2021

Blind Direct Walking Distance Judgment Research: A Best Practices Guide

Crystal Massoglia
Michigan Technological University, camassog@mtu.edu

BLIND DIRECT WALKING DISTANCE JUDGMENT RESEARCH: A BEST PRACTICES GUIDE

By

Crystal A. Massoglia

A REPORT

Submitted in partial fulfillment of the requirement for the degree of

MASTER OF SCIENCE

In Computer Science

MICHIGAN TECHNOLOGICAL UNIVERSITY

2021

This report has been approved in partial fulfillment of the requirement for the Degree of MASTER OF SCIENCE in Computer Science.

Department of Computer Science

Report Advisor:	*Dr. Scott Kuhl*
Committee Member:	*Dr. Charles Wallace*
Committee Member:	*Dr. Will Cantrell*
Department Chair:	*Dr. Linda Ott*

To my family.

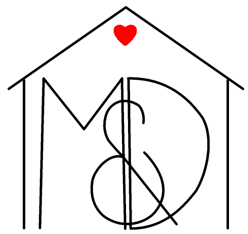


Table of Contents

Contribution

The idea and general outline for my report was collaboratively developed by myself and my advisor, Dr. Kuhl. I surveyed the existing literature on distance judgment research in VR, organized the information, and drafted my report. Dr. Kuhl and I collaboratively edited and improved my report.

Preface

Prior to the COVID-19 pandemic outbreak, this report was originally Distance Judgments on Shallow Slopes in Real and Virtual Environments. For this research topic, I was going to investigate the differences, if any, in how people judge distances on a shallow slope in a real environment and a replica virtual environment. The slope was located on the second floor of the SDC building and had an incline of around three degrees. Unfortunately, the pandemic struck and the school closed part way through this project. Since I planned on graduating in Fall 2020 and the school was not going to reopen or allow in-person human subjects research in the foreseeable future, with the help of my advisor, Dr. Kuhl, I changed my project to Blind Direct Walking Distance Judgment Research: A Best Practices Guide. Here, I present a review of past research pertaining to distance perception, particularly experiments using blind direct walking. Using this and Dr. Kuhl's laboratory procedures as evidence, I wrote up a research procedures guide aimed at new researchers conducting blind direct walking distance judgment experiments.

Abstract

Over the last 30 years, Virtual Reality (VR) research has shown that distance perception in VR is compressed as compared to the real world. The full reason for this is yet unknown. Though many experiments have been run to study the underlying reasons for this compression, often with similar procedures, the experimental details either show significant variation between experiments or go unreported. This makes it difficult to accurately repeat or compare experiments, as well as negatively impacts new researchers trying to learn and follow current best practices. In this paper, we present a review of past research and things that are typically left unreported. Using this and the practices of my advisor as evidence, we suggest a standard to assist researchers in performing quality research pertaining to blind direct walking distance judgments in VR.

1 Introduction

Virtual reality (VR) has proven to be useful for many things, including research, entertainment, training, and engineering. For example, VR may be used to model a true-scale building or mine for future real-world implementation. However, if a supporting structure is placed at an incorrect distance relative to neighboring supporting structures, the real-world building or mine could collapse. Making virtual environments that closely resemble real environments, including the ability to judge distances, is therefore paramount. It is widely known that distances are accurately estimated in real environments, but virtual environments tell a different story.

As nearly 30 years of past work has shown, distances are often compressed in VR. The full reason for this is unknown. Past research has studied many things such as eye height (Leyrer et al., 2011; Leyrer et al., 2015), physical qualities of vision such as peripheral vision (Jones et al., 2011), field of view (FOV) and brightness in the periphery (Li et al., 2016; Li et al., 2018; Li et al., 2015), and how participants are allowed to view the environment (Lin et al., 2011a), the quality or type of the graphics (Grechkin et al., 2010; Kunz et al., 2009; Thompson et al., 2004), the calibration of the head-mounted display (HMD), the image (Kuhl et al., 2006; Kuhl et al., 2009; Li et al., 2014; Li et al., 2015; Zhang et al., 2012), and the horizon line (Messing and Durgin, 2005), the affects of avatars (Leyrer et al., 2011; Lin et al., 2011a; Lin et al., 2011b; Ries et al., 2008), physical properties of the HMD (Buck et al., 2018; Grechkin et al., 2010; Willemsen et al., 2004; Willemsen et al., 2009), environmental context (at least in real environments) (Lappin et al., 2006; Witt et al., 2007) and richness (Nguyen et al., 2011), and the effects of feedback (Mohler et al., 2006). Some of these factors, such as FOV and brightness in the periphery (Li et al., 2016; Li et al., 2018; Li et al., 2015; Lin et al., 2011a), have been shown to significantly affect distance perception in VR but do not account for all of the observed compression.

Other factors, such as the quality of the graphics (Grechkin et al., 2010; Kunz et al., 2009), have not shown a significant effect on the observed compression, at least for blind direct walking. Some factors, such as the physical properties of the HMD, have shown inconclusive results; specifically Willemsen et al., (2004) and Grechkin et al., (2010) have shown that this does significantly affect distance judgments while Willemsen et al., (2009) and Buck et al., (2018) showed otherwise. This research and the conclusions are detailed and summarized, at least up to the year 2013, in the paper by Renner et al., (2013).

Much of this research lacks complete documentation of the procedures, such as hardware calibration processes, the inclusion or exclusion of practice, given instructions, and other procedural and practical details, used during the experiment. Excluding such details, due to their perceived unimportance or inclusion expanding papers beyond page limits, makes it difficult to accurately reproduce or compare experiments. However, if there was an agreed-upon standard, researchers could instead direct people to the standard they followed rather than include detailed procedural sections in publications.

The purpose of this paper is to give guidelines for all of the things that typically go unreported. New researchers to the field may be unaware of some of the details that established research laboratories regularly do but do not explicitly report. In a sense, this is a "best practices guide" that suggests a standard by which to conduct, specifically, blind direct walking experiments. We want to provide a guide so that all researchers may perform quality research.

2 Why Blind Direct Walking?

Blind direct walking, illustrated in Figure 1, is when a participant is shown a target for some amount of time, is blindfolded, and then attempts to walk to where she believes the target was. This distance judgment method has a couple important advantages.

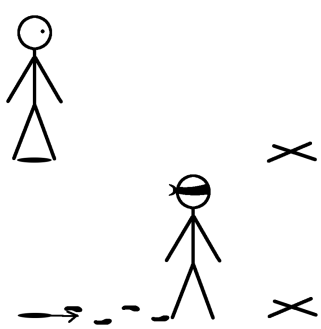


Figure 1: Blind direct walking is when a person looks at a target for some time, puts on a blindfold, and attempts to walk to that target.

First, it is performed in action space, or the space between two and thirty meters from an observer (Cutting and Vishton, 1995). Some distance judgment research has been done in personal space, or the space less than two meters from an observer, and in vista space, the space beyond thirty meters from an observer (Cutting and Vishton, 1995). These different space-types are represented in Figure 2. Action space is the

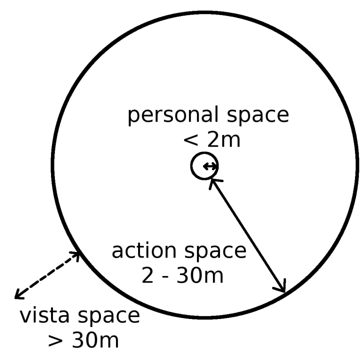


Figure 2: These are the three different types of space coined by Cutting and Vishton, (1995). Personal space is defined as the space less then two meters from a person. Action space is defined as the space between two and 30 meters from a person. Vista space is defined as the space beyond 30 meters from a person.

most commonly used space for distance judgment research, as well as for personal gaming and entertainment purposes, which makes it well documented and relevant to real life. Additionally, much past work has shown that egocentric distance judgments are accurate up to 25 meters in real environments when blind direct walking was employed (Renner et al., 2013), implying that it is a good measure of perception of common real-world distances. Second, Renner et al., (2013) stated that it is the most common and useful type of egocentric distance judgment method as it is easily reproducible and simple to accomplish in HMDs. For these reasons, this paper focuses on blind direct walking.

