiment 5: Wearing an HMD

This experiment tests whether wearing an HMD contributes to participants' distance misperception. While Experiment 1 has shown that participants in a seated position can throw to real environment targets with a high level of accuracy, there may be influences from simply wearing the HMD, as seen by Willemsen et al. (2004, 2009). The quesiton of this experiment:

Q8 Will wearing the HMD but viewing real targets cause participants to underthrow?

Design

We investigate a single factor at two levels: throwing in the real environment, with and without wearing the HMD. In all conditions, the target was placed at a distance of 3.5 meters from the participant.

Procedure

Participants were first asked to wear our HMD as a blindfold while throwing at a real world target. They were instructed to lift the HMD to view the target, to fix its position in their mind, and then lower the HMD and throw a beanbag towards the target. During this task, the HMD showed a black screen. Participants repeated this task five times. Participants were then asked to remove the HMD entirely and throw five beanbags at the target, with no blindfold.

Results

There was a significant difference in the percent error for HMD-as-blindfold (M=-12.7, SD=11.1) and no blindfold (M=-3.8, SD=8.1) conditions (t(80)=4.33, $p < 0.001$). This suggests that there may be under-throwing happening for some participants due to the physical aspects of wearing an HMD, or to throwing while blindfolded (**Q8**). This result is further discussed in Section 5.9 below.

5.9 Discussion

The results of these experiments demonstrate that while warp had an effect on distance misperception, none of the warp intervals we tested resulted in a perfect match to real world distance estimation for all participants. It appears that corrections may need to be personalized, as shown in Experiment 4. In post experiment interviews, the PSW manipulation was not noticed as distinct from changes in target distance, except for one participant who had a sense of being moved forward or backward in the scene. No one reported discomfort or disorientation as an effect of the manipulation. Three participants noticed distortions that may have been swim, but did not feel they made the task more difficult.

While this chapter's aim was not to understand distance misperception but rather to create an intervention method to overcome it, our results may still help illuminate contributing factors.

Distance misperception has been shown to be mitigated when virtual environments are more similar to familiar real world environments Interrante et al. (2006); Phillips et al. (2009); however, as shown in Interrante et al. (2008), stretching known environments can nullify these corrections. While the two environments used in Experiment 1 were not shown to be a significant factor for perceived distance, it is possible that distortions like swim, or simply seeing the same environment under several warp multipliers, may alter a participant's sense of presence, thus affecting distance misperception. Additionally, a significant effect was found between stereo acuity (measured as whether participants could discern the shapes in a single panel of a TNO test) and distance misperception in the data from Experiment 4, F(1,192) = 31.3, $p < 0.001$: 3 participants who saw only indistinct shapes tended to throw more accurately. This tiny sample of only three makes us wary of making any strong claims, though Armbrüster et al. (2008) did also find evidence that stronger stereo acuity corresponds with greater distance misperception. Future work might further explore the relationships of presence and stereo acuity to distance misperception.

Experiment 5 found evidence that simply wearing the HMD caused under-throwing. The distance misperception seen when using the HMD as a blindfold when throwing to real environment targets (M=-12.7,SD=11.1) was about half that seen when throwing to virtual environment targets (M=-28.0, SD=19.2), indicating that simply wearing an HMD is not the sole source of distance misperception. These findings parallel the results of Willemsen et al. (2009), though that work used a different perceived distance measure, blind walking.

The fact that no significant difference was found in Experiment 2 between view dependent corrections and view independent corrections is surprising. This provides evidence that any distortions PSW may introduce do not have negatively effect distance perception. In one sense this result is exciting, in that PSW may not introduce meaningful swim distortion. However, it also implies that swim distortions were not the cause of the fixed warp multiplier used in Experiment 1 inducing less of a change in perceived distance than was expected.

While the method described in Experiment 3 was based on the successful work of Kuhl et al. (2009), the fact that many users did not create warp multipliers greater than 1.0 – that is, multipliers that would counter underestimation – suggests that the calibration was not successful. It is unclear if this is because the real-to-virtual matching task is simply difficult to perform accurately, or because of interference from some other perceptual artifact. If the latter, this may imply that a perfectly calibrated scene would not match the retinal image of a real environment scene depicting the same distances – or, if the former, that we would not be able to recognize such a match. Either suggests some fundamental difference in the way virtual and real world environments are perceived. Future work might further explore the feasibility of calibrating to real environment objects, possibly using two-alternative forced choice tasks. Chapter 6 explores other methods of creating a personalized warp calibration.

5.10 Chapter Conclusions

This chapter presents the Perceptual Space Warping method in an attempt to correct for distance misperception in virtual environments. While the method significantly decreases underestimation, presenting the same warp multiplier to all participants does not fully alleviate misperception effects. Perceptual artifacts suspected to be introduced by PSW are not found to have a significant effect on distance misperception. Wearing a head mounted display and viewing real targets is seen to induce about half of the distance misperception seen when viewing virtual targets, which mirrors results seen elsewhere in the literature when using a different distance estimation method. As participants appeared to respond differently to different warp multipliers, personalized warps appear to be necessary. However, a calibration method depending on aligning virtual and physical objects was not found to improve distance misperceptions. The next chapter explores other methods of generating a personalized perceptual space warp.

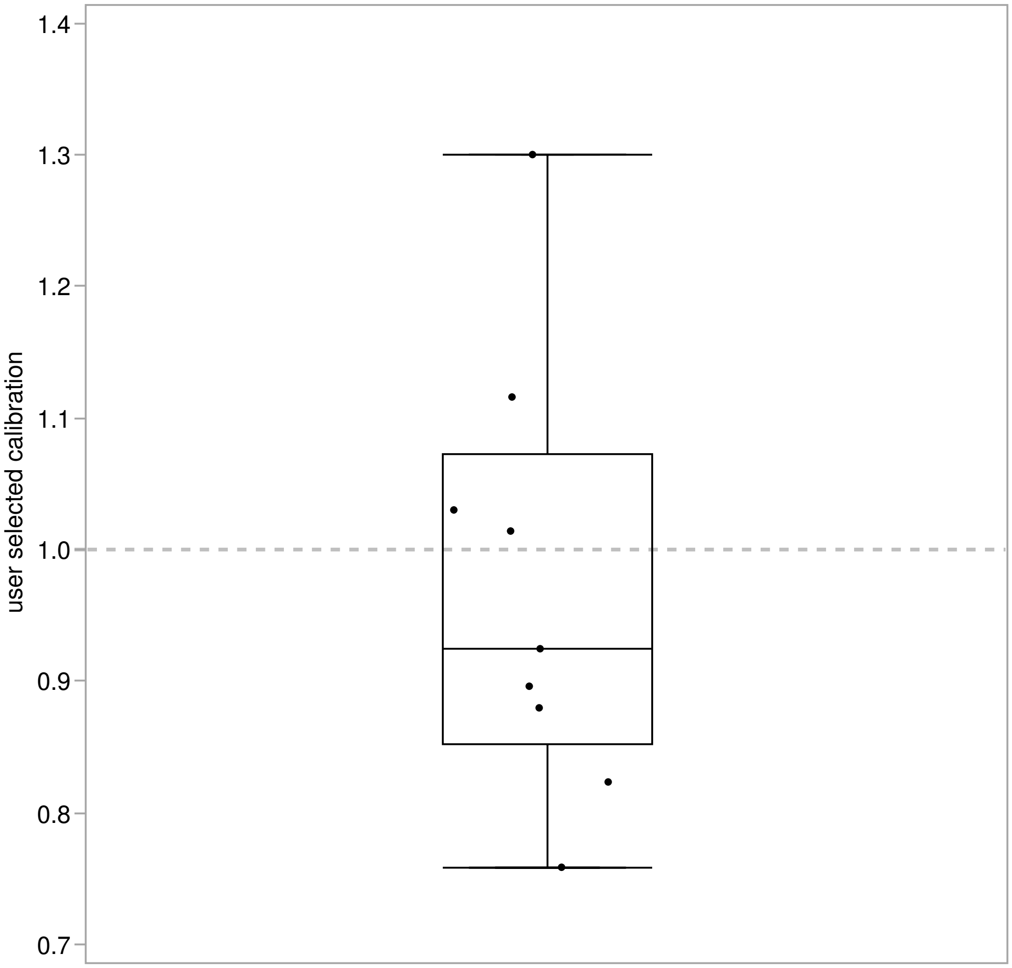

Figure 5.6: User selected warp multipliers. Note that a multiplier of 1 indicates no effect on the scene; >1 indicates pushing things further out in the scene, and <1 indicates pulling things closer. To correct for distance misperception, we expect >1.

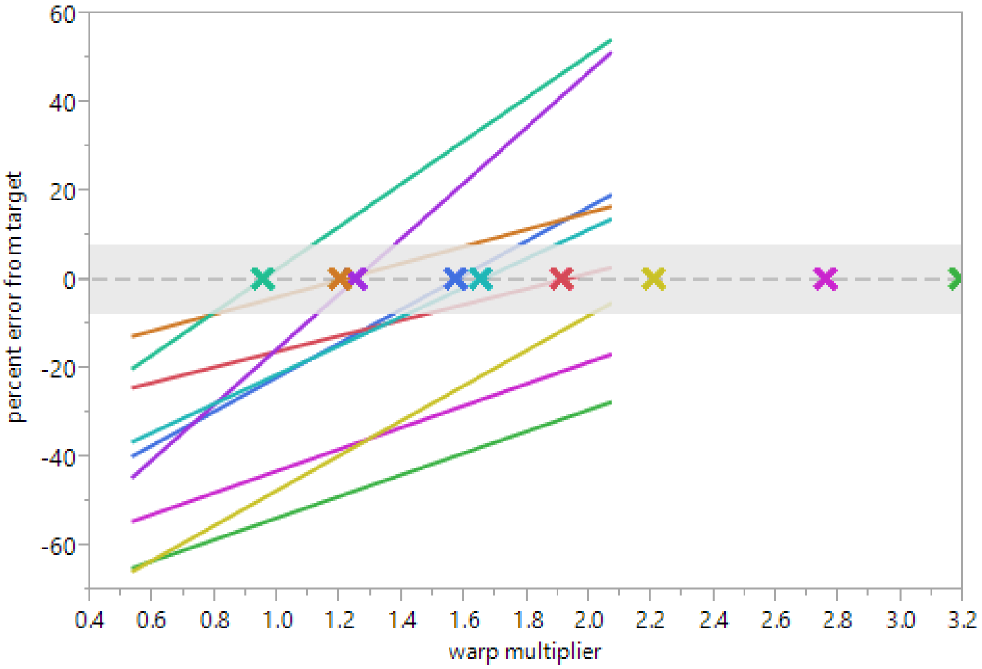

Figure 5.7: Linear fits of participants' throwing error over different w_m, with a different color used for each participant. Each X indicates the w_m expected to yield 0 error for that participant.

6 PERSONALIZED PERCEPTUAL SPACE WARP

Chapter 5 introduced a technique called *perceptual space warp*, and suggests that it could be calibrated to mitigate an individual's distance misperceptions. This study explored this possibility, providing evidence that one proposed calibration method can successfully mitigate distance misperception, with no detrimental effects on two other perceptual tasks. [1]

6.1 Overview

Immersive display technologies that drive virtual, augmented, and mixed reality systems have been shown to induce distance misperceptions relative to real environments (Renner et al., 2013). These misjudged distances have been cited as a concern for training applications in virtual reality (VR) (Wright, 1995), and have been shown to have some capability to transfer to behavior in real environments (Waller and Richardson, 2008; Witmer and Sadowski, 1998; Proffitt et al., 2003; Interrante et al., 2006; Plumert et al., 2005; Ziemer et al., 2009; Interrante et al., 2008). There is also some evidence that these misperceptions vary by viewer (Peer and Ponto, 2016a; Ponto et al., 2013).

One method to "fix" the discrepancies between the perceived and intended virtual environments is to stretch the world out away from the participant, such that perceived distances match intended. Peer and Ponto (2016a) attempted this, and found that while stretching the world by a fixed amount resulted in closer to real-world performance for a blind throwing task, it did not resolve distance msiperception to the degree expected. This work aims to further explore this result by changing the technique in two key ways. First, warps are calculated on an individual basis based on a blind throwing task that has been shown to accurately

---

[1] This chapter is adapted from Peer and Ponto (2019)

measure perceived distances (Peer and Ponto, 2017, 2016a; Sahm et al., 2005). Following the suggestions of Peer and Ponto (2017), an individual's distance misperception is taken as the difference in task performance when viewing a real environment (RE) and a virtual environment (VE), in a measure called Relative Percent Error (RPE). Secondly, two types of warps are explored: one with a linear multiplier based on distance ($warp_a$), and a second which also includes a static offset ($warp_{ab}$). Conceptually, this offset is similar to the PO_z attribute referenced by Ponto et al. (2013), in which the cameras representing the virtual eyes are placed at some distance inside of the participant's head.

It was noted by Peer and Ponto (2016a) that these warps may induce noticeable distortions. Though that work reported no noticeable adverse effects, it may be that these distortions aren't noticeable when performing blind throwing tasks, but would have an effect on other tasks. That is, "fixing" distance misperception using a single blind directed action task might "overfit" a manipulation to this single task, to the possible detriment of others. In this work, we observe the effect of the personalized warps on two perceptual matching tasks, similar to those used by Ponto et al. (2013).

This leads to the following questions:

Q1 Will a personalized warp reduce error in distances perceived when viewing virtual environments?

Q2 Will the extra component of the model for the $warp_{ab}$ condition improve performance over the $warp_a$ condition?

Q3 Will warps degrade performance in other perceptual tasks performed in the virtual environment?

6.2 Related Work

Renner et al. (2013) present a survey of the work on distance misperception in VR, finding an average reported underestimation of 26% across 30 papers. As most papers do not compare real and virtual task performance, this average may overestimate the actual misperception observed (Peer and Ponto, 2017).

Ponto et al. (2013) attempted to use perceptual matching tasks to calibrate eye position on two axes to the individual participant, and establish a relationship between their eye-pose parameters, their perceptual matching tasks, and perceived distance. One eye pose parameter discussed in Ponto et al. (2013) may function similarly to the idea of changing the center of the eye as modeled in VR rendering as discussed in Jones et al. (2016).

Minification of the presented image through manipulating geometric field-of-view (gFOV) has been shown to influence perceived distance (Li et al., 2015; Kuhl et al., 2009; Steinicke et al., 2011b). PSW differs in that it can be used without knowledge of device field-of-view (dFOV), and alongside calibrations correctly matching device parameters. PSW, being a simple multiplier, correlates more directly to a change in distance, which may make it easier to employ; it also allows for non-uniform multipliers, in the case of non-uniform misperception across distances, or deliberate manipulation of the VE (to focus attention, for artistic effect, etc).

Our attempt at calibration is similar in intent to Steinicke et al. (2011b), where participants were asked to match virtual and real stimulus by adjusting gFOV, and select gFOVs larger than the actual FOV of the HMD by up to 50%; our experiment differs in that it uses PSW rather than gFOV to manipulate the scene, presents a different virtual environment, and evaluates perceived distance once calibrated.

The presented work attempts to mitigate perceptual issues by warping the space around the user, as first proposed Peer and Ponto (2016a). In that work all users were assigned the same warp multiplier, meant to remove the effects of distance misperception for a throwing task. While this warp multiplier was shown to significantly reduce distance misperception, the effect was less pronounced than anticipated. The current work improves upon this previously presented method by:

1. Creating a per-person warp multiplier that attempts to match virtual world performance to real-world performance
2. Creating a warp that includes a multiplier as well as an offset

6.3 Method

This experiment attempts to quantify the distance misperception experienced by individuals in VR, and develop a customized set of parameters to drive a manipulation that elicits correct perceived distance. The effects of this attempted correction are then evaluated using three perceptual tasks: the same distance estimation task used to calibrate the manipulation, and two other perceptual matching tasks.

Warps

We manipulate the scene by applying a vertex shader that shifts the position of things in the scene relative to the user's position and gaze direction. This is in the spirit of the perceptual space warping introduced in Peer and Ponto (2016a); here, we'll simply call such manipulations *warps*. Two different warps were tested in this study.

$\boldsymbol{Warp_a}$ was determined by creating a linear regression based on the difference between the distances a participant threw to in real and virtual conditions. The regression was forced to pass through the origin, meaning that it was assumed that objects at very small distances from the participant would not show distance effects. This warp is implemented using a vertex shader that transforms vertices by increasing their distance from the viewer on the viewspace z-axis. Formally, we define this transformation as:

$$V_{out} = M^{-1} \times ((M \times V_{in}) \times (1, 1, w_a, 1)) \tag{6.1}$$

where V_{in} is an unwarped vertex position in 3-space, V_{out} is the warped resulting position, M is the modelview matrix, and w_a is the *warp multiplier*. By modifying w_a, one adjusts the magnitude and direction of the warp; a w_a of 1 causes no change, < 1 pulls objects closer, and > 1 pushes them further away.

Warp$_{ab}$: used a similar method as was used to determine *warp$_a$*, but added an additional component for world space offset. To accomplish this, the regression model was no longer forced to go through the origin, meaning objects at near distances could be offset via the warp. The warp is implemented using a vertex shader, which can formally be described as:

$$V_{out} = M^{-1} \times ((M \times V_{in}) \times (1, 1, w_a, 1) + (0, 0, w_b, 0)) \tag{6.2}$$

V_{in} is an unwarped vertex position in 3-space and V_{out} is the warped resulting position, M is the modelview matrix, and w_a is the *warp multiplier* and w_b is *distance offset*. As w_b is applied to all vertices regardless of their distance from the participant, one would hope that this value would be small; there is some question as to whether the position of a viewer's eyes is modeled correctly in current VR rendering (Jones et al., 2016; Ponto et al., 2013), and this might compensate for this sort of subtle misalignment.