Despite the advantages blind direct walking has, there are some key disadvantages. For instance, increased walking speed has shown to cause significant inaccuracies in estimations (Renner et al., 2013). Cognitive influences, such as counting steps or calculating the number of reference objects between the participant and the target, may instead measure cognition rather than perception. Further, longer distances are difficult to accomplish due to limited laboratory space, cable length limitations, or other technological or physical limitations, and participants could trip and fall on uneven terrain. Fortunately, there are ways to reduce or eliminate these disadvantages.

Other types of distance judgment methods, which will be briefly discussed next, include verbal reports, blind triangulated walking, timed imagined walking, affordance judgments, blind throwing, reaching, and perceptual matching tasks (Renner et al., 2013). Starting with verbal reports, the participant states aloud the estimate of the distance between himself and the target. Philbeck and Loomis, (1997) showed that, for real environments, both blind direct walking and verbal reports measure distance perception in a proportionate way. However, significant concerns remain about what verbal reports actually measure, the desired perception or the undesired cognition (Renner et al., 2013). Further, people may not fully understand how long the distance unit, such as feet or meters, is and may default to measuring in round numbers. This can be mitigated through the presentation of a reference measurement or encouraging rounding to a specific decimal length; though ideally, the participant should not require any further hints or instructions. As mentioned, blind direct walking is also susceptible to certain cognitive influences, but there are ways to avoid these influences. Thus, verbal reports are useful as an additional test to study perception differences or when space is limited (Renner et al., 2013).

Figure 3 shows blind triangulated walking, where a participant views a target for some amount of time, is blindfolded, walks obliquely to the target until told to stop, then turns and either takes a few steps or points toward the target. The distance is

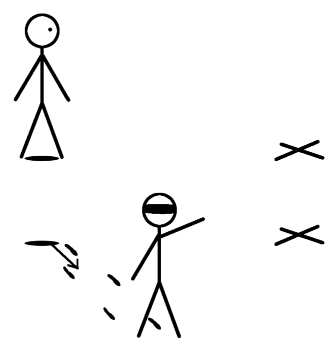


Figure 3: Blind triangulated walking is when a person looks at a target for some time, puts on a blindfold, walks obliquely to the target until told to stop, turns, then either takes a few steps or points toward the target.

calculated via triangle calculations. This method attempts to eliminate the potential cognitive influences found in blind direct walking, can be done in a limited space, and has been shown to be accurate up to 20 meters in real environments (Renner et al., 2013). However, this method produces more variance in measurements than blind direct walking (Renner et al., 2013) since it depends on the participant accurately updating the location of the target relative to the walking path. If incorrectly updated, the resulting slight deviation in desired turning angle would correspond to a large error in distance and not accurately reflect the desired distance indication of the participant.

The remaining methods are timed imagined walking, where a stationary participant starts a stopwatch, imagines walking to a previously seen target, and stops the stopwatch when she imagines that he is standing on the target, blind throwing, where a blindfolded participant throws an object to a previously seen target (Sahm et al., 2005), perceptual matching tasks, where a participant uses a reference object to determine the size of or distance to some target object, reaching tasks, where a participant places her finger or hand where a target was previously shown, and affordance judgments, where a participant determines if an action is possible after being shown a situation. Though these methods may be useful in small spaces, none of them are frequently used (Renner et al., 2013). Further, each has its own key disadvantage. As noted by Renner et al., (2013), some people may be better at imagining environmental updating than others when imagining a walk, meaning that undesirable variables are introduced in timed imagined walking; perceptual matching tasks may be less prone to cognitive influences but are difficult to accomplish in VR since only memorized or virtual reference objects may be used; reaching tasks are accurate in personal space but are impractical in action space; affordance judgments are the least useful of all methods and Renner et al., (2013) recommends against using them.

3 Experimental Design Decisions

3.1 Experimental Conditions

The most obvious and crucial consideration when setting up the experimental design is the conditions of the research. This includes determining what exactly is being tested, if a control or comparison needed, and the best way to accomplish the research. All design and experimental decisions should be based on the conditions of the experiment and what exactly is being tested. Though we intend this paper to help shorten procedure sections, important deviations may require detailed explanations in the final report.

3.2 Design Structure

Experiments often test a control condition and one or more other experimental conditions, implying that participants must be split among the conditions in some unbiased manner. There are three main experimental structures - within-subjects, between-subjects, and a combination of within- and between-subjects techniques, which we will call combo. Each presents significant advantages and disadvantages and is appropriate for certain experimental types.

Charness et al., (2012) and Greenwald, (1976) define the within-subjects design as the design structure in which all participants test under multiple or all conditions of an experiment. They state that the main advantage of this design structure is the statistical power it affords since many data points across all conditions may be captured with fewer total participants. Further, all conditions have the same number of participants, which may aid in analysis and comparison. They also state that this eliminates the need to randomly assign participants to conditions. However, there are some serious disadvantages. Charness et al., (2012) and Greenwald, (1976) state that participants who test many conditions may guess what is being tested for and alter their behavior to reflect what they believe the experimenter desires or deliberately make different choices based on previous trials. They add that some testing orders may show this effect more significantly than other orders. Finally, they mention that previous trials act as practice for subsequent trials, though they acknowledge that allowing sufficient time to pass between conditions may help remedy this. Even so, these disadvantages are severe enough that both Charness et al., (2012) and Greenwald, (1976) recommend against the within-subjects design structure.

Charness et al., (2012) and Greenwald, (1976) define the between-subjects design as the design structure in which each participant completes testing under only a single condition. They state that the main advantage of this design is that it does not expose data to the biases and carry-over effects present in the within-subjects design. However, the between-subjects design structure still has some disadvantages. Charness et al., (2012) state that more subjects are required to improve statistical power since, by nature, this design only captures one data point per subject. They add that this may create noisy data and miss important trends. Further, they mention that this design structure requires that participants be randomly and evenly separated among conditions as to not bias the results. These disadvantages are often minor and easier to deal with than those for within-subjects designs. Due to this coupled with the advantages presented, both Charness et al., (2012) and Greenwald, (1976) recommend using the between-subjects design.

Charness et al., (2012) and Ziemer et al., (2009) define the combo design as the design in which all possible permutations of conditions are generated, and participants complete all conditions in one randomly assigned permutation order. This implies that many data points per participant are possible with fewer subjects and statistical power is high, just as for the within-subjects design, since all participants complete all conditions, and the first condition tested from each permutation constitutes a between-subjects experiment. Charness et al., (2012) recommends using the combo design structure for experiments with a small number of conditions because there will be a low number of total permutations. This design structure is impractical for a large number of conditions because a very large number of participants is required in order to sufficiently test all permutation orders.

For more about each design, as well as an analysis method for the combo design, we refer the reader to Greenwald, (1976), Charness et al., (2012), and Erlebacher, (1977). We generally use the between-subjects design due to the advantages stated above. Ultimately, it is up to the experimenter to consider which design structure will work best for the research project.

4 Experimental Design Components

4.1 Experimental Location

4.1.1 Real

As of the writing of this paper we cannot fully virtualize our bodies, so a real-world environment (RE) is required to complete any experiments. A suitable RE supports action space and any necessary equipment, which will be discussed next, as well as a good initial real position, which will be discussed later. As blind direct walking experiments are completed in action space, distances in some range between two and 30 meters must be supported. Moreover, though distances are compressed in VR, some participants will overestimate when walking (Buck et al., 2018; Grechkin et al., 2010), necessitating extra "safety meters." If no safety meters are allotted, overestimating participants must be physically stopped by the experimenter, which may cause them to become overly cautious for the remaining trials. Due to the cognitive inconsistency, it is most conservative to throw this biased data out. Therefore, if a good space is found but the distance between one well-defined edge of a space to another well-defined edge, denoted wall-to-wall distance, is not long enough to allow for safety meters, try using the distance from one well-defined corner of the space to the opposite well-defined corner, denoted diagonal distance, since the diagonal of a space is longer than the wall-to-wall distance.

Similarly, the space must be wide enough to account for possible veer. Though veer seems to generally be unconcerning (Philbeck and Loomis, 1997), it is safest to take extra precautions to avoid collisions. There are creative solutions to help participants self-correct if the space is not physically wide enough, such as hanging a curtain one meter away from a wall (Philbeck and Loomis, 1997) or increasing the volume of the noise in the ear corresponding to the potential collision side (Grechkin et al., 2010; Nguyen et al., 2008). To our knowledge, there are no studies detailing the consequences of the use of self-correction methods. However, these methods still suggest to the participant that he is walking incorrectly, which may exhibit the same negative side effects detailed above. As such, the more cautious option is to use a space that is physically wide enough and avoid using self-correction methods.

The ideal RE allows for setup and use of any necessary equipment, including HMDs, tracking system parts, computers, measuring devices, and other equipment needed during experiments. Further, since transporting equipment for every experiment is impractical and may damage the equipment, the RE should have an on-site or nearby storage location. Because of these requirements, the most ideal RE in terms of size and reduced distractions may be unsuitable for virtual-world experiments, but it could still be used for real-world experiments. In the case that two different REs are used, it is most conservative to maximize consistency between the spaces to avoid introducing undesirable variables in the experiments.

There may be other factors that impact the choice of RE, including the amount of light entering the space, time of day, temperature, ambient noise, amount of foot or vehicle traffic, mechanical operations or failures, and many other things. All of these factors may distract or perturb participants. Some things, such as window glare, can be easily taken care of. Unfortunately, many of these things, such as the temperature of the room and mechanical operations, are difficult to control. We recommend addressing the problems that can be controlled, but otherwise, there is not much that can be done.

4.1.2 Virtual

The virtual environment (VE) could simply be a modeled replica of the RE. This would automatically make the VE long and wide enough, as well as provide maximum consistency between real-world and virtual-world tests, assuming a real-world experiment will be done. Often, VEs are not replicas of the RE, so the type of VE must be carefully considered. For starters, consider the terrain of the real location, both the texture, such as slope and bumps, and the type, such as grass, gravel, or tiles. Matching the terrain of the VE to that of the RE avoids confusing participants based on foot-feel and possibly enhances immersion. If experimental conditions dictate otherwise, then matching at least the texture may help prevent injury. For example, showing a participant a steep, gravel decline rather than a flat, gravel road may cause the subject to trip if she forgets he is walking on a flat, grassy field.

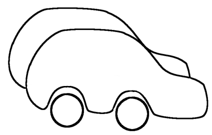
Similarly to the RE, the VE should support at least the same action space range as the RE, along with a few safety meters, so that the participant does not believe she will run into a virtual obstacle. This is especially important if not using a size-matched replica of the real space, which should already have enough room. Further, we recommend modeling a space that could realistically exist. Wide open spaces, such as infinite grassy planes, and spaces with a myriad of depth cues, objects, and patterns are not common real-world places and could be overwhelming to participants. Some spaces that could realistically exist include a standard classroom, a city block, or a cabin. Additionally, though we recommend modeling spaces that could exist in reality, they need not be photorealistic as the quality of the graphics does not contribute to the observed compression when blind direct walking is used (Grechkin et al., 2010; Kunz et al., 2009). As will be explained in Section 4.2.1, high-quality models may be the most desirable anyway. However, since many research laboratories may not have the time or ability to make realistic or high-quality models, publishing any good models for use by other research laboratories would be helpful.

There may also be some special considerations that depend heavily on the context of the research when deciding on the VE, such as the size of the VE versus the size of the RE and nonpictorial, pictorial (Cutting and Vishton, 1995), and environmental depth cues (Renner et al., 2013). To begin, a VE of a greater size than the RE has the advantage of assuaging participant concerns about running into virtual, and hence perceived real, obstacles and improving confidence while walking blind. Further, size-different VEs may not have a significant effect on distance estimates (Nguyen et al., 2008). However, people may tend to overestimate distances if aspects of the VE, such as familiar targets, are minified since familiar size cues dictate that smaller objects imply longer distances (Nguyen et al., 2008). This should not be concerning if the RE is long enough.

Cutting and Vishton, (1995) described nonpictorial depth cues as clues from our own physicality that we use to determine distances, including accommodation, convergence, and motion parallax. Since the eyes of a person must focus, or accommodate, on the lenses of the HMD to view the image, but also converge on the perceived three dimensional image, the conflicting accommodative and convergent senses may cause visual discomfort (Hoffman et al., 2008; Kramida, 2016; Lambooij et al., 2009). Additionally, the 24 hour study by Steinicke and Bruder, (2014) showed that prolonged exposure to HMDs can make the accommodation sense appear to distort the image and increase simulator sickness to unbearable levels after periods of motion. Though there is research being done to help alleviate this problem (Kramida, 2016), currently only changes to the graphics can help reduce this sensory conflict, specifically by introducing various reliable depth cues to decrease the influence of accommodation (Hoffman et al., 2008). Most research should not last long enough that sensory conflicts become problematic.

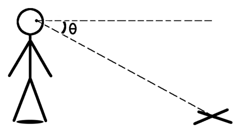
Cutting and Vishton, (1995) described pictorial depth cues as distance clues resulting from static objects in the environment and their positions relative to an observer and other objects. These include occlusion, relative size and density, the angle of declination, and aerial perspective; examples are shown in Figure 4. The more of these objects that exist in the VE, the more complex the VE becomes. Nguyen et al., (2011) found that participants appear to judge distances more accurately in less complex environments, but due to potential biases in their test environment, more research is needed to confirm these results. Contrarily, Renner et al., (2013) summarized many past works that showed that complex environments with vertical lines or regularly patterned floors garnered better results than simple, sparse, or grassy environments. Considering that the quality of the graphics does not seem to affect distance estimates for blind direct walking, the complexity or simplicity of the environment may not be a breaking factor for research. As before, we recommend using a space, including the pictorial depth cues, that could realistically exist.

Environmental context may affect distance perception (Lappin et al., 2006; Norman et al., 2018; Witt et al., 2007). For instance, in a real-world bisection study by Lappin et al., (2006), participants performed better in an open lawn (simple environment) than a lobby (complex environment) or hallway (familiar size and linear perspective cues, but a poor FOV). In contrast, Norman et al., (2018) found that both younger and

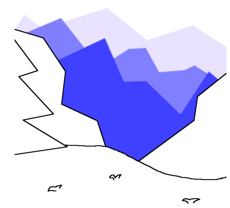


(a) The closer car occludes the further away car.

(b) The closer set of flowers appears larger and farther apart than the further set of flowers.



(c) The angle of declination.



(d) In aerial perspective, the further mountains appear bluer and paler.

Figure 4: Examples of the pictorial depth cues as described by Cutting and Vishton, (1995).

older adults perform better in indoor environments than outdoor. Witt et al., (2007) showed that when targets were placed the same distance away from an observer, the targets with empty space beyond them appeared closer than those with a wall beyond them, but only if both types of targets were viewed. Though these were for the real-world only and, to our knowledge, no similar virtual-world experiments have taken place, these results imply that certain environmental depth cues may require consideration.