The warps in this experiment are meant to match virtual to real-world performance. To this end, the regressions were fit to perceived distance in virtual environments, relative to mean perceived distance in real environments. To do this, perceived distance in the two environments was measured using a blind throwing task (see §6.3), yielding a perceived

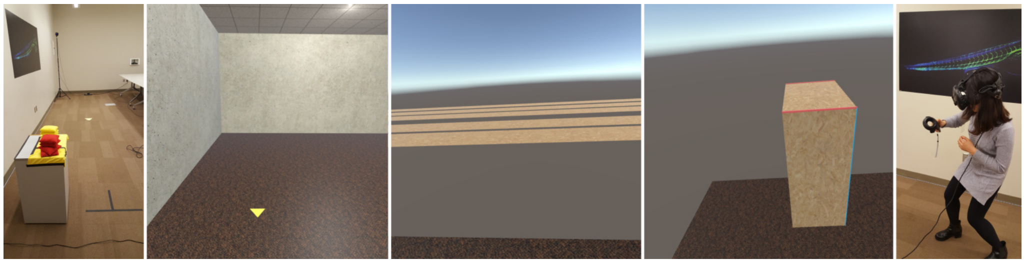

Figure 6.1: The environments and tasks used during the experiment. From left to right: The real environment for throwing; the virtual throwing environment; the planks task; the cube task; a collaborator demonstrating the cube task.

distance at given target distances, where target distances are randomly modified by ± 0.5m from the three base target distances (2m, 3m, 4m), as described in §6.3. The real environment measurements were then binned at three intervals – the three base target distances – and the mean error in perceived distance for each RE base target distance was calculated. Finally, VE measurements were adjusted by the mean error of their matching RE base target distance; regressions were fit to this adjusted data to yield warp parameters with the intention of shifting perceived distance in the virtual environment to match that of the real environment.

Tasks

The experiment asked participants to perform three tasks: blind throwing, plank leveling (*planks*), and block squaring (*cube*). Blind throwing was used to measure the degree of error in perceived distance to calibrate the warps; all three were used to evaluate the effects of the warp manipulations.

Blind Throwing

We follow the protocol for blind throwing as established in works by Sahm et al. (2005) and Peer and Ponto (2016a, 2017); in Peer and Ponto (2017) in particular, it was shown to be a directed action task that did show distance misperception, and showed real world performance nearest to zero error. When performing this task, participants begin with their eyes closed. A sound then prompts them to open their eyes and view a triangular target placed at some distance between 1.5m and 4.5m, in either the real or virtual environment. After three seconds, participants are prompted to close their eyes, and throw a beanbag towards the target. The experimenter then registered the landing position of the target, removed the beanbag, and the next target was prepared – manually placed in RE conditions, using an LED light strip positioned to one side of the throwing space to indicate both the next target position and the position of a controller held near the target, such that matching the two lights indicated an aligned target.

The primary error measure used for blind throwing is relative percent error (*RPE*) as described in Peer and Ponto (2017). First, perceived distance is taken as the distance of a beanbag's landing point, projected onto the axis in the ground plane running forward in the direction the participant has been asked to face towards – the axis on which targets are placed. Error in perceived distance is taken to be the difference between perceived and target distance, such that error is negative when perceived distance is less than target distance, and percent error is this value divided by the target distance; for RPE, we adjust the target distance by the mean error seen in real environment conditions. For this experiment, this RE mean error is binned by the three base target distances described in §6.3; VE measures are adjusted by the RE mean error belonging to the base target distance nearest to the VE target distance.

Planks

The planks task mirrors the perceptual matching task used for distance estimation in Ponto et al. (2013), in which perceived slope is used as a proxy measure for distance perception (Ooi et al., 2001; Proffitt et al., 1995). This task is intended to isolate stereo vision as a depth cue. In this style of task, participants are asked to level a tilted plane.

As in Ponto et al. (2013), 5 planks with a depth of 0.52m were placed at one meter intervals, with a random perturbation of $\pm 0.25m$. Planks are positioned 0.72m above the floor. Our implementation differs in that the planks have no height (0.001m) and infinite width (10000m), in hopes of further isolating the desired depth cue.

The planks are "tilted" by some angle, such that an angle of 0 aligns them along a flat plane, a negative angle orients them as a staircase leading down, and a positive angle as a staircase leading up. The staircase effect is due to forcing the planks to remain parallel to the desired flat plane, as Ponto et al. (2013) found that aligning the planks to the tilted plane indicated by the selected angle induced unintended depth cues through motion. In this work, the intended z position of individual planks is preserved by choosing their y position by casting a ray from their centers in the y direction, and positioning them at the intersection of this ray and the tilted plane.

When executing this measure, participants stood in place and held a tracked controller. When holding the trigger, moving the controller up or down would work as a sort of open-air joystick, causing the planks' orientation to change by some constant increment determined by the distance of the controller on the y-axis to its position when the trigger started being held down. This was described to participants as holding the trigger and "pulling" the planks up or down. The intent is to not provide too direct a mapping between the planks' movement and the participant's proprioceptive sense of distance.

At the beginning of each planks trial, the planks were tilted by a random angle between -5 and 5 degrees. The error measure for the planks task is the angle selected; any deviation from 0 is error. For the sake of observing whether a warp makes a task more difficult, we are concerned less with the directionality of the error as with its magnitude, so we also investigate absolute error, measured as the absolute value of degrees from 0.

Cube

This task mirrors the perceptual matching task used to judge perception of shape by Ponto et al. (2013). Participants are presented with a floating box with each dimension randomly sized, between 0.15m and 0.6m. One dimension is held fixed, and participants are asked to manipulate the other two such that all sides are the same size; that is, they are asked to make the box into a cube. Ponto et al. (2013) found that this task can be affected by manipulating the way a virtual environment is presented.

In our implementation, the box floated 1m above the ground. The z dimension was held fixed. Colored guides were placed on the two top edges running along the x dimension (red), and two opposing edges running along the y edges (blue), such that one edge of each color should always be in view as participants moved through the space. A similar open-air joystick control scheme as used for the planks task was used here, with movement along the box's x and y axes causing the cube's respective dimension to grow or shrink; participants were told to "pull" in the red or blue direction to make the cube grow or shrink in that direction.

The error metric for the cube task is as in Ponto et al. (2013): an L2 norm, calculated thusly:

$$\mathrm{Error} = \sqrt{(l_x - l_z)^2 + (l_y - l_z)^2} \tag{6.3}$$

Where l_x, l_y, l_z is the cube's length along the subscripted x, y, and z axes, respectively.

Participants

13 participants were recruited from a local university campus. One participant was removed from the current analysis, as extreme responses to the planks task and experimenter observation during execution suggest they may have experienced difficulty in completing the task. The remaining 12 participants ranged in age from 20 to 28 (M: 23, SD: 2.5), 6 female and 6 male.

Materials

The experiment was administered in an 8m by 6m conference room, with a roughly 8m by 3m lane cleared for the experiment. The lighthouse tracking system of an HTC Vive was set to track a 6m long portion of this space, with the lighthouses positioned roughly 6.6m apart, connected by an optical sync cable and set to optical sync master and slave modes. Vive controllers were used as input devices during the two perceptual matching tasks, and to register beanbag landing positions during throwing tasks. An LED lightstrip was positioned to the left side of the space, and was used to indicate to the experimenter where real-world throwing targets should be placed, and whether Vive controllers were tracking accurately during beanbag landing registration. The LEDs were positioned roughly 2.5cm apart.

Design

This experiment followed a within-participant design, with each participant exposed to all conditions.

Three rounds of blind throwing were performed. Conditions of two factors were presented: *Viewing Condition* and *Target Distance*. *Viewing Condition* has four levels: the real environment (*real*), a virtual environment with no warp applied (*no warp*), a VE with a $warp_a$applied, and a VE with a $warp_{ab}$applied. *Target Distance* has three levels: 2, 3, and 4 meters. All target distances were used in all rounds of throwing. Target distances were always presented in five randomly ordered sets of the three levels, such that each target distance was presented five times. The actual distance presented was adjusted slightly from the base target distance, by ± 0.5m.

The first two rounds of throwing were pretests, *real* then *no warp*, over all distances for a total of 15 trials each. The third round of throwing was a post-test, and randomly presented both $warp_a$ and $warp_{ab}$ conditions over all distances, for a total of 30 trials, or 15 for each *Viewing Condition*.

Two perceptual matching tasks were performed. Both presented conditions of one factor, *Warp Method*, with three levels: *none*, *a*, *ab*. Both tasks presented scenes using the three methods in random order. Both tasks presented each warp method 5 times, for a total of 15 trials each.

Procedure

After reviewing and signing the consent form, participants were shown to the testing area. They were then asked to stand with their heels on a line on the ground made of tape, marking the zero distance point for the throwing task. Participants were asked to face forward along the axis on which targets would be placed, marked by another line of tape running forward between their feet. These tape lines were reproduced in the virtual environment used for throwing.

The throwing protocol was explained and the sounds indicating when to open and close their eyes were demonstrated. A target was placed near the center of the space, and participants were allowed to practice throwing beanbags at this target until they indicated confidence in their ability.

Participants were then asked to close their eyes, and performed the real environment round of pretest blind throwing. After this, the experimenter explained the use of the HMD, providing instructions on adjusting its fit. Once the headset was adjusted for comfort and clarity, the virtual environment pretest throwing was performed.

Once both pretest phases were complete, the per-participant parameters for the corrective manipulations were calculated. The post-test VE throwing phase was then executed.

Next, participants were asked to remove the HMD, and the use of the controller and the goal of the planks task was explained. Participants then wore the headset and performed the planks phase. After this, the cube phase was explained, then performed.

After completing the main experiment participants filled out a short demographic survey, and were asked to complete a colorblindness test using Ishihara plates, and a stereoblindness test using a random-dot stereogram. Finally, a post-experiment interview and debrief was then performed.

6.4 Results

Analysis

Throwing Task

Table 6.1: Summary of throwing trials.

Viewing Condition	N	Percent Error (PE) Mean	Percent Error (PE) SD	Relative Percent Error (RPE) Mean	Relative Percent Error (RPE) SD
real	180	-3.77	9.20	–	–
no warp	180	-10.81	12.49	-6.96	13.71
warp$_a$	180	-6.59	10.84	-2.79	10.74
warp$_{ab}$	180	6.23	20.23	10.03	19.83

Table 6.2: ANOVA of throwing trials. Df_n and Df_d are the numerator and denominator degrees of freedom, F is the F-value, p is conditional probability of the F-test, η_p^2 is partial eta-squared. PE is percent error, RPE is percent error in virtual environment trials relative to real environment trials.

Effect	On	Df_n	Df_d	F	p	Sig.	η_p^2
Condition (C)	PE	3	33	8.87	.008	**	.45
Distance (D)	PE	2	22	30.47	$< .001$	***	.73
C x D	PE	6	66	8.11	.001	**	.42
Condition (C)	RPE	2	22	10.55	.006	**	.49
Distance (D)	RPE	2	22	20.54	$< .001$	***	.65
C x D	RPE	4	44	2.86	.035	*	.21

Table 6.3: Pairwise t-tests comparing conditions from throwing trials. Holms-Bonferroni was used to correct for multiple comparisons.

Conditions		Metric	Df	t	p	Sig.
real	no warp	PE	179	-8.98	$< .001$	***
real	warp$_a$	PE	179	-3.0087	.003	**
real	warp$_{ab}$	PE	179	6.478	$< .001$	***
no warp	warp$_a$	RPE	179	-4.4355	.005	**
no warp	warp$_{ab}$	RPE	179	-10.432	$< .001$	***
warp$_a$	warp$_{ab}$	RPE	179	-10.691	$< .001$	***

Results of repeated-measures ANOVAs on the effect of viewing condition and target distance on percent error and relative percent error can be seen in Table 6.2; all factors show significant differences between groups. The effects of target distance are somewhat surprising and will be revisited in the discussion, but will not be further examined here.

Pairwise t-tests on viewing condition, as shown in Table 6.3, show that the RE and VE pretest conditions (*real* and *no warp*) are significantly different, which suggest participants exhibited distance misperception.

Each combination of *no warp*, $warp_a$, and $warp_{ab}$ were also found to be significantly different from one another, suggesting warp had some effect on perceived distance.

Table 6.1 and Figure 6.2 summarize the throwing trials. In *no warp* conditions the mean percent error (-10%) and relative percent error (-7%) are lower than seen elsewhere – percent error of 25% (Renner et al., 2013), or relative percent error of -18% (Peer and Ponto, 2017). This suggests participants experienced lower than average distance misperception. Changes in relative percent error between *no warp* to $warp_a$ conditions show mean error roughly halving, indicating the $warp_a$ condition had a positive effect on perceived distance (**Q1**). The $warp_{ab}$ conditions show significant overestimation, which suggests they did not have the intended effect (**Q1**, **Q2**).

Perceptual Matching Tasks

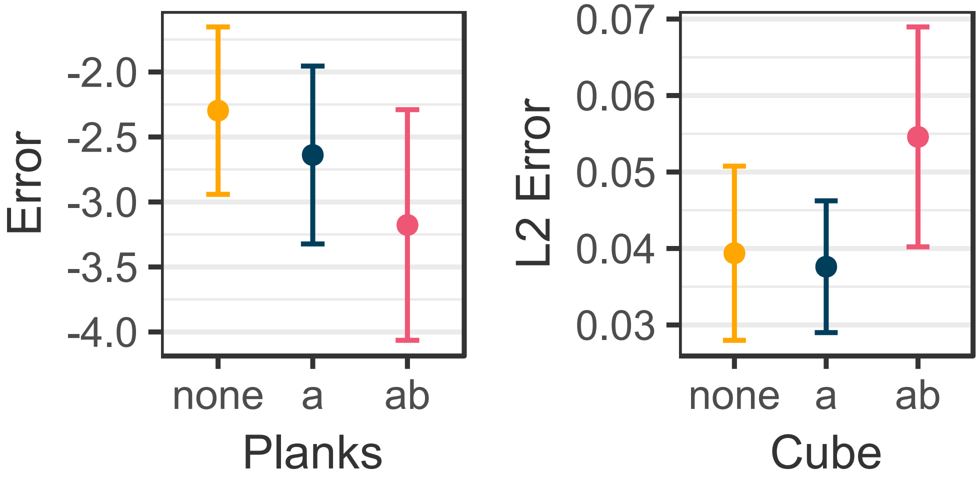

Figure 6.3: Mean error when performing the two perceptual tasks. On the left, degrees error when performing the planks task. On the right, L2 error when performing the cube task. Error bars represent 95% confidence intervals.

Table 6.4: Summary of the perceptual matching tasks' results.

		Planks (degrees)		Cube (m)	
		Error		L2 Error	
Warp	N	Mean	SD	Mean	SD
none	60	-2.30	2.49	0.039	0.044
a	60	-2.64	2.64	0.038	0.033
ab	60	-3.18	3.43	0.055	0.056

Table 6.5: ANOVA results for the two perceptual matching tasks. Headings as in Table 6.2, with the exception of dependent variable here being the error metric from the referenced task.

Task	Metric	Df_n	Df_d	F	p	Sig.	η_p^2
Planks	Error	2	22	1.38	.269		.11
Cube	L2	2	22	3.98	.033	*	.27

Table 6.6: Pairwise t-tests comparing warp methods during the cube task. Holms-Bonferroni was used to correct for multiple comparisons.

		Task	Df	t	p	Sig.
none	a	Cube	59	0.28	.778	
none	ab	Cube	59	-1.88	.064	
a	ab	Cube	59	-2.26	.027	*

Results of repeated-measures ANOVAs on the effect of warp method on the error metrics of the perceptual matching tasks can be seen in Table 6.5. Warp method is found to have a significant effect on the cube task's L2 error, but not on the planks task's degree error (**Q3**).

Pairwise t-tests comparing the error metrics of the various warp methods for the cube task are shown in Table 6.6; only $warp_a$ and $warp_{ab}$ conditions are found to be significantly different (**Q3**).

A summary of the observed error in perceptual matching tasks can be found in Table 6.4, as well as Figure 6.3. The planks task sees a trend towards increasing error and variance in both $warp_a$ and $warp_{ab}$ cases, relative to the *no warp* case (**Q3**). All planks conditions show significantly larger errors than seen in Ponto et al. (2013). The cube task sees similar amounts of error in *no warp* and $warp_a$ cases (**Q3**). $Warp_{ab}$ shows an increase of 0.016m mean error, as well as increased variance (**Q3**). The cube task shows mean errors of a similar magnitude as Ponto et al. (2013).

Discussion

In regards to the questions proposed in §6.1, results suggest that the $warp_a$ technique reduced distance misperception in the throwing task (**Q1**) and showed no significant influence in the two perceptual tasks (**Q3**). The $warp_{ab}$ technique, however, not only increased error in throwing, it caused overestimation (**Q1**, **Q2**); oddly, though $warp_{ab}$ does increase mean error in the perceptual tasks, it had no statistically significant influence (**Q3**). Overall, $warp_a$ may be a viable technique for mitigating distance misperception in VR; $warp_{ab}$, as implemented here, is not, but allows observation of the effect of extreme warp on the three tasks.

Table 6.7: Summary of warp parameters chosen to correct for each participant's differing amount of observed distance misperception.

		$Warp_a$		$Warp_{ab}$	
parameter	N	Mean	SD	Mean	SD
a	12	1.107	0.069	1.281	0.209
b	12	-	-	-0.343	0.276

That $warp_{ab}$ had such a large detrimental effect on throwing trials was unexpected. To fit as a proxy of eye position as in the model proposed by Ponto et al. (2013), we would hope that the values of the offset from $warp_{ab}$ be small, on the order of centimeters in magnitude. However, as can be seen in Table 6.7, calibration from the linear regression selected offsets which were in general quite large, in tens of centimeters – the mean w_b of -0.34 would shift the user's position backwards by 34cm. Not only is this well out of the range of candidate eye positions, it would result in significant changes to the scene at near distances when the a multiplier does less to cancel out the effect. For example, when looking directly at one's feet, the ground would appear to be almost 34cm higher.