4.1.3 Pre-Experiment

We normally perform all pre-experimental procedures, including completion of consent forms, pre-screening procedures, instructions, pre-questionnaires, putting on wearable items, and practice blind walking (see Section 6 for more), in a location other than the main laboratory space. This could mean completing them in a different room, in the hallway outside of the main laboratory with the door shut, or in another location entirely. We do this for two main reasons. First, subjects who view the laboratory space before beginning experiments may get a sense of how large the space actually is and their location in the space, which could bias results. Second, it is common practice for experimenters to disallow participants from viewing the laboratory space before virtual experiments. Since virtual participants do not view the laboratory space before trials begin, neither should real participants for consistency.

4.2 Hardware

4.2.1 HMD

Several options for display hardware are available, including Display Walls, HMDs, CAVE, stereoscopic desktop monitors, and BOOM2C (Renner et al., 2013). Due to infrequency of use (Renner et al., 2013) and immobility, the CAVE, desktop monitors, and BOOM2C will not be discussed here. Display Walls are made up of several television screens or computer monitors put together to create a giant wall that displays the virtual environment. One major advantage to having a wall of displays is the wide FOV it affords, much larger than HMDs (Grechkin et al., 2010). Further, Display Walls do not require observers to wear equipment, so this system allows observers fully unencumbered and wireless motion, unlike wired HMDs. Additionally, these systems can be sterescopic or not, depending on the type of screen used. Despite these advantages, however, there are some major disadvantages. First, Display Walls are not as widely used as HMDs (Renner et al., 2013), making reproducing experiments more challenging. Display Walls must also be moved into place for viewing and removed for every blind direct walking trial (Gaines and Kuhl, 2020). This adds an undesirable time delay between viewing the targets and the start of the walk. Depending on the size of the Display Wall, moving it alone or with only a few experimenters could be dangerous, necessitating additional coordination in experiment time slots. Further, Grechkin et al., (2010) noted that participants completing a timed imagined walking task significantly underestimated distances for closer targets, indicating that they may have been hyper-aware of the Display Wall. This suggests that Display Walls may be more prone to immersion breaks. Moreover, implementing motion parallax is difficult since participants must wear some head-tracking apparatus, and a separate tracking system that is not blocked by the wall is required. Most importantly, the size of the Display Wall is antagonistic to portability.

The most well-known and widely used (Renner et al., 2013) display hardware system is the HMD, which is a helmet-like system worn on the head with lenses near the eyes that present the virtual world to the user. Ideally, this system is “light-tight” and does not allow light or other outside stimuli from entering the edges of the system through gaps, ensuring a more immersive experience. This system has many important advantages. One major advantage is that since HMDs are widely used in research, reproducing and comparing experiments is easier. Further, many systems are relatively cheap and easily available for research, training, and entertainment purposes. Many manufacturers are currently creating wireless versions, meaning that movement is less inhibited and less prone to tangling as compared to their wired counterparts. Additionally, motion parallax and stereoscopic vision are easily achieved in HMDs. Lastly, they are portable and often come with a tracking system built-in, meaning that HMDs can be used in a variety of REs. Due to the importance of these advantages, we recommend using a HMD to present the VE to participants.

Despite these major advantages, HMDs still have several limitations, including the accommodation convergence conflict, which was detailed in Section 4.1.2, a limited FOV, possible light in the periphery, certain physical properties of the HMD, connecting wires, graphics quality, motion parallax limitations, latency, and interpupillary distance (IPD) capacities. Some of these have been shown to cause a significant effect on distance estimates, while others have not, and others may be mitigated or eliminated entirely. Starting with the FOV, as illustrated in Figure 5 normal human peripheral vision allows over a 200-degree FOV whereas HMDs provide a FOV of between 60 degrees, in older models, and 110 degrees, in contemporary models. This

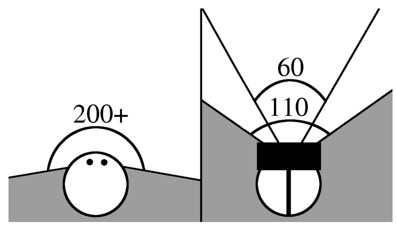


Figure 5: Humans have 200+ degrees of peripheral vision normally while HMDs afford only 60 to 110 degrees of peripheral vision.

restriction has shown significant effects on distance estimates in VR (Li et al., 2016; Li et al., 2018; Li et al., 2014; Li et al., 2015) and possibly the real world (Li et al., 2015), though Creem-Regehr et al., (2005) found that a restricted FOV has no affect on distance estimates if head movement is allowed. However, Li et al., (2014) and Li et al., (2015) showed that a properly calibrated, wide-FOV HMD may allow participants to more accurately judge distances in VR without any further manipulations. Further, peripheral light, such as from brightening the peripheral area of the display (Li et al., 2018) or from light that leaks in through gaps between the HMD and the face of the user (Jones et al., 2013; Jones et al., 2011), can affect distance estimates. Blocking windows and turning the RE lights off helps mitigate light leakage. There are cases where the HMD seal is good, but this may cause dry eyes (Steinicke and Bruder, 2014). Fortunately, experiments often are short enough that this is not a problem.

Since HMDs are placed on the head and require many electrical and optical components, they exert excess force and differing moments of inertia on the head, as shown in Figure 6, which may contribute to the compression seen in VR (Grechkin et al., 2010; Willemsen et al., 2004), at least when coupled with the restricted FOV (Buck et al., 2018; Willemsen et al., 2009). A lighter, more compact, or wireless system may aid distance estimation. Currently, wireless HMDs are beginning to take the spotlight. They boast enhanced maneuverability and safety due to the lack of wires, which may improve immersion. On the other hand, wireless systems may be more prone to network connection issues, which could be detrimental to presence in the virtual environment. Gonçalves et al., (2020) showed that participants experienced

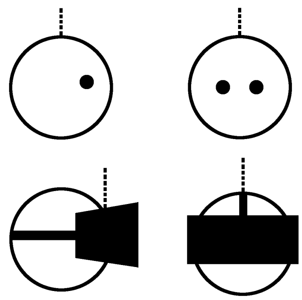


Figure 6: The center of mass of the head while wearing a HMD is approximately centered between left and right but is more forward than the normal human head.

similar levels of presence in both wired an wireless HMDs, but they did note some important factors, such as a fear of tripping over wires and network instability, that may require additional study. Further research on the differences between wired and wireless HMDs may help draw more definitive conclusions.

Screen resolution in HMDs is lower than retinal resolution, meaning that, regardless of the photorealism of the scene, the quality of the graphics viewed in HMDs is lower than what our eyes are capable of seeing. Fortunately, past studies (Buck et al., 2018; Grechkin et al., 2010; Kunz et al., 2009; Thompson et al., 2004) suggest that the quality of the graphics does not affect distance judgments when blind direct or triangulated walking is employed. There is a question, however, on how the graphics quality influences the sense of presence in VEs. Zimmons and Panter, (2003) found no significant effect of the quality of the graphics on the sense of presence in their pit room VE citing that the sense of danger looking over a six meter pit may have eclipsed the desired factor. Contrarily, Slater et al., (2009) did find that the quality of graphics had a significant affect on the sense of presence in a similar pit room VE. More research must be done on a wider variety of VEs to make further conclusions. Since VR attempts to replicate reality, it may be most useful to use as realistic of graphics as the modeler is capable of creating. Also, as mentioned, publishing good models for use by other research laboratories would be helpful.

HMDs do allow for motion parallax, though depending on the setup, motion parallax can be missing or distorted in VR, which is not necessarily a bad thing. According to Cutting and Vishton, (1995), the effectiveness of motion parallax as a depth cue decreases with distance. Additionally, Jones et al., (2008) showed that motion parallax does not contribute to the observed compression in VR. Since VR attempts to replicate reality, the inclusion of motion parallax is desirable, but it is not necessarily detrimental if it is distorted or missing all together.

Latency, or the time delay between when a person moves and when that motion appears in the display, can affect performance in VR (Ellis et al., 1999). Lower latency, or shorter delay times, correspond to higher presence in the VE (Meehan et al., 2003).

Since frames refresh at a fixed rate, which is around 80 Hz or better for contemporary systems, motion delay times should ideally match the frame refresh rate, but some lag in latency is acceptable. Latency lag of 15 ms, the average detectable latency difference threshold provided by Mania et al., (2004), or longer may necessitate additional manual improvements to avoid performance degradation.

Finally, the IPD is often static in HMDs, but some HMDs do allow for adjustments. Since the lenses are collimated, so long as the eyes are mostly centered, the user should be able to see the image. Additionally, current evidence suggests that the ability or disability to adjust the IPD of the HMD does not affect distance judgments (Buck et al., 2018).

With all of the limitations and potential fixes detailed, and the fact that HMD technology is rapidly improving and has become cheaper and more widely available, which HMD is the “best” for research? There are plenty of options when deciding which to use. A study by Buck et al., (2018) compared several HMD systems and found that newer, commercially available systems outperformed the older models. More specifically, their results showed that the weight of the HMD and the FOV likely contribute to the observed compression in VR, but the screen resolution, IPD, and refresh rates of HMDs likely do not. Though the specifications are an important factor, reproducibility of experiments in research is arguably more important. The most effective way to assure this is to use a popular system that is widely available and has good specifications, even if not the “best” specifications.

4.2.2 Tracking System

The real world appears to move and change according to the position and orientation of an observer within it. If the displayed image is not a panorama or other static image, the same should happen to the virtual world displayed in a HMD. The tracking system monitors the movements and orientation of a user and sends that data to the program, which will update the display image accordingly. Contemporary HMDs come with a tracking system built into the headset, which holds two major advantages. First, only one system is needed to both display and update the image during an experiment. Second, the built-in tracking system will monitor the real-world eye position of the participant. This is important because the eyes take in visual information, so distance perception begins at the eyes. Therefore, any targets should be placed or rendered their specified distance as measured from the eyes of the participant, and the final position after the walk also corresponds to the eye position. Measuring from and to the eye position is difficult in reality, so more about this will be discussed in Section 5.1.2.

The major disadvantage to a built-in tracking system, demonstrated in Figure 7, is that the tracking area is often too small to cover the required walking distance, so the final position of the participant cannot be automatically taken. Instead, the

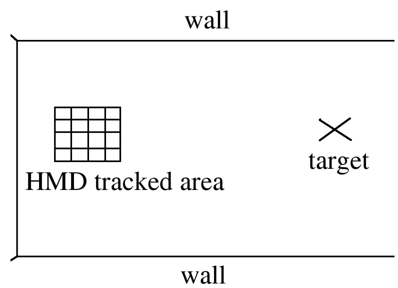
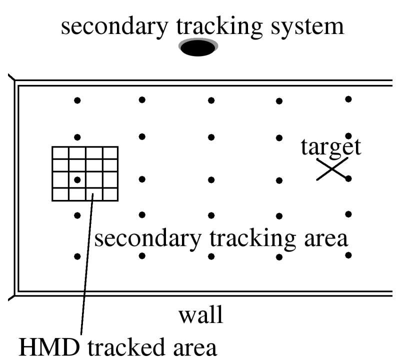
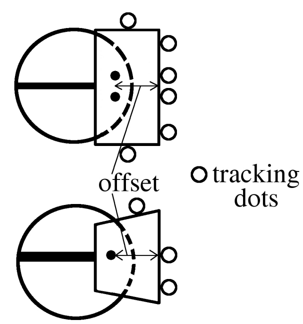


Figure 7: The built-in tracking system of the HMD may be too small to track the full walking distance.

experimenter could physically measure the distance the participant walked, which presents its own difficulties that will be discussed in Section 5.1.2. Further, since a participant is wearing a HMD, a tracking system would provide a more accurate final location measurement. As such, a secondary tracking system, such as the wall-mounted system represented in Figure 8a, used to monitor only position may be employed. This separate tracking system will likely require the use of tracking dots,



(a) A secondary tracking system mounted on a wall may cover the entire walked area.



(b) A visualization of how the tracking dots may be placed on the HMD, as well as the offset between the dots and the eyes.

Figure 8: A secondary tracking system may help with tracking the designated target area, but extra tracking dots may need to be added to the outside of the HMD.

which, as shown in Figure 8b, are attached to the outside of the HMD and will be offset from the eyes of the participant. This minor disadvantage is eliminated by carefully calculating the offset between the position of the dots and the position of the eyes. If the experimenter chooses to use a secondary tracking system, we recommend using a system that will work best for the purposes of the research.

4.2.3 Calibration

It is important to properly calibrate the hardware used in an experiment. If factors such as tracking system correctness, eye height, the horizon line, the optical lenses, and the FOV are improperly calibrated, distorted images, deterioration of presence in the virtual world, and avoidable inaccuracies in distance estimations may result. For starters, to ensure that the tracking system is correct, first check the tracking grid. The tracking grid is an invisible grid structure inside of which the tracking system accurately monitors the position and orientation of the HMD. This grid will have an origin and an orientation with respect to the room. Figure 9 represents a tracking grid that is properly oriented and placed with respect to the room. When

(a) Front view.

(b) Side view.

Figure 9: An example of correct alignment of the tracking grid with respect to the room and floor.

the HMD is being worn by a user, the HMD should be some reasonable distance off of the ground and the same everywhere in the tracked space. We check this by placing the HMD at various predetermined heights in different locations within the tracking grid. If the measured height is unreasonable, as it would be in either grid location in Figure 10, or varies anywhere in the tracking grid, as it would in Figure 11b, then we reposition the grid so that the height of the grid is as close to the ground plane as possible. Further, if any position of the HMD with respect to the origin of the grid

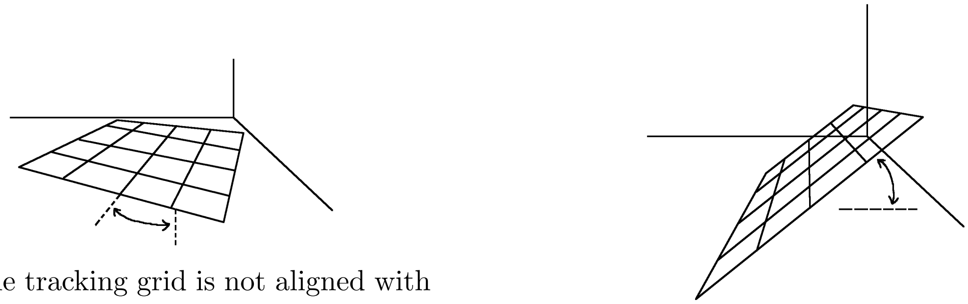
(a) The height of the HMD will be negative if below this grid.

(b) The height of the HMD will be positive even if placed on the ground.

Figure 10: Examples of the tracking grid floating above and sinking below the floor.

is unexpected, as would be the case if the grid appeared as Figure 11a, we tweak the position of the tracking area until it is properly aligned.

If a separate tracking system is required, calculate the offset between the tracking dots and the eyes of the user. This is very important because the tracking system



(a) The tracking grid is not aligned with respect to the room.

(b) The tracking grid is not aligned with respect to the floor.