This likely created a situation in which closer objects were overly displaced, as shown in the 2m column of Figure 6.4. It is interesting to note that while fairly extreme overestimation was shown in the throwing task, this did not translate to selection of significantly different negative angles of inclination in the planks task as would be indicated by Ponto et al. (2013). This may suggest that the throwing task and planks task draw from different depth cues, or that $warp_{ab}$ changes the cues used during one but not the other. This may also be due in part to differences between this experiment and that of Ponto et al. (2013). The earlier experiment used a CAVE, allowing for a visible physical reference object; the HMD used here did not. It may also be due to the differences in task implementations: the planks here are infinitely wide, rather than fixed width and jittered on the x-axis; the planks here have no apparent height, rather than 0.46m.

Several participants were observed looking along the planks to where the infinite length met the horizon, particularly under extreme $warp_{ab}$ conditions. Participants did not report this helping their judgments, however; when interviewed, it seemed that looking at an infinitely long stimulus that would not be influenced by the warp may have afforded a more comfortable or less cognitively demanding stimulus, though they hesitated to assign it to either when asked directly. This gives some slight indication that extreme warp may have some detrimental effect outside of task performance as measured in this work, one that participants may not be completely aware of themselves but do seek to alleviate.

It is also interesting that despite $warp_{ab}$ inducing particularly strong overestimation at the closest distance of 2m, it did not have an effect of similar magnitude on the cube task, which was viewed from distances exclusively closer than 2 meters. This may suggest that under extreme warps, a different strategy was employed – either different depth cues were used, or cues entirely separate from depth were somehow employed. This suggests users might adapt, even to extreme warps. Participants did not report consistently using any alternative strategies. Also worth noting is that participants had no trouble walking around the cube in either warp condition.

While $warp_a$ showed throwing task performance that better matched real-world conditions, it does not induce fully matched performance. Looking again to Figure 6.4, it seems $warp_a$ performed better at shorter distances; this may indicate that the multipliers were biased by there being so little distance misperception evident at the 2m distances in *no warp*, which is itself an anomaly. We see amounts of distance misperception similar to Interrante et al. (2006), who made the serendipitous discovery that a virtual environment matched to the real environment occupied by a participant appeared to induce greatly reduced underestimation – percent error on the order of -10%, in both real and virtual trials. The virtual environment used here was a spatially matched but sparse environment, in the style of Peer and Ponto (2017), which work uses both blind throwing and relative percent error and saw somewhat more underestimation (-18%) than we do here (-7%). It may be that the virtual environment used in the current work is somehow too good a match and runs afoul of the matched environment effect seen by Interrante et al. (2006), at least partially, and particularly at near distances. However, we do see more underestimation in virtual trials than real trials, and so our participants do seem to have experienced distance misperception.

$Warp_a$ also caused no significant detriment in performance for either perceptual matching task; for the cube task, this suggests that $warp_a$ did not introduce distortions that interfered with the task. For the planks task, improved distance estimates in the throwing tasks would be expected to correspond to reduced error in planks trials; however, no significant difference was found, and the mean error is slightly worse in $warp_a$ conditions than *no warp*. Similarly to the $warp_{ab}$ results, this may be due to differences between our implementation of the task and those of previous work (Ponto et al., 2013). It may also indicate that the depth cues used by the planks task either are not or are negatively influenced by the warp manipulation; further work is needed to decide.

At 4m, the two warp methods seem to bracket a relative percent error of 0; further, $warp_a$ provides the least mean error at 2m, and steadily more as distance increases; $warp_{ab}$ provides less mean error than $warp_a$ at 4m, with $warp_{ab}$ error increasing as distance decreases. This may indicate that some combination of the two might be better than either alone, that some nonlinear or piecewise linear warp function might be optimal. Further work might also explore $warp_{ab}$ at longer distances to see if the trend of decreasing mean error continues; conversely, adding pre-test measures at even closer distances might prevent extreme w_b values.

6.5 Chapter Conclusions

Of the two warp methods tested, $warp_a$ seems the most immediately viable, as it improved distance estimates in throwing trials and did not disrupt estimates in the cube task. However, it did not significantly influence the planks task, which was expected to be influenced by the sense of depth manipulated by the warp. The $warp_{ab}$ method used resulted in extreme warp parameters, which caused overestimation in throwing trials, but had no significant influence on either perceptual task. Overall, the basic premise of mitigating distance misestimation in virtual reality using personalized manipulations does seem viable. More work is needed to explore means of calibrating and implementing warps across a range of distances, and to observe the effects of warps on other classes of tasks than those explored here. Using similar calibration methods with manipulations other than warps may also be possible, with their own strengths and caveats.

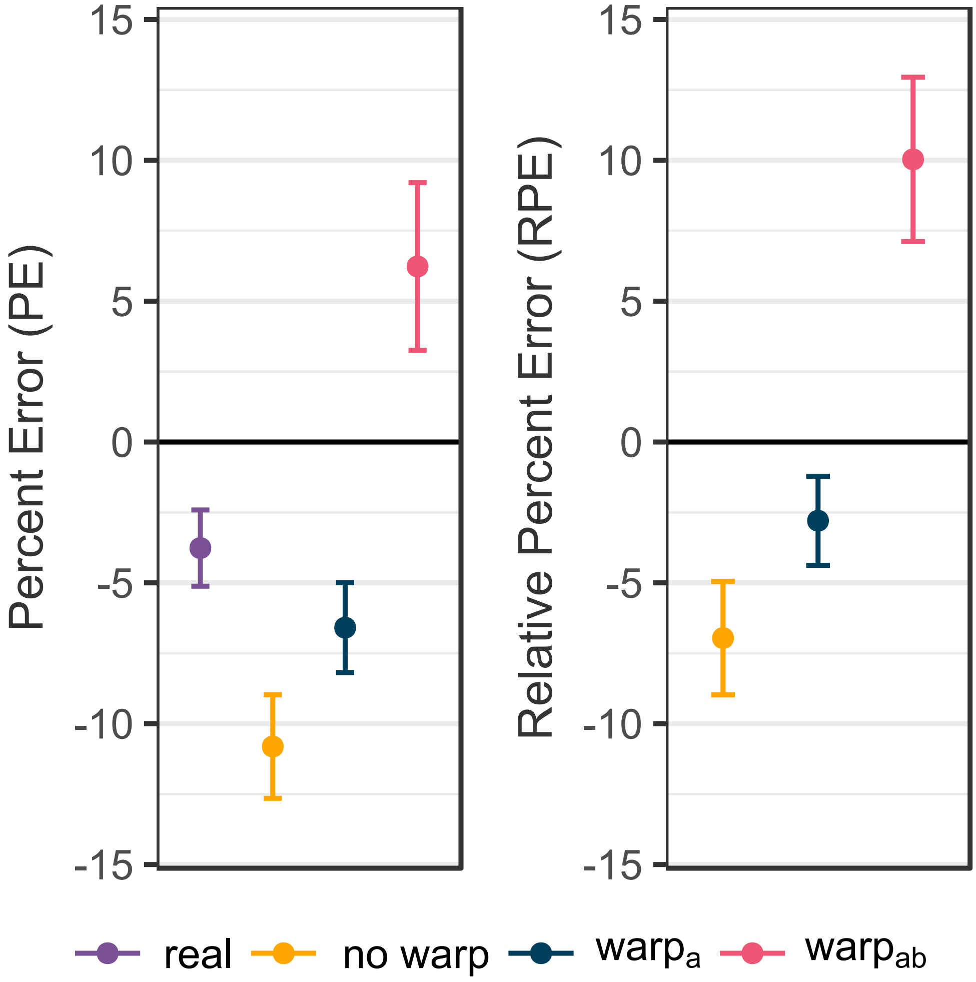

Figure 6.2: For the throwing task in each viewing condition, mean percent error (left) and mean relative percent error (right). Error bars represent 95% confidence intervals. Negative error represents underestimation, positive error represents overestimation.

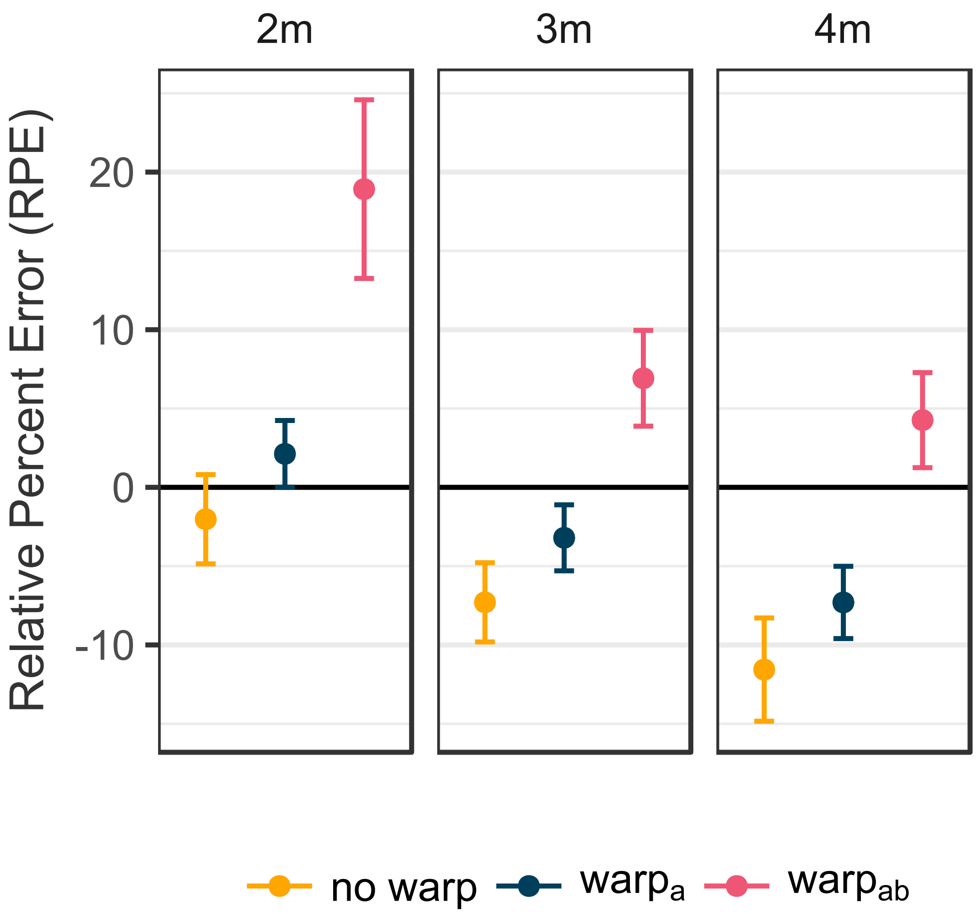

Figure 6.4: Mean relative percent error when performing the throwing task, split horizontally by the three base target distances. Error bars represent 95% confidence intervals.

7 MEASURING INDIVIDUAL DIFFERENCE : VISUAL ACUITY AND STEREO ACCURACY

Individual difference is evident in measurements of perceived distance. These are measurements of an effect; is it possible to measure some aspect of individual difference that might be closer to the cause? Visual and stereo acuity are two properties of human perception that are known to vary between individuals, are strong candidates for influencing spatial perception in stereo visual displays, and that might be expected be mediated by the limitations of current VR display technology. This chapter outlines methods to measure visual acuity and stereo accuracy as mediated by immersive displays, and a preliminary user study provides evidence that visual acuity is mediated by display resolution, and that measuring stereo accuracy requires further calibration to individual difference in stereo perception. [1]

7.1 Overview

Virtual and augmented reality experiences are driven by immersive displays, meant to present visual cues that elicit the sense of depth and presence we experience in our neither virtual nor augmented "real" lives. However, these displays have limits that can be expected to mediate the intended visual experience – screen resolution, lens distortion and fixed focal length, as well as any misalignments between eye and lens due to poor fit or adjustment, should all be expected to have some effect. These effects are difficult to predict.

In head-mounted displays (HMDs), for example, lenses cause pixels to be distributed non-uniformly across the visual field. With detailed enough

---

[1] This chapter is adapted from Peer and Ponto (2020)

knowledge of the properties of said lenses and the position of the viewer's eyes, this pixel distribution could be computed. However, HMD lenses are generally poorly documented. User eye position can be approximated using interpupillary distance (IPD), which should be expected to range between 50-75mm (Dodgson, 2004). If HMDs do not allow lenses to be moved to accommodate, if users don't make the adjustment, or if they inaccurately measure their IPD, a misalignment should be expected. It should also be expected that some amount of shift will happen in the fit of the HMD each time it is put on, and over time during use; movement of the HMD from the expected fit will introduce misalignments. Even given perfect alignment, normal eye movement may cause meaningful misalignments (Jones et al., 2016). Measuring, modelling and predicting the possible distortions induced by such potentially complicated systems may not always be practical.

Instead, it may be better to observe the mediating effects of the display on the intended visual stimulus, as perceived by the viewer. In this work we attempt to quantify some of these effects by providing tests of visual acuity and stereo accuracy that can be administered using a variety of immersive displays. We provide proof of their feasibility by testing their use with two HMDs, the HTC Vive and the Valve Index, and provide a preliminary analysis of the relative effects of the two display platforms on visual acuity and stereo accuracy.

7.2 Related Work

Computer-based methods of measuring visual acuity exist in the literature. Bach (1996) describes the Freiberg Visual Acuity Test. It provides a screen-based method to measure several properties of the human visual system, using a variety of stimuli and possible units of measure; a web-based version is available (Bach, 2019a). Fidopiastis (Fidopiastis et al., 2005;

Fidopiastis, 2006) created a similar test for use in the development of a retro-reflective immersive stereo display, though this implementation no longer seems to be available for use. Bach and Fidopiastis both suggest the use of Landolt C optotypes presented at differing orientations and sizes, but differ in their preferred method of selecting stimuli to be presented and subsequent analysis: Bach favors an adaptive staircase method, best PEST (Lieberman and Pentland, 1982), as it allows for reduced trials and directly provides a final estimate of acuity (Bach, 2006); Fidopiastis favors the method of constant stimuli, citing concerns over inaccurate estimates from staircase methods due to accumulating errors (Kaernbach, 2001; Pinkus and Task, 1998). These works also differ in one other way worth noting: Bach trusts anti-aliasing to smooth the edges of small stimuli, where Fidopiastis prefers to tailor the presented stimuli to forms and sizes that can be rendered in perfect alignment to the display's underlying pixels. In this work, we employ Landolt Cs, the method of constant stimuli, and allow the HMDs' rendering pipelines to render small stimuli as they normally would.

Both Bach and Fidopiastis have also developed tests of stereo perception. Bach et al. (2001) describes a stereo acuity test, a screen-based test similar to FRaCT, though it no longer seems available via website. This test presents a line that is presented at some distance from a surrounding frame, and viewers are asked to indicate whether the line is in front of or behind the frame; a best PEST method is again favored to choose levels of stimuli to be presented and to provide the final estimate. The stimuli is phrased in terms of raw stereo disparity, and the measure is taken as stereo acuity. Fidopiastis et al. (2005) creates a similar test for use in immersive displays, an extension of earlier work by Rolland et al. (2002) and inspired by the Howard-Dollman two-peg test (Howard, 1919). The original Howard-Dollman test presents two physical rods and asks participants to align them to the same depth; this allows for several nonstereo cues, such as changes in the apparent size of the rod as it moves towards and away from the viewer. The solution found in these more recent immersive display methods is to present multiple objects of differing shapes to foil known-size cues, as fabricating and positioning new shapes is comparatively easy in the virtual space (Rolland et al., 2002; Fidopiastis, 2006). Fidopiastis describes their stimuli in terms of distance between two objects, and so terms their measure stereo accuracy rather than acuity. Fidopiastis again favors the method of constant stimuli. In this work, we adopt a method most closely related to that of Fidopastis – we adopt the method of constant stimuli and measure stereo accuracy. Our stimuli is similar to a Howard-Dollman device in that it is presented in the virtual world rather than through raw stereo disparity, but rather than presenting multiple objects to foil known-size cues, we use a method of presenting only stereo cues that may be novel for this style of test.

Problems of misperceived depth in immersive displays are well documented, and are sometimes called depth compression. In a 2013 survey of 30 papers studying this effect in virtual reality displays, Renner et al. (2013) found that distances were underestimated by an average 26%. These papers found many factors that appear to contribute to the effect, but none that could account for it fully. Armbrüster et al. (2008) found greater stereo acuity to correspond to greater distance misperception. Visual and Stereo acuity has otherwise been largely ignored in the literature, though poor visual or stereo acuity is sometimes used as an exclusion criterion for participants, with measurement method and data rarely reported in detail. The tools presented in this work might be used to further investigate the relationships between stereo accuracy, visual acuity, and distance misperceptions. Experiments like those discussed in Renner et al. also depend heavily on precisely engineered visual stimuli; visual acuity and stereo accuracy measures might be used to establish that carefully crafted stimuli are able to be received as intended. Unfortunately, we are not

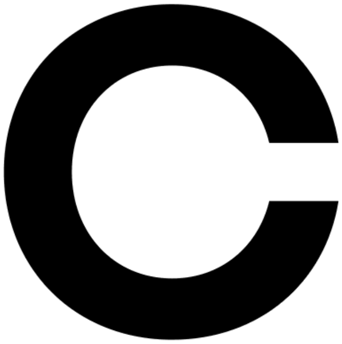

Figure 7.1: A Landolt C, used during visual trials.

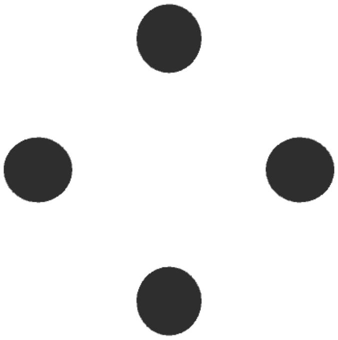

Figure 7.2: An example of the four dot stimuli of the stereo accuracy task.

aware of any currently available measurement tools for use in modern immersive displays; we hope the software described in this work might provide a common set of tools for the larger academic community to use for the measurement of visual acuity and stereo accuracy as mediated by immersive displays.

7.3 Method

Visual Acuity

A common method of measuring visual acuity is to present some stimulus that subtends a known angle on the viewer's visual field. One common type of stimulus is the Landolt C, designed such that a C 5 arc minutes tall and wide would have a 1 arc minute tall gap on one side. An example of the Landolt Cs used in this work can be seen in Figure 7.1.

The size of the C corresponds to the visual acuity needed to consistently discern which direction it is facing, as defined by the following formula:

$$VA_{decimal} = 1/\theta \tag{7.1}$$

Where θ is the size of the C's gap in arc minutes, and the result is decimal visual acuity ($VA_decimal$). This form directly correlates to the perhaps more commonly known Snellen Ratio, where 20/20 vision corresponds to a decimal acuity of 1.

When used to measure visual acuity, a series of Cs are shown in various orientations and sizes; viewers are asked to indicate the way the C is "facing", which is the orientation of the C's gap relative to C's center. The C in Figure 7.1 is facing to the right. In our test, we allow the C to face in four directions: up, down, left, and right. We limit these to four for ease of user input.

In the tool described in this work, the C is held at a fixed position 6m directly in front of the participant's viewpoint, such that it stays centered in the participant's view independent of head movement. The size of a C subtending a given angle was determined as follows:

$$H = 2 * D * Tan(\Phi/2) \tag{7.2}$$

Where D is the distance in meters between the participant's viewpoint and the C, Φ is the desired subtended angle, in radians, and H is the height of the C, in meters, needed to achieve the intended subtended angle.

Various strategies exist for choosing the sequence of sizes to show to best identify the viewer's visual acuity. In this work, we implement an automated test using the method of constant stimuli as described in later in this section.

Stereo Accuracy

To test stereo accuracy, we display four unlit spheres that appear to the viewer as four dots. Similar to a Howard-Dolman test (Howard, 1919;

Fidopiastis, 2006; Rolland et al., 2002), one dot is placed at a different distance than the others; for the sake of this work, this dot is always closer to the viewer. Each dot subtended 0.25 degrees in the participant's visual field; their scale was adjusted to hold this constant apparent size, regardless of the distance they were presented at, such that changes in size could not influence perceived distance. Dots were also kept in the same location in the participant's field of view, such that parallax and other movement-based depth cues were not available.

The smallest distance between the near and far dots at which the participant could consistently identify the near dot is taken as the measure of their stereo accuracy. In this work the far dots are always held at 6m, but this is a limit of the scope of the work and not inherent to the method.

Similar to our visual acuity stimulus, many strategies can be employed for choosing the sequence of distances to show in order to best find the participant's stereo accuracy. The strategy we've implemented is described in Section 7.3.

Automated Test: Method of Constant Stimuli

The method of constant stimuli involves establishing a range of interest within the space of possible stimuli, dividing it into evenly spaced intervals, and sampling multiple times at each interval (Fidopiastis, 2006; Bach et al., 2001). For our two stimuli, we established our expected ranges of interest in pre-experiment pilot trials, and held them constant for each participant. For the automated visual acuity test, Cs of a range of different sizes and facing random directions were presented; for the automated stereo accuracy test, a range of different distances between near and far dots was presented, and the direction of the near dot was selected at random.

For each trial, participants are asked to indicate the direction that the C is facing, or the direction of the nearest dot. The percent of correct responses at each interval is then calculated, and a curve is fit to these; the intercept between this curve and the halfway point between random chance and 100% on the y-axis is used to determine the point of the participant's visual acuity or stereo accuracy on the x-axis. Both of our tests present four choices with one correct, giving a 25% chance of randomly choosing the correct direction. The halfway point between 25% and 100% is 62.5%, and is used in this work as the threshold to determine visual acuity and stereo accuracy. Figure 7.3 demonstrates this process.

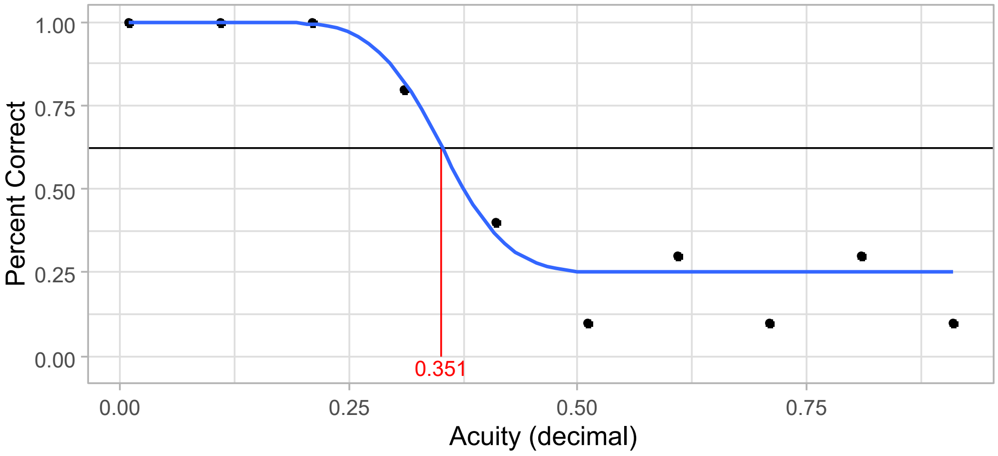

Figure 7.3: An example of the analysis of the data generated by the method of constant stimuli, using data from one participant's automated visual acuity test using the Vive. A horizontal line indicates the 62.5% threshold, with the red line indicating the intersection between the curve and this threshold indicating the participant's visual acuity.

Participants

10 participants were recruited from a local university campus. Participants ranged in age from 20 to 38 years old (M: 24.7, SD: 5.7), 3 female, 5 male,

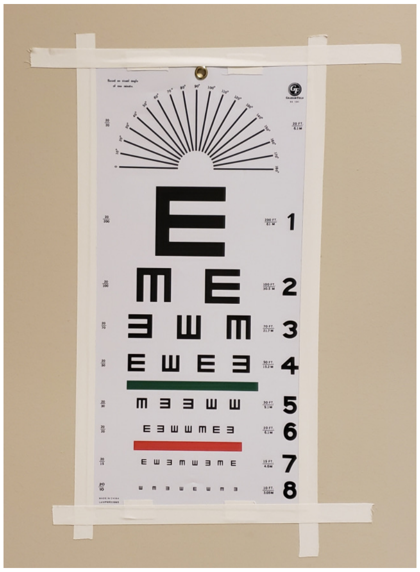

Figure 7.4: The Tumbling E eye chart used in the experiment.

Figure 7.5: The Pupillary Distance Meter used in the experiment.

and 2 who declined to provide a gender.

Materials

A tumbling E eye chart, depicted in Figure 7.4, was used to determine participant's baseline visual acuity. Administering this eye test works similarly to the Landolt C used elsewhere in the experiment; participants indicate the direction the Es face, with each line providing Es that subtend specific angles that correlate to different visual acuities. A Sunwin Digital Pupillary Distance Meter was used to measure participant's interpupillary distance (IPD). The unit used provides measurements at 0.5mm intervals. The device can measure a variety of vergence distances; far IPD was measured by setting the device to "infinity". Two headsets were used to present our stimuli: the HTC Vive and Valve Index. The Vive provides a resolution of 1080 x 1200 pixels per eye; the Index provides a resolution of

1440 x 1600 per eye. The Unity game engine was used to present stimuli and to receive and log responses from participants.

Design

This experiment uses a within-participant design, with each participant performing both measures in both HMDs: the Vive and the Index. Order of presentation of HMD was counterbalanced, with half of participants using the Vive first.

The automated visual acuity trials varied the decimal acuity of presented stimuli between 0.01 and 1.01 (logMAR of 2 to -0.004) at 10 intervals of 0.1 decimal acuity. A randomly ordered series of these 10 intervals was presented 10 times, for a total of 100 trials per HMD. The automated stereo accuracy trials held the far dots constant at 6 meters from the participant, and varied the distance between the near dot and far dots within a range of -1 to 0 meters, with negative indicating movement towards the participant, at 20 points spaced at 0.05m intervals. A randomly ordered series of these 20 intervals was presented 5 times, for a total of 100 trials per HMD.

Procedure

After entering the experiment space, participants were first asked to read and sign the consent form, and given five dollars compensation for their time. A tumbling E visual acuity chart was then used to establish participant's baseline visual acuity. Participants were asked to stand 10 feet away from the chart, and it was explained that to read a line from the chart participants should verbally indicate the direction the Es face, from left to right; the topmost line was used as an example. Participants were then asked to read the bottommost line, corresponding to a visual acuity of logMAR 0 (normal corrected vision). Participants who were to wear eyeglasses when using the HMDs also wore them during this test. All participants were able to successfully read this line, indicating normal or corrected to normal vision.

Participants were led to a seat, and remained seated for the rest of the experiment. The participant's interpupillary distance (IPD) was then measured. A Sunwin Digital Pupillary Distance Meter was used. When using the device, participant's hold one side to their eyes, with the metal bar on top and nosepiece on the bottom aligning the device. Participants are asked to look at an illuminated green dot within the device, which is placed at a distance corresponding to zero vergence by lenses in the device. The experimenter peers through the other end of the device, adjusting two sliding controls on the top to align vertical lines with the center of the participant's pupils. The device is able to present monocular stimulus and to present a variety of vergence distances; for this experiment, far IPD with a stereo stimulus was measured.

The keyboard to be used for input was then presented to the participant. It was explained that participants would use the arrow keys to indicate directions, and that they wouldn't be able to see the keyboard while wearing the HMDs; The location of the arrow keys on the keyboard was indicated. The first HMD was then presented to the participant. When a new headset was presented, the methods of adjusting its fit were explained. The experimenter also adjusted the HMD's IPD to match the participant's, as measured earlier. Participants donned the HMD and adjusted for comfort and clarity. When participants indicated they were ready, participants were shown the visual acuity task, and it was explained that the participant should press the direction key corresponding to the direction in which the C was "facing" – that is, the direction of the gap in the circle. Eight practice trials were presented at a range of 0.01 to 1.01 decimal acuity (2 to 1.4 logMAR), which presented stimuli large enough that all participants were able to correctly identify the indicated direction. Participants were instructed that during the full test, some of the Cs would be hard or impossible to see; in this case, they should give their best guess or, if they had no guess, simply press any direction. The automated visual acuity test was then administered.

After the visual acuity test, the stereo accuracy test was presented. Participants were shown the first of 8 practice trials, and the experiment was explained: four dots corresponded to four directions, and one dot should appear closer to the participant than the others; the participant should press the directional key matching the direction of this closest dot. In these practice trials, the distance between the close dot and the far dots varied between 2 and 3 meters, which we expect to present a detectable difference for any participants who are stereo sensitive. After the practice trials were complete, the automated stereo accuracy test was administered. After all tests were completed in the first HMD, participants were asked to remove the HMD and complete a brief demographic survey. The participant was then presented with the second HMD, and the tests were presented in an order similar to that for the first headset, excluding practice trials: automated visual, then automated stereo. Participants were then thanked for their participation, and given the opportunity of a brief debrief. Each automated round of testing averaged about 5 minutes, and the entire experiment, about 30 minutes.

7.4 Results

Analysis

Analysis was performed using R. Curve fitting as described in Section 7.3 was achieved using glm with an mafc.probit link provided by the "psyphy" package, which is designed for the analysis of multiple-alternative forced choice tasks with one correct answer, as in both of the tests discussed in this work.

Visual Acuity

The curves fit to the results of the automated visual acuity tests can be seen in Figure 7.6. Curves were well fit, suggesting the range selected contained all participants' visual acuity.

Visual acuities as measured by the automated test were converted to logMAR before calculating means and standard deviations (Bach, 2019b). The Index saw a mean acuity of 0.56 logMAR (SD 0.15), which corresponds to decimal acuity of 0.27. The Vive saw a mean acuity of 0.69 logMAR (SD 0.10), which corresponds to a decimal acuity of 0.20. "Normal" vision is generally considered to be a decimal acuity of 1; both headsets presented notably worse than normal visual acuity, the Vive slightly worse than the Index.

A paired-samples t-test was performed to compare visual acuities in Vive and Index conditions, and a significant difference was found; t(9)=-3.947, p=0.003.

Stereo Accuracy

The curves fit to the results of the automated stereo accuracy tests can be seen in Figure 7.7. Curves overall are not well fit, as evidenced by their relatively flat rather than S-shapes: it appears that some participants have stereo accuracies between -0.25m and 0m, which would be found with more accuracy using smaller intervals over that range; it appears that some participants may have stereo accuracies of -1m or less, which cannot be detected by the current range of stimuli presented. Four participants' fits did not afford valid intercepts with the 65.2% threshold, and were removed from calculation of summary statistics: for the remaining six participants, the mean stereo accuracy when using the Vive was -0.35 meters (SD 0.29), and for the Index was -0.37 meters (SD 0.14).

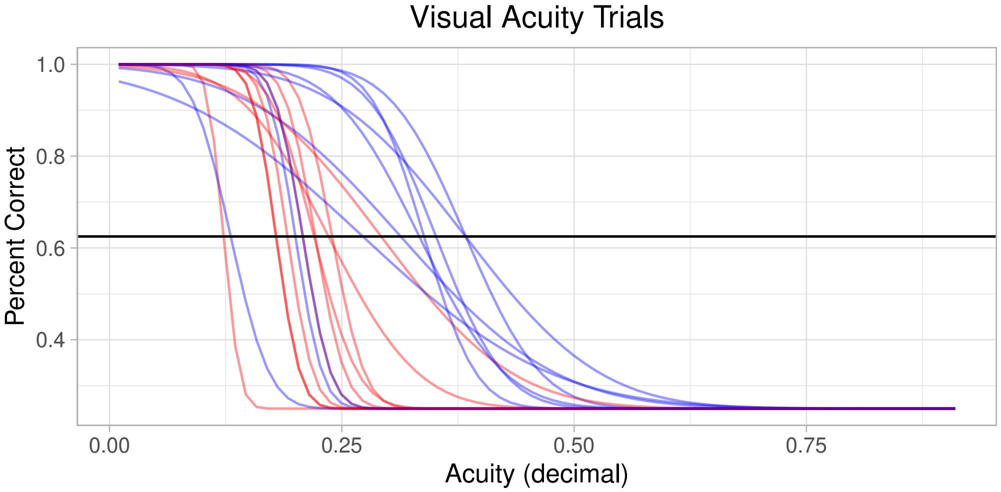

Figure 7.6: Curves fit to each participant's automated visual acuity trials. Color indicates headset: red for the Vive and blue for the Index. A horizontal line runs across the 62.5% threshold; the point at which each curve intersects this line indicates that participant's visual acuity as measured when using that headset. A decimal visual acuity of 1, here at the right, indicates normal vision; these curves indicate reduced visual acuity in both headsets, but greater acuity in the Index overall.

A paired-samples t-test was performed to compare stereo accuracies in the Vive and Index conditions, and no significant difference was found ($t(6)=0.373$, $p=0.72$).

Discussion

Visual acuity was significantly reduced in both headsets, but slightly more so in the headset with the lesser resolution; this is not entirely surprising. Stereo accuracy results suggest our 1m range of test stimuli was not suitable in general, and will need to be tailored to the individual. Several participants were not able to reliably detect the correct dot during test trials, though they were able to during practice trials; this suggests they

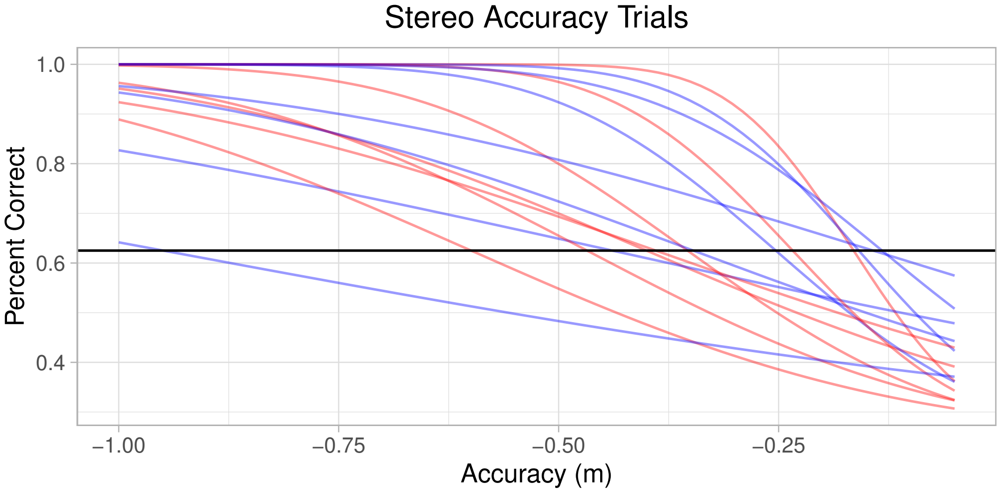

Figure 7.7: Curves fit to each participant's automated stereo accuracy trials. Color indicates headset: red for the Vive and blue for the Index. A horizontal line runs across the 62.5% threshold; the point at which each curve intersects this line indicates that participant's stereo accuracy as measured when using that headset. As stereo accuracy approaches 0, here at the right, participants are able to discern the difference between objects that are closer together; a stereo acuity of -1 would indicate participants have difficulty discerning a one meter difference between objects. Curves' fits are poor overall; 4 participants were removed from this chart, as their curves achieved particularly poor fits.

were able to use stereo cues, but for larger differences in position than tested. Other participants may have been able to detect the correct dot with smaller differences than tested. This was a surprising result, and in future work the stereo test should be paired with an out-of-headset stereo acuity or accuracy pretest to identify if this effect is more due to human or headset.