Figure 11: Examples of how the tracking grid may be twisted with respect to the direction it is expected to face.

will trace the markers rather than the eyes. For example, if the tracking system markers are ten centimeters above the eyes, it could cause the software to render the graphics as if the user were ten centimeters taller than he actually is, which, as detailed below, is detrimental to distance estimates in VR (Leyrer et al., 2011; Leyrer et al., 2015). In addition to this vertical error, similar errors can occur horizontally or forward and backward in depth. This problem can be easily tested if the origin of the tracked space is well-defined and known in the real world. First, imagine the three-dimensional point in the HMD representing the bridge of the nose between the eyes of a user. Then, set the HMD on the ground so that this point is directly above the origin. Verify that an appropriate value with some small height above the origin is obtained. If the system is properly calibrated, rotating the HMD 90 degrees repeatedly should result in a similar position value from the tracking system after each rotation.

It is important that the camera height in the virtual world be correct. If the camera height is too high for a particular user in the virtual world, then the observer will view the world as taller than she actually is and distances will be significantly shortened (Leyrer et al., 2011; Leyrer et al., 2015). Curiously, Bian and Andersen, (2013) found that a 1.1 meter increase in eye height in the real world expanded distance judgments rather than shortened them. To our knowledge, this discrepancy has yet to be addressed. Also in the virtual world, if the camera height is too low, then the observer will view the world as shorter than he actually is and distances will be significantly expanded during blind direct walking tasks (Leyrer et al., 2015). This expansion is not seen for verbal reports (Leyrer et al., 2011), but since this paper deals with blind direct walking, it is important that the height be correct. If the tracking system is calibrated, the height value from the tracking system can be used. In some cases, an experimenter may wish to measure the eye height of a user prior to the experiment to verify that the eye height values from the tracking system are accurate. To accurately measure eye height, first place a long piece of tape vertically on a real wall, as shown in Figure 12. Next, have the participant stand very close to

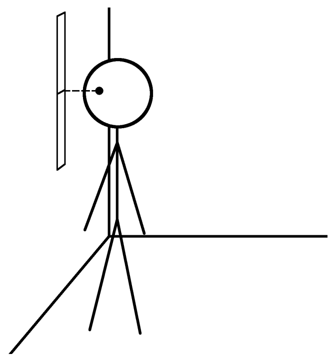


Figure 12: Have a participant stand close to a vertical piece of tape on the wall with his head level, and place a line on the tape at her eye height.

the wall with normal posture and look straight ahead with her head level. Place a pencil mark on the tape corresponding to the height of the center of the eyes. Measure this height with a tape measure and compare to the height measured by the tracking system.

The horizon line, if visible, should similarly be at a the eye height of a particular person. For example, you can create an artificial horizon in a real room by measuring the eye height of a user and placing a horizontal piece of tape on the walls at that eye height. If the user then viewed a virtual scene in the HMD while standing up, the virtual horizon should align with the artificial horizon in the real room. The correctness of the horizon line is important for scenes that have few other available depth cues, such as infinite grassy planes. We recommend avoiding such environments, unless low-cue environments are being studied specifically, as they do not commonly occur in real life. But if one will be used, we refer the reader to Messing and Durgin, (2005) who provide a method to properly render a horizon line when it is impossible to render an infinite plane in the software.

The pitch of the HMD was not found to be a contributor to the underestimation of distances in VR (Kuhl et al., 2009), but we recommend that the pitch be properly calibrated to ensure maximal accuracy and correctness of the image. Calibration of the yaw and roll may also be important but, to our knowledge, have not been studied in the distance judgment literature. This calibration is particularly important if an orientation sensor will be attached onto the HMD and it is uncertain if the orientation sensor and the optical axis of the HMD are aligned. If the orientation sensor is integrated with a widely used consumer HMD, it is likely safe to assume that the manufacturer has handled this calibration. Otherwise, we refer the reader to Kuhl et al., (2009) who provide a full procedure to properly calibrate the pitch of the HMD. The correction will be slightly different for different users, but most individual values will hover around the averages.

Pincushion distortion is when the image appears to bend inward toward the center of the display, such as when pins are placed in a pincushion, and is most evident near the edges of the image. According to Kuhl et al., (2009), pincushion distortion does not contribute to the compression in VR, but they recommend correcting it anyway since our eyes do not have this distortion during real vision, and correction will provide a more realistic image. We refer the reader to Kuhl et al., (2009) for correction details. The authors state that pincushion distortion must be corrected before the FOV is calibrated, described next, since the distortion is nonlinear. However, most modern consumer HMDs, such as the Oculus Rift or HTC Vive, come with software development kits that automatically include pincushion distortion correction.

There are two FOV numbers that VR developers need to be aware of. The first is the FOV used to render the graphics, which is sometimes called geometric FOV or GFOV, and the second is the fixed FOV of the display device, denoted DFOV. It is important that developers set the GFOV to match the DFOV. If the GFOV is smaller or larger than the DFOV, the image will be minified or magnified, respectively (Kuhl et al., 2006; Kuhl et al., 2009; Li et al., 2014; Li et al., 2015). We refer the reader to Kuhl et al., (2009) who detail a method to calibrate the horizontal and vertical FOV of the display and to Li et al., (2015) who provide a similar method of calibrating the horizontal FOV. The default FOV settings provided by manufacturers of mass-produced consumer devices are likely correct. However, manufacturer supplied DFOVs made by smaller manufacturers may be less reliable since they have fewer users double-checking the values. In addition, confirming that the graphics rendered on the screen are correct through a calibration process can help catch cases where a programming mistake unintentionally changed the GFOV in a significant way.

Once the HMD is properly calibrated, make sure the program and tracking system recognize the initial position and properly measure target distances. To do this, choose a target and a distance (see Section 4.4 for more). Starting from the initial real position (see Section 5.1.1), measure the distance in the RE and place the chosen target on the ground. Stand with the eyes over the initial real position, and display the VE in the HMD along with the corresponding target and distance. Repeatedly lift and lower the HMD to view the real target and the virtual target, as shown in Figure 13. If the two targets do not align on the same position, recalibrate the tracking system as necessary.

4.2.4 Masking Noise

Uninhibited auditory cues, such as echoes, footsteps, ventilation systems, or the voice of the experimenter, may assist participants in localizing their position in the real world. Though we found no studies about the benefits or drawbacks of drowning out ambient noise, it is believed to be important to reduce or eliminate these cues through the use of headphones with masking noise, noise-canceling headphones, ear plugs,

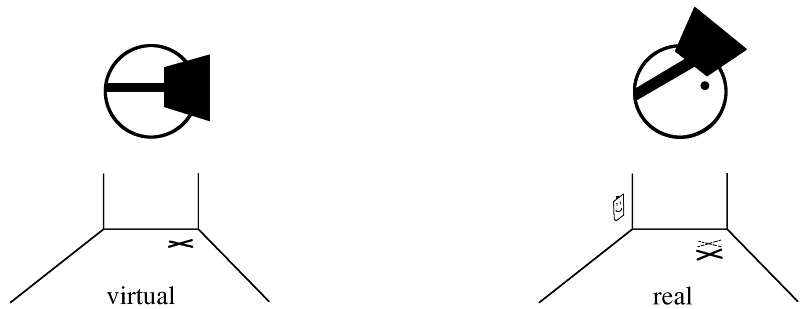


Figure 13: Calibrate the virtual target distance by repeatedly lifting and lowering the HMD and ensuring the real and corresponding virtual targets align. In this case, the dotted 'x', or the remembered virtual target position, is not aligned with the real target.

over-ear hearing protectors, or any other noise-blocking or noise-canceling apparatus. We use headphones with some masking noise since they allow the participant to hear verbal instructions and are easy to use during both virtual and real testing.