Beyond the headset comparison results, this experiment shows that participants can perform our two tasks successfully – that the basic form of the tasks and stimuli seem to be viable. During the debrief, participants provided feedback on the experiment. Most participants felt they performed better in the Index on both tasks, though one felt the larger pixels of the Vive exaggerated differences and so made the tasks easier. Participants indicated they found the tasks easy to perform, though some mentioned that the high number of "impossible" trials were discouraging; "easy" trials like those used during practice could be added occasionally to prevent this effect, as in Bach's FRaCT (Bach, 1996). The most common complaint of participants was discomfort due to the high contrast of the black-on-white stimuli and the brightness of the white screen, with the Vive being noticeably brighter. We suspect that a reduction of the number of trials needed would address any concerns of comfort: future work might look to accomplish this by replacing the method of constant stimuli with something like best PEST.

7.5 Chapter Conclusions

It is our hope that the visual acuity and stereo accuracy tests presented here can be useful to the larger community of researchers; if measuring individual differences between participants is made convenient enough, it may become a more common area of study.

8 ALIGNING VIRTUAL AND REAL SCENES

A common task in developing visual stimuli and implementing perceived distance measures for distance misperception experiments is the alignment of virtual and real coordinate spaces. Through the course of my work, it also became clear that there was a source of positional error in the consumer-grade tracking system we used, which may render its use in research impractical. This chapter described efforts to quantify the behavior of this error and correct for it. It also suggests a generally applicable method for aligning virtual and real coordinate spaces that can be adapted to correct for this or similar sources of tracking error. [1]

8.1 Overview

Alignment of real and virtual space is often desirable in virtual reality. At a basic level, accurate tracking of user movement and eye position are requirements for the rendering of immersive environments, and simulation of interaction within them. Some scenarios may involve interacting with physical objects while viewing the virtual world, and so further require a high-degree of accuracy in aligning the virtual and physical space. The virtual spaces described by tracking systems are arbitrarily aligned to real space, as determined by placement of tracking hardware and initial calibrations. An alignment between the real and virtual space can be established through exacting measurements, which is practical for long-term static installations in a single location; recent consumer grade tracking systems, however, lend themselves to quick deployment in new spaces and dynamic temporary arrangements. A correspondingly quick method of alignment is desirable. These alignments are generally a translation in

---

[1] Portions of this chapter are adapted from Peer et al. (2018)

3-space and a rotation around the y or "up" axis, such that the positions of any two points on the real and virtual floor planes match.

The tracking system of the HTC Vive has also been shown to display systemic errors (Niehorster et al., 2017; Peer et al., 2018; Luckett et al., 2019), most notably an error in tracked object height that can be described as a rotation around the x and z axes of the tracked space. Call this error *Tilt*. This sort of error has an impact on attempts to accurately align virtual and real spaces, and may be a threat to the basic levels of tracking fidelity necessary to facilitate immersive rendering and interaction. This misalignment appears to change when tracking of the headset is temporarily lost, and when the SteamVR software is initialized. The degree and direction of misalignment as yet appears to be random, but may be on the order of tens of centimeters. Niehorster et al. (2017) conclude that this error renders the Vive unusable for experiments when loss of tracking is a possibility. They show a correction is possible by individually measuring the tracked position of 36 points on a 9 x 4 grid, then finding a transformation that minimizes the root mean square error between the measured and ideal grids. They conclude this is an impractical method to employ mid-experiment, which is reasonable; however, we suggest that practical methods of correction do exist.

We propose two methods using only three tracked points, which align the virtual and physical spaces, account for the errors seen by Niehorster et al. (2017), and are practical for use during experiments. Both methods spring from a basic assumption: A plane can be derived from three points. If three tracked points are arranged on a known way in the physical space, deviations from their physical positioning in the virtual tracked space can be taken as error in describing this plane; this can be used to derive an *alignment* between tracked and real space (a translation and a rotation around y), and will also correct for the Tilt errors in height seen by Niehorster et al. (2017) (rotations around x and z), a correction we'll call *untilt*.

8.2 Method 1

Method 1 is a proof-of-concept and initial exploration of automated alignment, and the nature of the error introduced by Tilt. Additional steps would be needed to use Method 1 to correct for Tilt in real time, which are explored in Method 2.

Materials and Apparatus

A 3m x 3m grid was marked on the floor at 0.5m intervals, yielding a 7 x 7 grid of 49 points. Grid points were marked by black lines on pieces of white tape, and a 4m x 4m square around the grid was marked in white tape. An HTC Vive purchased in 2016 was used. The Vive's lighthouses were mounted on tripods at a height of 2.4 meters from the floor, and placed at opposing corners of the 4m square around the grid. The Vive headset was located on a table just outside of the grid. The Unity game engine and SteamVR library were used to acquire tracking information from the Vive system, using a coordinate system where negative z is forward and y is up. Three Vive Trackers were affixed to PVC pipe using 3D printed fittings, arranged to form two perpendicular 0.5m vectors on the ground plane. This apparatus can be seen in Figure 8.1.

Measurement

The tracked location of individual grid points was measured by individually placing the central Vive tracker at each point, with the other trackers positioned at neighboring grid points to assist in alignment. The tracked position of the Vive tracker was sampled for three seconds, and the average position was taken as the measured position at that grid point. The tracker was then moved to another point, and another 3 second measurement was taken; this process was repeated for all grid points.

Correction

Though our eventual goal is real-time automated correction using the Vive Tracker apparatus herein described, this work approximates its function using the measured location at grid points (0,0,0), (0.5,0,0), and (0,0,0.5). It should be noted that these measurements all come from the same Vive Tracker, and may lack some intra-device variance; they also may be less well aligned than three different pucks rigidly affixed to enforce a right angle as depicted in Figure 8.1. However, they serve well as a proof-of-concept.

Using these points, the necessary alignment transformation can be derived. Given our three points marking the intended origin, x-axis, and z-axis:

$$\begin{aligned} P^o &= \mathtt{grid}[0,0,0] \\ P^x &= \mathtt{grid}[0.5,0,0] \\ P^z &= \mathtt{grid}[0,0,0.5] \end{aligned}$$

We can derive two vectors describing the directions of the axes:

$$\begin{aligned} \vec{x} &= P^x - P^o \\ \vec{z} &= P^z - P^o \end{aligned}$$

And normalize both vectors, to yield $\hat{x}$, $\hat{z}$. If $\hat{\imath}$ is the normalized form of some vector $\vec{\imath}$, this operation becomes:

$$\hat{\imath} = \frac{\vec{\imath}}{||\vec{\imath}||}$$

The cross product of these vectors yields a unit vector on the intended y-axis:

$$\hat{y} = \vec{z} \times \vec{x}$$

And these three vectors directly describe a 4x4 rotation matrix needed to align the virtual and real spaces:

$$\mathbf{M}_{\text{rot}} = \begin{pmatrix} \hat{x}_x & \hat{x}_y & \hat{x}_z & 0 \\ \hat{y}_x & \hat{y}_y & \hat{y}_z & 0 \\ \hat{z}_x & \hat{z}_y & \hat{z}_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We also need a translation to align the scene to the origin:

$$\mathbf{M}_{\text{pos}} = \begin{pmatrix} 1 & 0 & 0 & -P^o_x \\ 0 & 1 & 0 & -P^o_y \\ 0 & 0 & 1 & -P^o_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

And composing these transforms through matrix multiplication yields our final alignment transformation:

$$\mathbf{A} = \mathbf{M}_{\text{rot}} \mathbf{M}_{\text{pos}}$$

The Matrix A describes the translation and rotation about y needed to align the virtual and real scenes, and the rotations about x and z that describe the degree of Tilt currently in effect.

Results

As a proof-of-concept, these section discusses applying the matrix A by multiplying points by this matrix directly; this should be expected to be an imperfect correction, as Tilt appears to affect only y position, and applying a rotation on x and z, as in A, should also be expected to affect x and z position. For a discussion of a more complete correction, see Method 2 later in this chapter.

Measurements were taken of the tracked position of 49 points on a 7 x 7 grid, with grid points at 0.5m intervals. An alignment matrix was derived from the measures taken in one corner of the grid, as if our apparatus had been placed there (as discussed in §8.2). A partial alignment matrix was also derived, using only $\hat{x}$ and P^o to describe a 2D translation and rotation about the y-axis. The partial alignment allows the real and virtual coordinates to align for the sake of generating figures, without correcting for Tilt errors in height.

Table 8.1 and Figures 8.3 and 8.2 show the results of these measures, with the *uncorrected* entries having the partial alignment matrix applied, and the *corrected* entries having the full alignment matrix applied. Improvements between *uncorrected* and *corrected* cases are due to the elimination of the errors described in Niehorster et al. (2017). The corrected case is more level, as seen in the more uniform coloring of Figure 8.3.

As seen in Table 8.1, the standard deviation, range, and total error in height are all reduced by an order of magnitude. Some error still remains; the range of error is perhaps most descriptive, as it shows participants in the corrected case would still experience almost 3cm of difference in, say, eye height when walking from one end of the tracked space to the other. This is an improvement over the 20cm in the uncorrected case, but may still influence experiments that require extremely accurate representation of space.

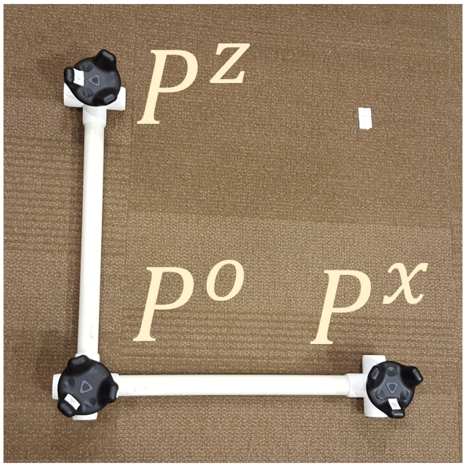

Figure 8.1: The proposed Vive Tracker apparatus.

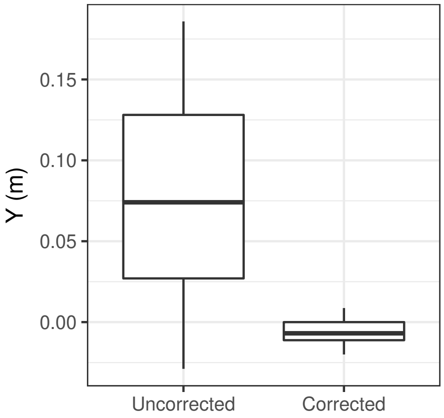

Figure 8.2: Difference in tracked object heights at the same physical height, using Method 1. Deviation from 0 is error.

8.3 Method 2

A second method of scene alignment and Tilt correction is currently being developed. This method is more general, in that it does not require three tracked objects arranged to describe a right angle on the x-z plane, but rather any three positions in tracked space whose real space counterparts

Table 8.1: Summary of variation in tracked object height seen using Method 1, in meters.

	Mean	SD	Range	Total Error
Uncorrected	0.075	0.061	0.214	3.91
Corrected	-0.006	0.007	0.028	0.38

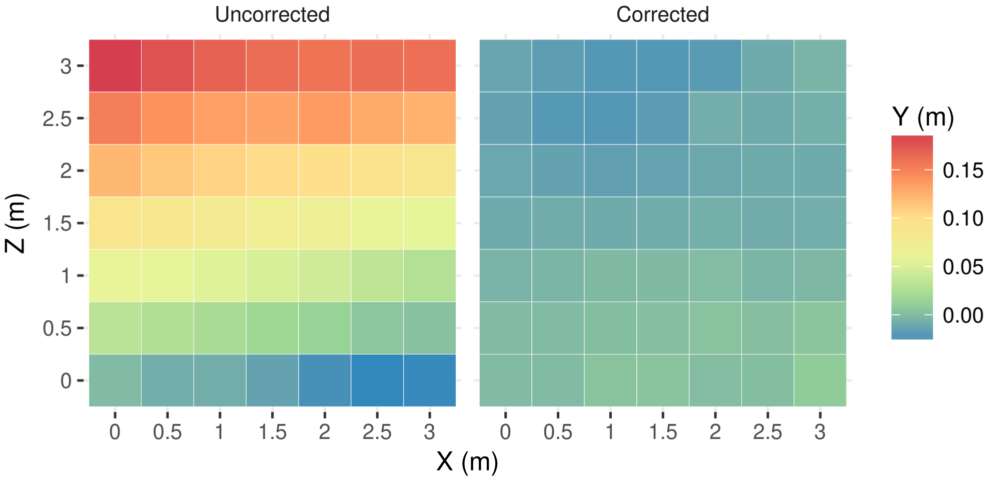

Figure 8.3: Error in tracked object height at various grid points in uncorrected and corrected cases, seen using Method 1.

are known. This allows for more flexible selection of tracked reference objects, including the possibility of using three SteamVR base stations so as to require no further instrumentation of the space. This method also explores real-time application of Tilt correction.

Correction

We have three coordinate spaces of interest: tracked space, which represents the current position of tracked objects as detected by the tracking system; real space, the physical world from which we can measure reference points; virtual space, which contains the virtual model of the reference points as measured in the real space, alongside whatever other virtual environment is rendered. In this phrasing of the problem, real space provides the measurements for the real reference points, and the virtual and tracked spaces are being aligned. The tracked space approximates the real space to some level of accuracy, but the relationship between the coordinate spaces of the virtual and tracked spaces are entirely arbitrary. We want to align the virtual and tracked spaces, and remove any error between the real and tracked spaces.

We begin with six points, three in each of the two spaces we want to align. Each point has a known matched counterpart in the other space: our goal is that a point $\vec{t}$ in tracked space match the position of expected point $\vec{v}$ in virtual space.

First, we calculate the centroid of our two sets of points, our constellations. The difference between the centroids provides the translation component of our alignment.

Calculating the rotation component requires further steps, with some additional complications to correct for Tilt. I'll first describe the basic process, then the modifications to address Tilt. Given our two sets of points and their centroids, we subtract the centroid of each set from their member points so that rotations will be calculated as centered at these centroids. We further subtract the centroid of the set containing the intended positions (those of virtual space) from each of the points of the set we want to change to be aligned (tracked space). We then apply the Kabsch algorithm (Kabsch, 1976) to determine a rotation that minimizes root mean square error, where error is defined as the difference in position between paired points. This provides the rotation component of our alignment, without accounting for Tilt.

This base method of calculating a rotation to align the tracked and virtual spaces does not take Tilt into account, which presents errors only in y position. We can adapt this basic application of the Kabsch algorithm to account for Tilt in a simple way: first, we project our points to the x-z plane, so that both constellations are on a 2D plane without the y-position information that might be influenced by the tilt artifact. The Kabsch algorithm is then used to determine the optimal rotation about the y-axis to match these 2D sets. This rotation provides the rotation component of our alignment, with any influence of tilt removed. The translation described by the 3D constellations' centroids and this rotation component derived using the 2D x-z constellations we will call our *alignment*; we expect this to position the virtual space to match the tracked and therefor real spaces, with the exception of error introduced by Tilt.

Tilt appears to manifest as an error in the y position of tracked objects, as if y coordinates were calculated relative to a titled ground plane. We can correct for this by first calculating the degree of Tilt currently in effect, and later applying a correction to the position of tracked objects every frame.

To calculate the degree of Tilt currently in effect, I propose the following method: First align the 3D constellations using the alignment calculated for their 2D x-z plane counterparts. The remaining difference between the constellations on y is assumed to be due to tilt. Calculate the difference between the tracked and virtual points on the y axis, and generate a new constellation using the tracked x-z positions and this difference on y. Call this constellation d, and any member of this constellation $\vec{d}$. If we similarly let the tracked constellation be t, members $\vec{t}$, and the virtual constellation be v and members be $\vec{v}$, then for each paired $\vec{t}$ and $\vec{v}$, a corresponding $\vec{d}$ is defined as:

$$\vec{d} = \{\vec{t}_x, \vec{t}_y - \vec{v}_y, \vec{t}_z\}$$

This new constellation captures the Tilt currently in effect at the x-z coordinates of the members of the tracked constellation. A plane fit to this constellation will describe this Tilt across the whole of the tracked space. To calculate this plane, first calculate the centroid $\bar{d}$ of the constellation d, then calculate the positions of the members of d when centered at this centroid, which we'll call d'. If we let n be the number of points in the constellation d (the same number as in t and v), and x_i denote the ith member of some constellation x, this becomes:

$$\bar{d} = \frac{\sum_{i=0}^{n} d_i}{n}$$
$$d'_i = d_i - \bar{d}$$

Assemble the constellation d' into a matrix, D, with each position in the constellation as a row in the matrix. Calculate the covariance matrix of this D, and take the eigenvalues and eigenvectors of the covariance matrix. The eigenvector corresponding to the smallest eigenvalue can be taken as the normal vector of a plane fit to the tilt described by d'. Call this $\vec{n}$; store it in a normalized form, $\hat{n}$.

Given this normal, a correction for tilt can be calculated for any x-z position in tracked space. The error on y introduced by Tilt is the y distance between the plane described by this normal and the x-z plane at some given x-z point. This can be calculated as the collision between the plane and a line l situated at the x-z point and aligned to the y axis.

One additional adjustment is needed - a simple adjustment with a less simple explanation. Imagine a set of tracked points directly on the physical ground. Assume the virtual ground plane is at a height of $y = 0$. The untilt correction can be imagined to rotate the tilted plane around the x- and z-axes to make it flat. Assuming perfectly flat physical ground and no other sources of tracking error, all points that are physically on the ground plane will have the same height, the same y position, in the tracked coordinate space once tilt is corrected; however, there's no guarantee that this height will be at 0. Rather, the height of the resulting plane will match that of the pivot point about which the tilted plane is rotated. An untilt correction must include a constant offset on y that is the height of this pivot. For the sake of this work, the first point in the tracked constellation is chosen as the pivot. Assuming no sources of error in tracked positions other than tilt, this choice is somewhat arbitrary; it may be that using the centroid of d may more evenly distribute error from sources other than tilt, or that using the origin of the tracked space may simplify calculations. Refinement of the choice of pivot is a possible avenue for future work.

Calculating a tilt correction while taking the pivot into account can be done in two phases: recording values related to the pivot one time when the alignment and degree of tilt is initially calculated, then an adjustment applied to the y position of tracked devices, calculated every frame. The position of the pivot should be recorded when the tilt is initially calculated. Here I choose the first tracked device as the pivot, and record its x, z, and difference in y between this tracked device's position and its corresponding virtual position counterpart; this is the first member of d described earlier, which we'll here denote as d^0. The y component is taken as a constant offset to be added to all untilt adjustments. The x-z components are taken as the pivot point p, the center of the untilt space - for the sake of determining untilt adjustments, all other x-z positions should have the pivot's x-z subtracted. Each frame, a given tracked device's position $\vec{a}$ is first cast to the x-z plane, then these coordinates are centered at the pivot.

$$\begin{aligned}
\vec{p} &= \{d^0_x, 0, d^0_z\} \\
\vec{a}'' &= \{\vec{a}_x, 0, \vec{a}_y\} \\
\vec{a}' &= \vec{a}'' - \vec{p}
\end{aligned}$$

A line-plane collision can then be used to find the distance between the tilted plane and the untilted ground plane. The collision is calculated between a line positioned at $\vec{a}'_i$ and facing in the direction $\vec{y}$. This distance, minus the y component of our pivot, is our untilt correction, the scalar u:

$$\vec{y} = \{0, 1, 0\}$$
$$u = \frac{-\hat{n} \cdot \vec{a}'}{\hat{n} \cdot \vec{y}} - d_y^0$$

This untilt correction u is added to the tracked device's y position to yield the corrected tracked object position:

$$\vec{a}^{corrected} = \vec{a} + \{0, u, 0\}$$

It should be emphasized that this correction must be applied to the position of all tracked objects, either every time the tracking system generates new positions, or immediately before positions are used (e.g. for rendering, for interaction, or for physics calculations). A computationally simpler version of this application step has been explored where the line-plane intersection is replaced by pre-computing the slope of the plane on x and z, but current tests were conducted using the method as described above.

Preliminary Evaluation

Preliminary examinations of this method have been conducted. An HTC Vive Pro HMD was used, and candidate reference objects included either three Vive 2.0 base stations, or three Vive 3.0 trackers. The real space locations of the base stations and pucks were measured relative to the walls of the room they were installed in, using a hand-held laser measure. Pilot trials suggested that changes in tracked base station locations indicated that the tracking system had been been reinitialized, and a new alignment and tilt correction would be needed; this is used to automate the process of acquiring new alignments. In a first environment only lighthouses were tested. 8 different alignments were observed, 4 triggered by inducing a loss of tracking by occluding the HMD behind a wall, and 4 triggered by restarting the SteamVR program. The distance of each alignment from the centroid of all alignments was calculated, showing a mean distance of 1.46cm (SD 1.31). Several confederates confirmed that when viewing the scene, the virtual, real, and tracked spaces did appear to be well aligned. In a second environment, lighthouse positions were consistently misreported by the tracking system, leading to increased error in alignment and particularly in untilt. This led to investigating the feasibility of using Vive trackers as reference objects. Early tests indicate favorable results.

Discussion

Some variance is seen between alignments made using Method 2 with base stations, though it may be within tolerable levels for many experiments. The variance may be due to error in the reported positions of the base stations, imprecision when measuring the physical positions of the base stations, or mischaracterization of the tilt artifact. In the first case, this may simply be the level of precision to be expected from this consumer-grade tracking system. We might better estimate the actual position of tracked objects by taking some estimate informed by several different tracked positions, though in the case of base stations this is complicated by new positions only appearing when the tracking system is reinitialized; acquiring multiple base stations' positions requires time and human intervention enough to be impractical. In our second environment, Vive trackers appeared to report position more consistently than base stations; they still allow for human error measuring real world positions at which to place the pucks, and while replacing the pucks if initial placement isn't permanent. In both environments, monitoring for changes in lighthouse position was the most dependable way to detect tracking loss and re-acquisition.

In the case that the physical space positions of the reference objects are imperfectly measured, more precise measurement methods might be employed, as in Luckett et al. (2019). If the tracked positions of the reference objects can be established as being accurate or at least consistent relative to one another, they might be used to inform the measurements of their physical counterparts: errors in the positions of the tracked and virtual reference points that cause the two constellations to be of different shapes will manifest as errors in the alignment and untilt procedure. There may be methods that fuse multiple sources of measurement to achieve better overall accuracy, either different classes of tracked object, or entirely different tracking systems. It might be possible to combine data from different consumer-grade tracking systems (e.g. SteamVR and cellphone AR) to achieve better accuracy, though the complexity of such a piecemeal system may quickly outweigh the cost of simply using a different, more accurate off-the-shelf tracking system.

If Tilt is seen to be a source of error even after the corrections suggested here, further investigation into the nature of tilt may be needed to better inform the assumptions that drive our untilt technique.

8.4 Chapter Conclusions

This chapter first proposes a method using 3 Vive trackers to address the general need for real and virtual space alignment, and account for the Tilt errors described by Niehorster et al. (2017); a proof-of-concept simulation shows that this method largely accounts for the observed error. Some error remained, ranging over 2cm; this may have been unavoidable noise in the Vive tracking system, with a second concern being that our three points didn't fully characterize the error, and additional trackers distributed more widely through the space might be needed. It should also be noted that directly applying the alignment matrix as described in this first method treats tilt as if it were a rotation, though it seems closer to a shear.

A second method using any matched pairs of at least 3 points in tracked and real space is also discussed. This method is more flexible in how tracked devices can be positioned, possibly avoiding further instrumentation of the space if base station positions are reported stably. The first method depends on how accurately three tracked devices can be rigidly aligned in a right angle on a flat plane, and requires fairly simple math; the second allows for an arbitrary constellation of reference points, depends on how accurately the position of real-world reference points can be measured, and requires slightly more computation.

It should be noted that the nature of the tilt error is still underexplored. For instance, it appears to impact only height (y) and not ground plane position (x and z); if further investigation were to show this assumption doesn't hold, the methods described here would not provide a correction. It is also unclear how the error affects users: anecdotal evidence suggests it goes largely unobserved, which is surprising. The degree of error is also highly variable, with the uncorrected case presented here in the discussion of Method 1 being chosen from the several recorded due to showing enough error that the correction was obvious. This amount of error is not atypical, but also not constant; the error seems to change every time the Vive headset acquires tracking. Though preliminary exploration of our second method shows a promising amount of stability in corrected positions across a variety of tilt conditions, more work is needed to quantify the expected behavior of this error, as well as how it is experienced by users.

9 DISCUSSION

This dissertation describes efforts exploring distance mispercepetion in virtual reality - to mitigate it, and to better understand it directly and through facilitating further research on the topic. In this chapter, I discuss the overall insights drawn from the work discussed in previous chapters, as well as the limitations of this work, and possible directions for future work.

9.1 Reflection and Limitations

Chapter 5 and Chapter 4 represent early explorations into the behavior of distance misperception as well as the methods used to measure it. Results from the user studies described in these chapters suggested that individuals experience different degrees of distance misperception, which influenced the rest of my work. As noted in Chapter 4, this might not be due to differences in individuals, such as visual or stereo acuity, but rather differences in fit of the HMD resulting in misalignment between the viewer's eyes and the HMD's lenses, or deviations of an individual HMD from manufacturer specifications Kuhl et al. (2009). Tracking artifacts, such as the Tilt described in Chapter 8, might also introduce per-session variability in the spatial cues presented. Another possibility is that our participant samples contained outliers relative to the larger population, that the sub-population we drew participants from is somehow biased towards these outliers. In all of these cases, the personalized perceptual space warp presented in Chapters 5 and 6 would still be viable, though the calibration may need to be repeated on a per session or per headset level rather than per person. The visual acuity and stereo accuracy measures suggested in Chapter 7 would have less strong motivation if all variability seen between participants were due to variance in HMD fit, but would still have some support elsewhere in the literature (Armbrüster et al., 2008).

The perceptual space warp technique presented in Chapter 5 is, in Chapter 6, combined with this insight about the importance of individual difference to provide a flexible method for mitigating distance misperceptions. User studies described in those chapters provide some evidence that perceptual space warp does not cause noticeable perceptual artifacts and is not a detriment to task performance in two other perceptual tasks. However, this evidence was gathered primarily in visually sparse environments, and with the chief exposure to the environment being blind tasks that require very little movement through the space. There may be other visual stimuli or tasks that would show some detrimental effects, or, something like the transfer effects seen by Richardson and Waller (2005, 2007) and Waller et al. (1998), it may take time for a participant to adapt to a warp enough that any possible transfer effects become evident. Perceptual space warp is introduced here as a novel technique, and as such many aspects of its potential use merit further investigation.

Chapters 7 and 8 both provide techniques that could facilitate future research, but only preliminary results as to their efficacy. Chapter 7's stereo acuity measure in particular lacks an accessible non-VR counterpart. Both tools presented in that chapter might benefit from adopting adaptive psychophysical methods to reduce total number of trials, and perhaps a better exploration of the space; for the stereo acuity measure, this or an iterative application of the method of constant stimuli seems necessary. For the alignment and untilt method presented in Chapter 8, further data could be gathered on the variability of alignments and tilt corrections seen over some larger number of acquisitions to better describe the expected performance of the technique, or to more fully establish the nature of the tracking artifact. As the nature of tilt is relatively unexplored, it is unclear whether aspects of a particular capture environment – the positions and orientations of the tracking system's base stations, the reflectivity of nearby surfaces, sources and intensity of lighting – are an influence, and so how well any gathered data may generalize. For the techniques of both Chapters 7 and 8, the best test of whether these techniques generalize to other research contexts would be to provide implementations to the larger research community, for testing in a wider variety of contexts.

Chapter 4 presents relative percent error, a method of defining distance misperception as the difference between an individual's perceived distance estimates made when viewing real and virtual target stimuli. This method significantly reduced the overall degree of distance misperception seen in the user study described in that chapter. It also reduced the difference seen between different methods of measuring perceived distance, though the differences remained statically significant. Having a method of reporting degree of distance misperception observed that affords comparisons across studies using different methods of measuring perceived distance seems valuable. Relative percent error is one candidate. The user study of Chapter 4 does show statistically significant differences between different measurement methods after conversion to RPE, but these methods were still more comparable than when described using raw percent error. It may be that describing error as a ratio of target distance is the wrong choice, that distance misperception may have a logarithmic, nonmonotonic, or otherwise more complex relationship to distance; it may be that some other transformation of the data better affords comparisons. Whether through refinement of RPE or exploration of other methods, I think it important that the research community establish better ways to afford cross-experiment comparisons.

9.2 Open Questions and Future Work

The limitations described in the previous section imply some future work, but a number of other research directions exist in regards to mitigating and understanding distance misperceptions. In this section, the set of questions that most directly follow from the work presented in this dissertation are discussed.

Further Exploration of Perceptual Space Warp

The technique of perceptual space warp introduced in Chapters 5 and 6 shows promise as a novel means of manipulating or mitigating distance misperceptions. As PSW is a new technique, there is ample room for future work to explore its behavior in a larger set of contexts than those observed here. One concern is that PSW may have some detrimental effect. Though user studies have provided some evidence that PSW does not cause meaningful distortions during blind throwing tasks or the two perceptual matching tasks of Chapter 6, a wider variety of tasks and virtual environments might be explored.

In Chapters 4 and 6, there was some evidence that error in perceived distance might not be linear across distances. There may also be instances where it is desirable that only specific distances be under the effects of PSW, maybe for artistic effect, or to avoid distortions at far distances while affording accurate interaction at arms' reach. PSW should be adaptable to warps more complex than a simple multiplier applied to all objects in a scene, though user studies would be needed to establish how well viewers tolerate different variants. The results of Chapter 4 showed different methods of measuring perceived distance to indicate different degrees of distance misperception. One reason to use these different methods of measurement is to accommodate different ranges of distances – one cannot throw a beanbag as far as one can imagine walking. Exploring a warp personalized using different measurement methods for different distance ranges might yield a warp better tuned for tasks that parallel the measurement methods used, for example a blind reaching task might be better suited to distances near arm's length than a blind throwing task.

Further work might explore the possibility of these more complicated perceptual space warps.

The Larger Impact of Distance Misperceptions

If the efficacy of PSW were to be established further, it might allow exploration of one of the largest open question in distance misperception research: when, outside of the lab, do distance misperceptions matter? The existence of distance misperception seems somewhat unbelievable to many who've experienced VR, partly because we are not generally consciously aware of it happening, but also because it doesn't always happen. If you were to perform a blind walking task while wearing an HMD, you would be expected to walk short due to distance misperception. When walking towards a target you can see, you would always be expected to stop at the target due to continuous visual updates – you can see whether the target is currently below your feet. That is to say, blind walking in a virtual environment consistently shows significant misperception, while walking with full sight shows none.

Though it might be argued that walking without a blindfold is the more general case and the blind visually directed action tasks used to detect distance misperception must be quite rare outside of controlled experiments in a lab, this hasn't been explored in the distance misperception literature. There may be tasks that exist somewhere in the middle, and there may be times when we perform relatively blind judgements as pieces of other tasks; there may be times when distance misperception surprises us, or goes undetected. This merits further study. Personalized perceptual space warp, as explored in Chapters 5 and 6 could be used to create conditions where distance misperception has or has not been mitigated – roughly, conditions with distance misperception, and without.

Comparing participant behavior under conditions with and without distance misperception might reveal what impact distance misperception has. It may have detrimental effects on task performance, inhibits sense of presence, increase cognitive load, induce fatigue, or correlate with simulator sickness. It may also have no observable effect outside of blind directed action tasks. Establishing the larger impacts of distance misperception would help inform when mitigations might be important to employ. This would allow taking into consideration the risk of introducing distortions through software adjustments like PSW, or whether vari-focal displays to remove the accomodation-vergance conflict from HMDs are worth the cost of development or manufacturing.

Work by Richardson and Waller establishes that participants can adapt to distance misperceptions, mitigating them in one context but possibly transferring them to others: they show that adapting to underestimations in distance judgments from the viewer to a target object (an egocentric distance) caused overestimations when judging the distance between two objects (an allocentric distance) (Richardson and Waller, 2005); they also show that adapting to underestimation in a virtual environment can lead to overestimation in a real environment (Waller and Richardson, 2008). If similar experiments using a personalized warp in the place of adaptation might show no transfer effects, this would provide evidence that PSW as a mitigation strategy generalizes across tasks. The most impactful immediate next step might be to replicate Richardson and Waller's virtual-to-real transfer experiment, but with the addition of a PSW-corrected virtual condition; if virtual-to-real transfer is not seen in corrected conditions, but is seen in uncorrected conditions, it would suggest that PSW could be used to remove distance misperception effects when using virtual environments to train for real-world tasks.

Establishing Common Methods

The methods of measuring visual acuity and stereo accuracy presented in Chapter 7 are intended as the first part of a larger battery of tests to quantify potential sources of the between-participant variance in distance misperception seen in the results of the user studies described in Chapters 4, 5, and 6. Chapter 8's method of aligning virtual and real spaces is meant to improve the accuracy of the kinds spatial stimuli and measurements used in distance misperception research. Chapter 3 suggests several ways in which future research might report results to better afford comparisons between experiments, including development of a standardized set of stimuli; Chapter 4 suggests using RPE to better capture an individual's misperceptions, and possibly to better afford comparisons between experiments.

The common theme of all of these contributions is that of creating a common set of tools for future distance misperception researchers that afford more comparable research. Other efforts have been made to establish a standard battery of tests for aspects of immersive display use (Lampton et al., 1994; Kirkley Jr, 2003; Bowman et al., 2001; Fidopiastis, 2006), though none see wide use today. I hope that providing the tools I've developed as source code will allow them to be adopted by the larger research community, with the larger hope that future work might establish a more complete standardized set of tools for perceptual research in immersive displays.

10 CONCLUSION

At a high level, this dissertation suggests that individual differences between participants have a meaningful influence on distance misperception, provides methods to measure and mitigate perceived distance in ways that account for individual differences, and suggests individual differences such as visual and stereo acuity while viewing VR should measured alongside future experiments, while providing tools to perform these measurements, and otherwise facilitate accurate experiments. More specifically, this dissertation describes a number of contributions to the study of distance misperception in virtual reality - practical, empirical, and methodological.

10.1 Practical Contributions

This dissertation provides a novel technique for manipulating perceived distance in virtual reality, called here *perceptual space warp*. This method can be calibrated to mitigate an individual's degree of misperception, and is flexible enough to be further extended to discontinuous and nonlinear misperception spaces, such as might occur at distances near the accommodative distance of a head-mounted display's lenses.

This work provides an implementation of tools for measuring visual acuity and stereo accuracy as mediated by an immersive display, in hopes that measuring these potential sources of individual difference in error of perceived distance might become more standardized and commonplace in future work.

This work also presents methods for aligning virtual and real scenes in general while also correcting for a tracking artifact seen in a consumer-grade tracking system, in hopes that research that depends on accurate measurement of distance and alignment of spaces can be made as accessible as consumer hardware without sacrificing rigorous accuracy.

10.2 Empirical Contributions

Across several user studies exploring distance misperception research method in general, as well as the novel tools and techniques proposed in this work, empirical data was gathered that provides evidence for several contributions of note.

Chapter 3 provides some evidence that distance misperception has not been reduced over time, or as HMD properties such as FOV, weight, or resolution have changed.

In Chapters 4, 5 and 6, evidence is provided that indicates individuals experience different degrees of distance misperception. Chapter 4 also presents evidence suggesting that the results of different perceived distance measures cannot be directly compared, even for a single individual, though they are more comparable when error in perceived distance is phrased as a difference between real and virtual behavior.