Many contemporary HMDs come with built-in headphones, making it easy to play any noise file directly from the computer. If the HMD comes with built-in headphones, using them is easiest since removing them is challenging and may break the headset. However, some HMDs do not come with headphones, and real-world experiments do not use HMDs. In these cases, we use a separate pair of headphones or noise-canceling headphones that are either the over-ear or on-ear style. Ear buds could also be used but may be difficult to clean between participants. If these separate headphones are wired, they cannot be connected to the computer since they would impede movement. Instead, connect wired headphones to a portable media player playing the masking noise, which will then be clipped to the clothing or held in the pocket of the participant. The wire may then be tied to the HMD wires, if applicable, or moved in some way as to not distract the participant.

Each RE will have a unique ambient noise, so unless noise-canceling headphones are practical and sufficiently mask the ambient noise, we recommend creating a unique masking noise file that suits the RE. We find that white noise is a good starting point. If this noise does not sufficiently block the ambient noise when played at a comfortable volume, try other colors of noise, like brown noise or pink noise, and use whichever one works best. Using the created masking noise file for all experiments, real and virtual, will also help maximize consistency between experiments.

Some studies, such as Razzaque et al., (2001), programmed noise or instructions so that it appeared to be coming from a radio within the virtual world in an attempt to improve immersion. However, creating a realistic radio that properly interacts with the VE is difficult, as is obtaining the same masking effect in the real world using a real radio. Real-world experiments will still require headphones to make the masking effect consistent. Since headphones will be used in both types of experiments, playing a masking noise without a radio seems to be the easiest approach.

4.2.5 Communication

Communicating with the participant during the testing phase is crucial for both the success of the research and the safety of the participant. The headphones worn by the participant for acoustic masking (see Section 4.2.4) can also be used to deliver verbal instructions via a microphone. If the headphones are connected to the HMD, then the microphone may be connected directly to the computer. Otherwise, the microphone may need to be connected in some manner to the participant, meaning the experimenter may have to walk around with the participant as to not hinder movement. Alternatively, the experimenter could talk loudly enough so that the participant can hear the experimenter over the masking noise. If the experimenter is following the participant throughout the experiment, more on this in Section 5.3, the participant will not be able to use the voice of the experimenter as a localizing cue. This simple alternative also reduces wires, the number of things that can go wrong, and microphone and volume setup time.

Other methods of communication exist, including playing prerecorded instructions (Razzaque et al., 2001) or displaying text on the HMD screen (Steinicke et al., 2009). These are as easy as pressing a key. However, these communication methods might be insufficient to handle emergency or unforeseen circumstances. For example, an experimenter may wish to remind the participant about a specific experiment instruction after he makes a mistake. Further, these communication methods, particularly displaying instructions to users, are infeasible during an analogous real-world experiment.

4.3 Avatars

Participants are often encouraged to look around in the virtual environment when viewing targets. This may include looking straight down to where a body should be. Figure 14 illustrates three possible views. Looking straight down reveals the body

Figure 14: When looking down, the body of the person is viewed in the real world while the body of an avatar, if used, is viewed in the virtual world (left). If no avatar is used, then looking down in VR results in viewing nothing (center) or a collar (right), if one is used.

of the person in the real world, but this is not true in VR. If an avatar, or virtual model that a person may control in some manner, is present, then looking down in VR reveals the body of the model. However, most research does not use avatars, so looking down reveals the virtual floor. Creem-Regehr et al., (2005) investigated how the lack of a body influenced distance perception by employing a collar on real-world participants that restricted their views out to 1.5 meters. This "collar trick" showed that the view of a body is not necessary to make accurate distance estimates. Even so, several studies (Leyrer et al., 2011; Mohler et al., 2008; Mohler et al., 2010; Phillips et al., 2010; Ries et al., 2008; Ries et al., 2009) have shown that the inclusion of an avatar significantly aids distance estimates in VR, though some (Lin et al., 2011a; Lin et al., 2011b; McManus et al., 2011) showed otherwise.

Using avatars adds technical complexity in tracking implementation and in the creation of realistic avatars. Full-body tracking may require additional sensors or markers be placed on the body of the participant. Inverse kinematics may be necessary when there is insufficient tracking information, such as when the knees are not tracked but the hips and ankles are, to reconstruct a full avatar. Errors in tracking or by the inverse kinematics algorithm may also introduce distracting errors in the VE. Construction of the avatars may or may not require consideration of many factors, including the height, shoe size, limb length, skin color, weight, gender, and attire of the participant. The overall quality and realism (Achenbach et al., 2017; Jo et al., 2017; Lin et al., 2011a; McManus et al., 2011; Ries et al., 2008) of the avatar may also have implications that should be considered. Experimenters using avatars should review other distance judgment studies that use avatars and possibly other studies on presence or body ownership with avatars (Freeman et al., 2020; Jo et al., 2017; Latoschik et al., 2017; Lugrin et al., 2015a; Lugrin et al., 2015b; Waltemate et al., 2018) in HMDs.

4.4 Targets and Distances

Obviously, the target is the end goal of a blind walk. From the few papers that detailed the targets they used, a wide variety of target sizes, shapes, colors, and materials have been employed, including traffic cones, cylindrical poles, boxes of light, and foam board cutouts. Unfortunately, most papers present little detail, if any, about employed targets. To our knowledge, no research has been done concerning target types or colors and their influence on distance perception; however, there are still some things to be aware of.

Target shape is an important consideration. Familiar objects, such as pop cans, traffic cones, and hockey pucks, may present familiar size cues. Participants who use such targets as references may bias their distance judgments. One way to avoid familiar targets is to use a target of constant angular size, such as the programmable light box used by Philbeck and Loomis, (1997). However, this requires the room be completely dark. The only other way to accomplish this in the real-world would be to painstakingly create targets that are of constant angular size at specified distances. This would require tedious and complex measurements and planning. Instead, creating several simple and differently shaped targets would be enough to avoid familiarity. Some shapes include a triangle, such as the virtual target we use in Figure 15, circle, square, and 'X,' but any shape will work. These can be made from foam board or cardboard in the real world and quickly modeled for the virtual world.

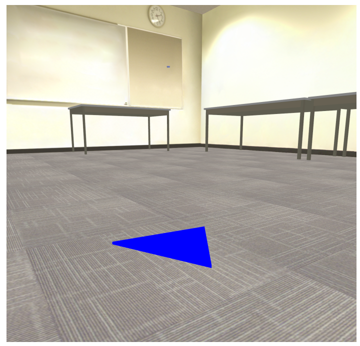


Figure 15: An example of a virtual target we use. It is a simple triangle in shape and can be easily replicated in the real world using foam board and construction paper.

Target color and material are also important considerations. Targets should stand out in the environment rather than blend in. However, some participants may be colorblind. For example, a red target would certainly stand out in green grass except if the viewer was red-green colorblind. Additionally, shiny or glossy material may induce glare and be difficult to look at. Matte colors and materials are easier to look at and easier to model and shade. The way we accomplish this is by dynamically coloring and diffuse shading virtual targets using a program during experiments, and covering real targets in colored construction paper.

The number of targets has practical implications. Since 13 papers out of a sample of 30 papers published between 1997 and 2018 used only one target, using exactly one target is both easy and common. However, doing so presents some significant drawbacks. For example, if each distance is shown multiple times, participants may use the relative change in size of the target to guess how far to walk relative to other distances. Also, if the same distance is repeated consecutively, participants may memorize where to walk rather than perceive the distance. These cognitive influences could skew results. Using multiple targets helps avoid this. For example, say s different shapes, z different sizes, and c different colors are chosen. If all combinations of target types are created, then there will be szc different targets to choose from, meaning a repeat trial can be avoided. Further, differently shaped, sized, and colored targets can be easily reused for future experiments whereas constant angular sized targets cannot. Therefore, creating multiple targets in this fashion is easy and effective for real and virtual experiments, and it helps avoid distance memorization and other undesirable cognitive influences.