Furthermore, in Chapter 6 evidence is provided suggesting that a perceptual space warp calibrated to an individual's degree of misperception can mitigate the misperception. Provided evidence also suggests that perceptual space warp does not introduce noticeable distortions or otherwise induce meaningful disruption of other perceptual tasks.

Evidence from Chapter 7 shows visual acuity to be mediated by display resolution. Preliminary evidence that visual acuity is mediated by display resolution differently by individual is provided, as well as preliminary evidence that stereo accuracy is mediated by head mounted displays differently by individual.

Evidence is provided in Chapter 8 that describes an artifact of SteamVR tracking which causes the y position of tracked objects to be incorrectly reported, as if anchored to a tilted ground plane. Further, preliminary evidence suggests that this SteamVR Tilt artifact can be corrected using methods use three tracked objects in a specific rigid alignment, or matched constellations of points in real and tracked space.

10.3 Methodological Contributions

Distance misperceptions in VR have been studied since at least 1995 (Wright, 1995), time enough that a plurality of technique and technology has been used to explore the problem. This variety makes it difficult to confidently compare results between experiments; it would be beneficial to standardize experimental methods used to study distance misperceptions going forward, to more broadly share our experimental data, and to standardize or share our visual stimuli.

Further, individual differences have largely been ignored as a potential influence of distance misperceptions. To both better compare results between experiments and better account for individual differences, this work proposes a method by which error in perceived distance can be calculated as the difference between percent error while viewing real and virtual stimuli, here called relative percent error.

I also propose that more individual differences, such as visual and stereo acuity, be measured as part of a battery of pre-tests in hopes of quantifying which measurable differences between individuals drive their differences in distance misperceptions.

10.4 Closing Remarks

Virtual reality is taking its first tentative steps to becoming a part of our everyday experiences, something you can find on store shelves alongside TVs and toasters. As VR is experienced by more people than ever before, it is important that it provide dependable stimuli – the immersive aspect of VR implicitly means that viewers trust their senses to VR displays and the authors of VR software, to a level not seen with 2D displays. When VR presents perceptual artifacts like distance misperceptions, it violates this trust in ways that could have far reaching effects on users' evaluation of the utility of VR. Distance misperception is a subtle effect, such that users may not be aware anything is wrong until an important error is made. In an immediate sense, these errors could be inconvenient (e.g. errors in an architectural model), or dangerous (e.g. incorrectly learning to fly a plane). If VR gains a reputation for being subtly undependable in its ability to utilise human spatial sense, user trust in one of VRs chief strengths would be damaged in a way that we may find very difficult to repair.

Truly eliminating distance misperception will likely not happen until we have a full understanding of the effect, and hardware advances to afford construction of more complete illusions. This dissertation presented a method to mitigate distance misperception that may find use in the interim, or over the longer term in cases where cost of hardware or CPU cycles doesn't allow for other solutions. It also presented data, techniques and tools that may help future researchers to accomplish that fuller understanding of distance misperception, and steer those hardware advances towards the hardware aspects of most impact.

Alexandrova, Ivelina V, Paolina T Teneva, Stephan De La Rosa, Uwe Kloos, Heinrich H Bülthoff, and Betty J Mohler. 2010. Egocentric distance judgments in a large screen display immersive virtual environment. In *Proceedings of the 7th symposium on applied perception in graphics and visualization*, 57–60. ACM.

Altenhoff, Bliss M, Phillip E Napieralski, Lindsay O Long, Jeffrey W Bertrand, Christopher C Pagano, Sabarish V Babu, and Timothy A Davis. 2012. Effects of calibration to visual and haptic feedback on near-field depth perception in an immersive virtual environment. In *Proceedings of the acm symposium on applied perception*, 71–78. ACM.

Andre, Jeffrey, and Sheena Rogers. 2006. Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Perception & Psychophysics* 68(3):353–361.

Armbrüster, Claudia, Marc Wolter, Torsten Kuhlen, Will Spijkers, and Bruno Fimm. 2008. Depth perception in virtual reality: distance estimations in peri-and extrapersonal space. *Cyberpsychology & Behavior* 11(1): 9–15.

Atchison, David A. 1995. Accommodation and presbyopia. *Ophthalmic and Physiological Optics* 15(4):255–272.

Bach, Michael. 1996. The freiburg visual acuity test-automatic measurement of visual acuity. *Optometry & Vision Science* 73(1):49–53.

———. 2006. The freiburg visual acuity test-variability unchanged by post-hoc re-analysis. *Graefe's Archive for Clinical and Experimental Ophthalmology* 245(7):965–971.

———. 2019a. Freiburg vision test ('fract'). `https://michaelbach.de/fract/`.

———. 2019b. Visual acuity "cheat sheet". `https://michaelbach.de/sci/acuity.html`.

Bach, Michael, Christina Schmitt, Miriam Kromeier, and Guntram Kommerell. 2001. The freiburg stereoacuity test: automatic measurement of stereo threshold. *Graefe's archive for clinical and experimental ophthalmology* 239(8):562–566.

Bergmann, Johanna, Elsa Krauß, Agnes Münch, Reiner Jungmann, Daniel Oberfeld, and Heiko Hecht. 2011. Locomotor and verbal distance judgments in action and vista space. *Experimental brain research* 210(1):13–23.

Bingham, Geoffrey P, Arthur Bradley, Michael Bailey, and Roy Vinner. 2001. Accommodation, occlusion, and disparity matching are used to guide reaching: a comparison of actual versus virtual environments. *Journal of experimental psychology: human perception and performance* 27(6):1314.

Bodenheimer, Bobby, Jingjing Meng, Haojie Wu, Gayathri Narasimham, Bjoern Rump, Timothy P McNamara, Thomas H Carr, and John J Rieser. 2007. Distance estimation in virtual and real environments using bisection. In *Proceedings of the 4th symposium on applied perception in graphics and visualization*, 35–40. ACM.

Bowman, Doug A, Donald B Johnson, and Larry F Hodges. 2001. Testbed evaluation of virtual environment interaction techniques. *Presence: Teleoperators & Virtual Environments* 10(1):75–95.

Bridgeman, Bruce, Andrea Gemmer, Trish Forsman, and Valerie Huemer. 2000. Processing spatial information in the sensorimotor branch of the visual system. *Vision research* 40(25):3539–3552.

Bruder, Gerd, Ferran Argelaguet, Anne-Héléne Olivier, and Anatole Lécuyer. 2016. Cave size matters: Effects of screen distance and parallax on distance estimation in large immersive display setups. *PRESENCE: Teleoperators and Virtual Environments* (00).

Bruder, Gerd, Andreas Pusch, and Frank Steinicke. 2012. Analyzing effects of geometric rendering parameters on size and distance estimation in on-axis stereographics. In *Proceedings of the acm symposium on applied perception*, 111–118. ACM.

Bruder, Gerd, Fernando Argelaguet Sanz, Anne-Hélène Olivier, and Anatole Lécuyer. 2015. Distance estimation in large immersive projection systems, revisited. In *2015 ieee virtual reality (vr)*, 27–32. IEEE.

Bruder, Gerd, Frank Steinicke, Kai Rothaus, and Klaus Hinrichs. 2009. Enhancing presence in head-mounted display environments by visual body feedback using head-mounted cameras. In *Cyberworlds, 2009. cw'09. international conference on*, 43–50. IEEE.

Buck, Lauren E, Mary K Young, and Bobby Bodenheimer. 2018. A comparison of distance estimation in hmd-based virtual environments with different hmd-based conditions. *ACM Transactions on Applied Perception (TAP)* 15(3):1–15.

Combe, Emmanuelle, Javier Posselt, and Andras Kemeny. 2008. Virtual prototype visualization: a size perception study. In *Proceedings of the 5th nordic conference on human-computer interaction: building bridges*, 581–582. ACM.

Creem-Regehr, Sarah H, and Benjamin R Kunz. 2010. Perception and action. *Wiley Interdisciplinary Reviews: Cognitive Science* 1(6):800–810.

Creem-Regehr, Sarah H, Jeanine K Stefanucci, William B Thompson, Nathan Nash, and Michael McCardell. 2015. Egocentric distance perception in the oculus rift (dk2). In *Proceedings of the acm siggraph symposium on applied perception*, 47–50. ACM.

Creem-Regehr, Sarah H, Peter Willemsen, Amy A Gooch, and William B Thompson. 2005. The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual indoor environments. *Perception* 34(2):191–204.

CuQlock-Knopp, V Grayson, Kimberly P Myles, Frank J Malkin, and Edward Bender. 2001. The effects of viewpoint offsets of night vision goggles on human performance in a simulated grenade-throwing task. Tech. Rep., DTIC Document.

Cutting, James E, and Peter M Vishton. 1995. Perception of space and motion, chapter perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. *Academic Pr* 46:69–117.

Dodgson, Neil A. 2004. Variation and extrema of human interpupillary distance. In *Stereoscopic displays and virtual reality systems xi*, vol. 5291, 36–46. International Society for Optics and Photonics.

Drascic, David, and Paul Milgram. 1996. Perceptual issues in augmented reality. In *Electronic imaging: Science & technology*, 123–134. International Society for Optics and Photonics.

Durgin, Frank H, Laura F Fox, Jed Lewis, and KA Walley. 2002. Perceptuomotor adaptation: More than meets the eye. *Psychonomic Society* 7: 103–104.

Ebrahimi, Elham, Bliss M Altenhoff, Christopher C Pagano, and Sabarish V Babu. 2015. Carryover effects of calibration to visual and proprioceptive information on near field distance judgments in 3d user interaction. In *3d user interfaces (3dui), 2015 ieee symposium on*, 97–104. IEEE.

Feldstein, Ilja T, Felix M Kölsch, and Robert Konrad. 2020. Egocentric distance perception: A comparative study investigating differences between real and virtual environments. *Perception* 49(9):940–967.

Ferris, Steven H. 1972. Motion parallax and absolute distance. *Journal of experimental psychology* 95(2):258.

Fidopiastis, Cali. 2006. User-centered virtual environment assessment and design for cognitive rehabilitation applications.

Fidopiastis, Cali, Christopher Fuhrman, Catherine Meyer, and Jannick Rolland. 2005. Methodology for the iterative evaluation of prototype head-mounted displays in virtual environments: Visual acuity metrics. *Presence* 14(5):550–562.

Frisby, John P, David Buckley, and Philip A Duke. 1996. Evidence for good recovery of lengths of real objects seen with natural stereo viewing. *Perception* 25(2):129–154.

Fukusima, Sergio S, Jack M Loomis, and José A Da Silva. 1997. Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology: Human Perception and Performance* 23(1):86.

Geuss, Michael, Jeanine Stefanucci, Sarah Creem-Regehr, and William B Thompson. 2010. Can i pass?: using affordances to measure perceived size in virtual environments. In *Proceedings of the 7th symposium on applied perception in graphics and visualization*, 61–64. ACM.

Geuss, Michael N, Jeanine K Stefanucci, Sarah H Creem-Regehr, and William B Thompson. 2012. Effect of viewing plane on perceived distances in real and virtual environments. *Journal of Experimental Psychology: Human Perception and Performance* 38(5):1242.

Gibson, James J, and Eleanor J Gibson. 1955. Perceptual learning: Differentiation or enrichment? *Psychological review* 62(1):32.

Gogel, Walter C, and Jerome D Tietz. 1979. A comparison of oculomotor and motion parallax cues of egocentric distance. *Vision Research* 19(10): 1161–1170.

Goodale, Melvyn A, and A David Milner. 1992. Separate visual pathways for perception and action. *Trends in neurosciences* 15(1):20–25.

Grechkin, Timofey Y, Tien Dat Nguyen, Jodie M Plumert, James F Cremer, and Joseph K Kearney. 2010. How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Transactions on Applied Perception (TAP)* 7(4):26.

Hoffman, David M, Ahna R Girshick, Kurt Akeley, and Martin S Banks. 2008. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of vision* 8(3):33–33.

Howard, Harvey J. 1919. A test for the judgment of distance. *Transactions of the American Ophthalmological Society* 17:195.

Interrante, Victoria, Brian Ries, and Lee Anderson. 2006. Distance perception in immersive virtual environments, revisited. In *Ieee virtual reality conference (vr 2006)*, 3–10. IEEE.

Interrante, Victoria, Brian Ries, Jason Lindquist, Michael Kaeding, and Lee Anderson. 2008. Elucidating factors that can facilitate veridical spatial perception in immersive virtual environments. *Presence: Teleoperators and Virtual Environments* 17(2):176–198.

Jones, J Adam, Darlene Edewaard, Richard A Tyrrell, and Larry F Hodges. 2016. A schematic eye for virtual environments. In *2016 ieee symposium on 3d user interfaces (3dui)*, 221–230. IEEE.

Jones, J Adam, J Edward Swan II, and Mark Bolas. 2013. Peripheral stimulation and its effect on perceived spatial scale in virtual environments. *IEEE transactions on visualization and computer graphics* 19(4):701–710.

Jones, J Adam, J Edward Swan II, Gurjot Singh, and Stephen R Ellis. 2011. Peripheral visual information and its effect on distance judgments in virtual and augmented environments. In *Proceedings of the acm siggraph symposium on applied perception in graphics and visualization*, 29–36. ACM.

Jones, J Adam, J Edward Swan II, Gurjot Singh, Eric Kolstad, and Stephen R Ellis. 2008. The effects of virtual reality, augmented reality, and motion parallax on egocentric depth perception. In *Proceedings of the 5th symposium on applied perception in graphics and visualization*, 9–14. ACM.

Kabsch, Wolfgang. 1976. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* 32(5):922–923.

Kaernbach, Christian. 2001. Slope bias of psychometric functions derived from adaptive data. *Perception & Psychophysics* 63(8):1389–1398.

Kellner, F., B. Bolte, G. Bruder, U. Rautenberg, F. Steinicke, M. Lappe, and R. Koch. 2012a. Geometric calibration of head-mounted displays and its effects on distance estimation. *Visualization and Computer Graphics, IEEE Transactions on* 18(4):589–596.

Kellner, Falko, Benjamin Bolte, Gerd Bruder, Ulrich Rautenberg, Frank Steinicke, Markus Lappe, and Reinhard Koch. 2012b. Geometric calibration of head-mounted displays and its effects on distance estimation. *IEEE transactions on visualization and computer graphics* 18(4):589–596.

Kelly, Jonathan W, Lucia A Cherep, and Zachary D Siegel. 2017. Perceived space in the htc vive. *ACM Transactions on Applied Perception (TAP)* 15(1): 1–16.

Kelly, Jonathan W, Lisa S Donaldson, Lori A Sjolund, and Jacob B Freiberg. 2013. More than just perception–action recalibration: Walking through a virtual environment causes rescaling of perceived space. *Attention, Perception, & Psychophysics* 75(7):1473–1485.

Kelly, Jonathan W, William W Hammel, Zachary D Siegel, and Lori A Sjolund. 2014. Recalibration of perceived distance in virtual environments occurs rapidly and transfers asymmetrically across scale. *Visualization and Computer Graphics, IEEE Transactions on* 20(4):588–595.

Kenyon, Robert V, Moses Phenany, Daniel Sandin, and Thomas Defanti. 2008. Accommodation and size-constancy of virtual objects. *Annals of biomedical engineering* 36(2):342–348.

Kenyon, Robert V, Daniel Sandin, Randall C Smith, Richard Pawlicki, and Thomas Defanti. 2007. Size-constancy in the cave. *Presence: Teleoperators and Virtual Environments* 16(2):172–187.

Kirkley Jr, Sonny Eugene Harrison. 2003. *Augmented reality performance assessment battery (arpab): Object recognition, distance estimation and size estimation using optical see-through head-worn displays*. Indiana University.

Klein, Eric, Oliver G Staadt, J Edward Swan II, Greg Schmidt, and Mark A Livingston. 2006. Egocentric medium-field distance perception in projection environments. In *Acm siggraph 2006 research posters*, 181. ACM.

Klein, Eric, J Edward Swan, Gregory S Schmidt, Mark A Livingston, and Oliver G Staadt. 2009. Measurement protocols for medium-field distance perception in large-screen immersive displays. In *2009 ieee virtual reality conference*, 107–113. IEEE.

Knapp, Joshua MacIvor. 2001. The visual perception of egocentric distance in virtual environments. Ph.D. thesis, ProQuest Information & Learning.

Kramida, Gregory. 2016. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE transactions on visualization and computer graphics* 22(7):1912–1931.

Kuhl, Scott A, William B Thompson, and Sarah H Creem-Regehr. 2006. Minification influences spatial judgments in virtual environments. In *Proceedings of the 3rd symposium on applied perception in graphics and visualization*, 15–19. ACM.

———. 2009. Hmd calibration and its effects on distance judgments. *ACM Transactions on Applied Perception (TAP)* 6(3):19.

Kunz, Benjamin R, Leah Wouters, Daniel Smith, William B Thompson, and Sarah H Creem-Regehr. 2009. Revisiting the effect of quality of graphics on distance judgments in virtual environments: A comparison of verbal reports and blind walking. *Attention, Perception, & Psychophysics* 71(6):1284–1293.

Lampton, Donald R, Bruce W Knerr, Stephen L Goldberg, James P Bliss, J Michael Moshell, and Brian S Blau. 1994. The virtual environment performance assessment battery (vepab): Development and evaluation. *Presence: Teleoperators & Virtual Environments* 3(2):145–157.

Leyrer, Markus, Sally A Linkenauger, Heinrich H Bülthoff, Uwe Kloos, and Betty Mohler. 2011. The influence of eye height and avatars on egocentric distance estimates in immersive virtual environments. In *Proceedings of the acm siggraph symposium on applied perception in graphics and visualization*, 67–74. ACM.

Li, Bochao, Anthony Nordman, James Walker, and Scott A Kuhl. 2016. The effects of artificially reduced field of view and peripheral frame stimulation on distance judgments in hmds. In *Proceedings of the acm symposium on applied perception*, 53–56. ACM.

Li, Bochao, Ruimin Zhang, and Scott Kuhl. 2014. Minication affects action-based distance judgments in oculus rift hmds. In *Proceedings of the acm symposium on applied perception*, 91–94. ACM.

Li, Bochao, Ruimin Zhang, Anthony Nordman, and Scott A Kuhl. 2015. The effects of minification and display field of view on distance judgments in real and hmd-based environments. In *Proceedings of the acm siggraph symposium on applied perception*, 55–58. ACM.

Lieberman, Harris R, and Alex P Pentland. 1982. Microcomputer-based estimation of psychophysical thresholds: the best pest. *Behavior Research Methods & Instrumentation* 14(1):21–25.

Lin, Qiufeng, Xianshi Xie, Aysu Erdemir, Gayathri Narasimham, Timothy P McNamara, John Rieser, and Bobby Bodenheimer. 2011. Egocentric distance perception in real and hmd-based virtual environments: the effect of limited scanning method. In *Proceedings of the acm siggraph symposium on applied perception in graphics and visualization*, 75–82. ACM.

Loomis, Jack M, and Joshua M Knapp. 2003. Visual perception of egocentric distance in real and virtual environments. *Virtual and adaptive environments* 11:21–46.

Loomis, Jack M, and John W Philbeck. 2008. Measuring spatial perception with spatial updating and action. In *Carnegie symposium on cognition, 2006, pittsburgh, pa, us*. Psychology Press.

Luckett, Ethan, Tykeyah Key, Nathan Newsome, and J Adam Jones. 2019. Metrics for the evaluation of tracking systems for virtual environments. In *2019 ieee conference on virtual reality and 3d user interfaces (vr)*, 1711–1716. IEEE.

Mohler, Betty J, Heinrich H Bülthoff, William B Thompson, and Sarah H Creem-Regehr. 2008. A full-body avatar improves egocentric distance judgments in an immersive virtual environment. In *Proceedings of the 5th symposium on applied perception in graphics and visualization*, 194. ACM.

Mohler, Betty J, Sarah H Creem-Regehr, and William B Thompson. 2006. The influence of feedback on egocentric distance judgments in real and virtual environments. In *Proceedings of the 3rd symposium on applied perception in graphics and visualization*, 9–14. ACM.

Murgia, Alessio, Paul M Sharkey, et al. 2009. Estimation of distances in virtual environments using size constancy. *The International Journal of Virtual Reality* 8(1):67–74.

Naceri, Abdeldjallil, and Ryad Chellali. 2012. The effect of isolated disparity on depth perception in real and virtual environments. In *2012 ieee virtual reality workshops (vrw)*, 107–108. IEEE.

Napieralski, Phillip E, Bliss M Altenhoff, Jeffrey W Bertrand, Lindsay O Long, Sabarish V Babu, Christopher C Pagano, Justin Kern, and Timothy A Davis. 2011. Near-field distance perception in real and virtual environments using both verbal and action responses. *ACM Transactions on Applied Perception (TAP)* 8(3):18.

Nguyen, Tien Dat, Christine J Ziemer, Timofey Grechkin, Benjamin Chihak, Jodie M Plumert, James F Cremer, and Joseph K Kearney. 2011. Effects of scale change on distance perception in virtual environments. *ACM Transactions on Applied Perception (TAP)* 8(4):26.

Niehorster, Diederick C, Li Li, and Markus Lappe. 2017. The accuracy and precision of position and orientation tracking in the htc vive virtual reality system for scientific research. *i-Perception* 8(3):2041669517708205.

Ooi, Teng Leng, Bing Wu, and Zijiang J He. 2001. Distance determined by the angular declination below the horizon. *Nature* 414(6860):197–200.

Parks, Theodore E. 2012. Visual-illusion distance paradoxes: A resolution. *Attention, Perception, & Psychophysics* 74(8):1568–1569.

Patterson, Robert, and Wayne L Martin. 1992. Human stereopsis. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 34(6): 669–692.

Patterson, Robert Earl. 2015. *Human factors of stereoscopic 3d displays*. Springer.

Peer, Alex, and Kevin Ponto. 2016a. Perceptual space warping: Preliminary exploration. In *Ieee virtual reality (vr)*, 261–262.

———. 2016b. Perceptual space warping: Preliminary exploration. In *2016 ieee virtual reality (vr)*, 261–262. IEEE.

———. 2017. Evaluating perceived distance measures in room-scale spaces using consumer-grade head mounted displays. In *2017 ieee symposium on 3d user interfaces (3dui)*, 83–86. IEEE.

———. 2019. Mitigating incorrect perception of distance in virtual reality through personalized rendering manipulation. In *2019 ieee conference on virtual reality and 3d user interfaces (vr)*, 244–250. IEEE.

———. 2020. Measuring visual acuity and stereo accuracy as mediated by immersive displays. In *2020 ieee conference on virtual reality and 3d user interfaces abstracts and workshops (vrw)*, 219–223. IEEE.

———. 2021. Distance misperception in vr: Considering 25 years of research. [Manuscript submitted for publication].

Peer, Alex, Peter Ullrich, and Kevin Ponto. 2018. Vive tracking alignment and correction made easy. In *2018 ieee conference on virtual reality and 3d user interfaces (vr)*, 653–654. IEEE.

Philbeck, John W, Adam J Woods, Joeanna Arthur, and Jennifer Todd. 2008. Progressive locomotor recalibration during blind walking. *Perception & psychophysics* 70(8):1459–1470.

Phillips, Lane, Victoria Interrante, Michael Kaeding, Brian Ries, and Lee Anderson. 2012. Correlations between physiological response, gait, personality, and presence in immersive virtual environments. *Presence: Teleoperators and Virtual Environments* 21(2):119–141.

Phillips, Lane, Brian Ries, Victoria Interrante, Michael Kaeding, and Lee Anderson. 2009. Distance perception in npr immersive virtual environments, revisited. In *Proceedings of the 6th symposium on applied perception in graphics and visualization*, 11–14. ACM.

Phillips, Lane, Brian Ries, Michael Kaeding, and Victoria Interrante. 2010. Avatar self-embodiment enhances distance perception accuracy in non-photorealistic immersive virtual environments. In *2010 ieee virtual reality conference (vr)*, 115–1148. IEEE.

Pinkus, Alan, and H Lee Task. 1998. Measuring observers' visual acuity through night vision goggles. Tech. Rep., AIR FORCE RESEARCH LAB WRIGHT-PATTERSON AFB OH HUMAN EFFECTIVENESS DIRECTORATE.

Plumert, Jodie M, Joseph K Kearney, James F Cremer, and Kara Recker. 2005. Distance perception in real and virtual environments. *ACM Transactions on Applied Perception (TAP)* 2(3):216–233.

Ponto, Kevin, Michael Gleicher, Robert G Radwin, and Hyun Joon Shin. 2013. Perceptual calibration for immersive display environments. *Visualization and Computer Graphics, IEEE Transactions on* 19(4):691–700.

Proffitt, Dennis R, Mukul Bhalla, Rich Gossweiler, and Jonathan Midgett. 1995. Perceiving geographical slant. *Psychonomic bulletin & review* 2(4): 409–428.

Proffitt, Dennis R, Jeanine Stefanucci, Tom Banton, and William Epstein. 2003. The role of effort in perceiving distance. *Psychological Science* 14(2): 106–112.

Rébillat, Marc, Xavier Boutillon, Étienne Corteel, and Brian Katz. 2011. Audio, visual, and audio-visual egocentric distance perception in virtual environments. In *Forum acusticum*, 482.

Renner, Rebekka S, Boris M Velichkovsky, and Jens R Helmert. 2013. The perception of egocentric distances in virtual environments-a review. *ACM Computing Surveys (CSUR)* 46(2):23.

Richardson, Adam R, and David Waller. 2005. The effect of feedback training on distance estimation in virtual environments. *Applied Cognitive Psychology* 19(8):1089–1108.

———. 2007. Interaction with an immersive virtual environment corrects users' distance estimates. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 49(3):507–517.

Riecke, Bernhard E, Pooya Amini Behbahani, and Chris D Shaw. 2009. Display size does not affect egocentric distance perception of naturalistic stimuli. In *Proceedings of the 6th symposium on applied perception in graphics and visualization*, 15–18. ACM.

Ries, Brian, Victoria Interrante, Michael Kaeding, and Lane Phillips. 2009. Analyzing the effect of a virtual avatar's geometric and motion fidelity on ego-centric spatial perception in immersive virtual environments. In *Proceedings of the 16th acm symposium on virtual reality software and technology*, 59–66. ACM.

Rolland, Jannick P, William Gibson, and Dan Ariely. 1995. Towards quantifying depth and size perception in virtual environments. *Presence: Teleoperators & Virtual Environments* 4(1):24–49.

Rolland, Jannick P, Catherine Meyer, K Arthur, and E Rinalducci. 2002. Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. *Presence: Teleoperators & Virtual Environments* 11(6):610–625.

Ryu, Jaeho, Naoki Hashimoto, and Makoto Sato. 2005. Influence of resolution degradation on distance estimation in virtual space displaying static and dynamic image. In *2005 international conference on cyberworlds (cw'05)*, 8–pp. IEEE.

Sahm, Cynthia S, Sarah H Creem-Regehr, William B Thompson, and Peter Willemsen. 2005. Throwing versus walking as indicators of distance perception in similar real and virtual environments. *ACM Transactions on Applied Perception (TAP)* 2(1):35–45.

Schor, Clifton M, and Ivan Wood. 1983. Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision research* 23(12):1649–1654.

Schor, Clifton M, Ivan C Wood, and Jane Ogawa. 1984. Spatial tuning of static and dynamic local stereopsis. *Vision Research* 24(6):573–578.

Siegel, Zachary D, and Jonathan W Kelly. 2017. Walking through a virtual environment improves perceived size within and beyond the walked space. *Attention, Perception, & Psychophysics* 79(1):39–44.

Sinai, Michael J, William K Krebs, Rudy P Darken, JH Rowland, and JS McCarley. 1999. Egocentric distance perception in a virutal environment using a perceptual matching task. In *Proceedings of the human factors and ergonomics society annual meeting*, vol. 43, 1256–1260. SAGE Publications.

Sinai, Michael J, Teng Leng Ooi, and Zijiang J He. 1998. Terrain influences the accurate judgement of distance. *Nature* 395(6701):497–500.

Singh, Gurjot, Stephen R Ellis, and J Edward Swan II. 2018. The effect of focal distance, age, and brightness on near-field augmented reality depth matching. *IEEE transactions on visualization and computer graphics*.

Singh, Gurjot, J Edward Swan II, J Adam Jones, and Stephen R Ellis. 2010. Depth judgment measures and occluding surfaces in near-field augmented reality. In *Proceedings of the 7th symposium on applied perception in graphics and visualization*, 149–156. ACM.

Slater, Mel, and Martin Usoh. 1994. Body centred interaction in immersive virtual environments. *Artificial life and virtual reality* 1:125–148.

Solini, Hannah M, Ayush Bhargava, and Christopher C Pagano. 2019. Transfer of calibration in virtual reality to both real and virtual environments. In *Proceedings of the human factors and ergonomics society annual meeting*, vol. 63, 1943–1947. SAGE Publications Sage CA: Los Angeles, CA.

Stanney, Kay, Ronald Mourant, and Robert Kennedy. 1998. Human factors issues in virtual environments: A review of the literature. *Presence* 7(4): 327–351.

Steinicke, Frank, Gerd Bruder, Klaus Hinrichs, Scott Kuhl, Markus Lappe, and Pete Willemsen. 2009a. Judgment of natural perspective projections in head-mounted display environments. In *Proceedings of the 16th acm symposium on virtual reality software and technology*, 35–42. ACM.

Steinicke, Frank, Gerd Bruder, Klaus Hinrichs, Markus Lappe, Brian Ries, and Victoria Interrante. 2009b. Transitional environments enhance distance perception in immersive virtual reality systems. In *Proceedings of* *the 6th symposium on applied perception in graphics and visualization*, 19–26. ACM.

Steinicke, Frank, Gerd Bruder, Klaus Hinrichs, and Anthony Steed. 2010. Gradual transitions and their effects on presence and distance estimation. *Computers & Graphics* 34(1):26–33.

Steinicke, Frank, Gerd Bruder, and Scott Kuhl. 2011a. Realistic perspective projections for virtual objects and environments. *ACM Transactions on Graphics (TOG)* 30(5):112.

Steinicke, Frank, Gerd Bruder, Scott Kuhl, Pete Willemsen, Markus Lappe, and Klaus Hinrichs. 2011b. Natural perspective projections for head-mounted displays. *IEEE transactions on visualization and computer graphics* 17(7):888–899.

Surdick, R Troy, Elizabeth T Davis, Robert A King, and Larry F Hodges. 1997. The perception of distance in simulated visual displays: A comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. *Presence: Teleoperators and Virtual Environments* 6(5): 513–531.

Swan, II, Adam Jones, Eric Kolstad, Mark A Livingston, Harvey S Smallman, et al. 2007. Egocentric depth judgements in optical, see-through augmented reality. Tech. Rep., DTIC Document.

Swan, J Edward, Gurjot Singh, and Stephen R Ellis. 2015. Matching and reaching depth judgments with real and augmented reality targets. *IEEE transactions on visualization and computer graphics* 21(11):1289–1298.

Thompson, William B, Peter Willemsen, Amy A Gooch, Sarah H Creem-Regehr, Jack M Loomis, and Andrew C Beall. 2004. Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence* 13(5):560–571.

Ukai, K. 2007. Visual fatigue caused by viewing stereoscopic images and mechanism of accommodation. In *Proceedings of the first international symposium on university communication*, vol. 1, 176–179.

Usoh, Martin, Ernest Catena, Sima Arman, and Mel Slater. 2000. Using presence questionnaires in reality. *Presence: Teleoperators & Virtual Environments* 9(5):497–503.

Walker, James, Ruimin Zhang, and Scott A Kuhl. 2012. Minification and gap affordances in head-mounted displays. In *Proceedings of the acm symposium on applied perception*, 124–124. ACM.

Waller, David. 1999. Factors affecting the perception of interobject distances in virtual environments. *Presence: Teleoperators and Virtual Environments* 8(6):657–670.

Waller, David, Earl Hunt, and David Knapp. 1998. The transfer of spatial knowledge in virtual environment training. *Presence* 7(2):129–143.

Waller, David, and Adam R Richardson. 2008. Correcting distance estimates by interacting with immersive virtual environments: Effects of task and available sensory information. *Journal of Experimental Psychology: Applied* 14(1):61.

Wann, John P, and Mark Mon-Williams. 1997. Health issues with virtual reality displays: What we do know and what we don't. *ACM SIGGRAPH Computer Graphics* 31(2):53–57.

Watt, Simon J, Kurt Akeley, Marc O Ernst, and Martin S Banks. 2005. Focus cues affect perceived depth. *Journal of Vision* 5(10):7–7.

Watt, Simon J, and Mark F Bradshaw. 2003. The visual control of reaching and grasping: binocular disparity and motion parallax. *Journal of Experimental Psychology: Human Perception and Performance* 29(2):404.

Willemsen, Peter, Mark B Colton, Sarah H Creem-Regehr, and William B Thompson. 2004. The effects of head-mounted display mechanics on distance judgments in virtual environments. In *Proceedings of the 1st symposium on applied perception in graphics and visualization*, 35–38. ACM.

———. 2009. The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments. *ACM Transactions on Applied Perception (TAP)* 6(2):8.

Willemsen, Peter, and Amy A Gooch. 2002a. Perceived egocentric distances in real, image-based, and traditional virtual environments. In *Virtual reality, 2002. proceedings. ieee*, 275–276. IEEE.

Willemsen, Peter, Amy A Gooch, William B Thompson, and Sarah H Creem-Regehr. 2008. Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoperators and Virtual Environments* 17(1):91–101.

Willemsen, Peter, and Amy Ashurst Gooch. 2002b. An experimental comparison of perceived egocentric distance in real, image-based, and traditional virtual environments using direct walking tasks. In *Proceedings of the ieee virtual reality 2002 (vr'02)*, 1–6. Citeseer.

Witmer, Bob G, and Paul B Kline. 1998. Judging perceived and traversed distance in virtual environments. *Presence* 7(2):144–167.

Witmer, Bob G, and Wallace J Sadowski. 1998. Nonvisually guided locomotion to a previously viewed target in real and virtual environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 40(3):478–488.

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

Wright, Robert H. 1995. Virtual reality psychophysics: Forward and lateral distance, height, and speed perceptions with a wide-angle helmet display. Tech. Rep., DTIC Document.

Wu, Bing, Teng Leng Ooi, and Zijiang J He. 2004. Perceiving distance accurately by a directional process of integrating ground information. *Nature* 428(6978):73–77.

Yang, Zhen, Jinlei Shi, Yi Xiao, Xiaojian Yuan, Duming Wang, Hongting Li, and Weidan Xu. 2020. Influences of experience and visual cues of virtual arm on distance perception. *i-Perception* 11(1):2041669519901134.

Young, Mary K, Graham B Gaylor, Scott M Andrus, and Bobby Bodenheimer. 2014. A comparison of two cost-differentiated virtual reality systems for perception and action tasks. In *Proceedings of the acm symposium on applied perception*, 83–90. ACM.

Zhang, Jun, Xiaohua Jia, and Yuan Zhou. 2012a. Analysis of capacity improvement by directional antennas in wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN)* 9(1):1–25.

Zhang, Ruimin, Anthony Nordman, James Walker, and Scott A Kuhl. 2012b. Minification affects verbal-and action-based distance judgments differently in head-mounted displays. *ACM Transactions on Applied Perception (TAP)* 9(3):14.

Ziemer, Christine J, Jodie M Plumert, James F Cremer, and Joseph K Kearney. 2009. Estimating distance in real and virtual environments: Does order make a difference? *Attention, Perception, & Psychophysics* 71(5):1095–1106.