Ordering targets in a particular way may further reduce the possibility of repeated trials and their side effects. One option includes mapping targets to trial numbers and distances. If enough targets were made, repeated trials could be eliminated, but this may mean creating a huge number of targets. Randomizing targets, on the other hand, both allows for a more manageable number of targets and reduces the chance of patterns naturally. Fully randomizing targets does not eliminate the chance of patterns, however. Though, the likelihood that all participants are shown the same targets in the same order is $(\frac{1}{szc})^t \times p$, where s, z, and c are as defined above, szc is the total number of targets, t is the total number of trials, and p is the number of participants, which is tiny for an experiment using even a small number of targets and trials. Further, the chance of consecutively repeating a target is only $\frac{1}{scz}$, which is also small for a reasonable number of targets. Other types of randomization include block randomization, pairwise randomization, and random permutations, but full randomization is the simplest. One way to accomplish this is to write a list of targets and randomly choose one in the program displaying virtual targets, and map real targets to numbers and roll an appropriate die. Whatever method is picked, randomly choosing one target to display from a list of many targets in each trial will significantly reduce the chances of introducing patterns and other biases.

In the virtual world, much like the real world, targets should be placed on the ground. Sometimes virtual programs may glitch and place targets in strange places or have them hover above or sink below the ground. Though this may be obvious, it is important to check virtual targets to make sure they are on the ground within an acceptable error.

The specific distances used are just as important as the targets. Since these experiments take place in action space, some distances between two and 30 meters is recommended. From a survey of 30 papers published between 1997 and 2018, three or four unique distances in the range between two and seven meters were the most common number and range of distances used, though up to seven unique distances have been used. Further, 25 of these studies performed multiple trials per distance. One option that seems to work well is to use four unique, whole-number distances seen two or three times each and three unique, half-number insertion distances seen randomly during the experiment one time each. For example, each of the distances 3, 4, 5, and 6 meters could be shown twice, and the distances 3.5, 4.5, and 5.5 meters could each be randomly shown once. These random insertion distances are different enough from the whole-number distances that they reduce the chances of participants memorizing distances and force the participants to pay attention to what they are doing.

The arguments for ordering distances are similar to those for ordering targets. Distances could be sequentially shown in some set order, but this may show unwanted patterns. Full randomization of distances reduces the chances of repeated trials naturally. Adding on to the statistical analysis for targets, suppose d distances are used. Since each distance will be shown only a set number of times, d will decrease by one after each trial. If all targets are in play for the duration of the experiment, then the chances of exactly repeating a consecutive trial, call it r, are $\frac{1}{szcd} < r \leq \frac{1}{szc}$. For experiments with multiple initial positions (see Section 5.1.1), we can add to this analysis the number of initial positions, call it i. Thus, the chances of exactly repeating a trial if the target, the distance, and the initial position are all randomly chosen are reduced to $\frac{1}{szcid} < r \leq \frac{1}{szci}$. Similarly to targets, one way to accomplish randomizing distances would be to randomly choose a distance from an eliminating list in the virtual program and pull a distance written on a piece of paper out of an opaque bag without replacement for real-world experiments. Whichever method is picked, randomly choosing the distances and targets, as well as the initial position if multiple are used, from lists of many choices will help mitigate biases in the data.

The above will work well later in the experiment when the participant gets a better understanding of what she needs to do. Early on, the participant will do weird things, possibly because he either did not pay attention to instructions or does not understand the task even if she says he does. We recommend allowing practice trials to weed out the early biased data and help participants gain a better understanding of the task. Curiously, from the same 30 paper survey, the majority of studies did not mention practice. Practice is very important, and we recommend always doing it. There is no reason why participants need to know about practice, so the first two or three trials of an experiment could simply be thrown out as the practice trials.

One way to choose the practice distances is to simply run the experiment. This entails choosing a set of distances and running them in whichever order until the experiment completes. However, throwing out the first few trials means that some distances will be seen fewer times than desired in the final data set. Further, there will likely be fewer practice trials than unique distances, so participants will be able to practice certain distances and not others, which could skew results. Instead, a separate list of practice-specific distances could be used. The program can be written to seamlessly switch between the lists at a specified trial number in the virtual world, and two different bags could hold the different distance lists in the real world, for example. In terms of the practice distances themselves, one option is to use the random insertion distances described above since their only purpose is to ensure participant focus and are not intended to be a part of the final data set. Another option is to use noisy distances. As illustrated in Figure 16, noisy distances are created by adding some random non-zero offset to a distance used in the actual experiment. Other viable choices for practice distances exist. In the end, the practice trials, though crucial to reducing biases and aiding in participant understanding, will be thrown out.

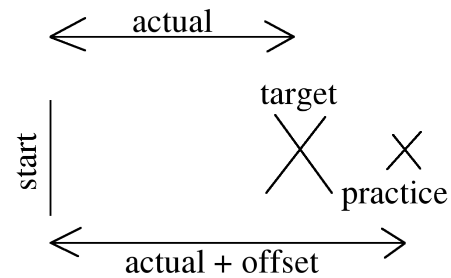


Figure 16: Noisy distances are experiment distances with an additional non-zero offset.

5 Experimental Design Procedures

5.1 Positioning and Measurement

5.1.1 Initial Position

We recommend choosing an initial real position such that a real-world comparison experiment could reasonably take place in the same RE, regardless if one will be executed presently or if the real and virtual tests will be completed in the same RE. Since, ideally, both real and virtual experiments will take place in the same RE, using the same initial position for both sets of trials maximizes consistency. Even if two different REs are used, either RE could be used for both real and virtual tests in a future experiment. Further, we recommend locating the initial real position at least one arm's length away from any walls or other obstacles, such as shown in Figure 17. Accidental collisions may cause participants to become overly cautious

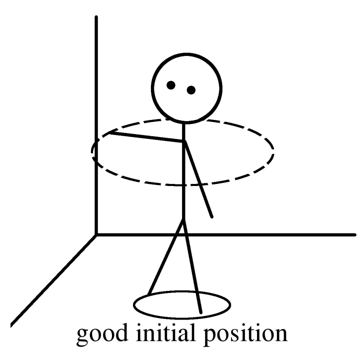


Figure 17: The initial position should be at least one arm's length away from obstacles to avoid accidental collisions while viewing.

and skew results. The extra space ensures that the participant will not accidentally bump anything.

Once a good initial real position is determined, we normally mark it with a piece of tape to aid in repositioning for subsequent trials. Aligning the eyes with the tape is most ideal since distance perception begins at the eyes and tracking systems measure the eye location. Though this is challenging to accomplish since the eyes approximately align with the balls of the feet, as shown in Figure 18, we find this to be the best approach.

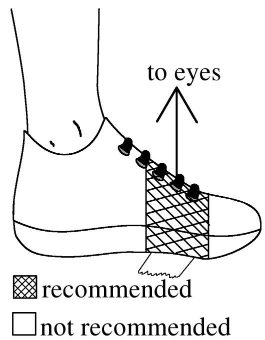


Figure 18: We recommend aligning the balls of the feet with the tape mark.

Figure 19 shows two other possible foot alignments - the toes and the heels. These

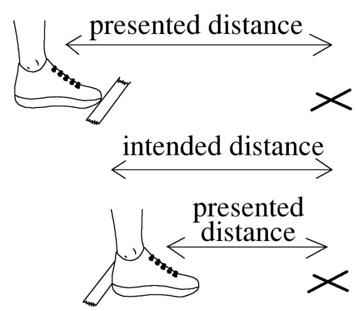


Figure 19: The toes or heel of the foot may be aligned with the tape instead, but this will change the presented distance with respect to the intended distance in the real world.

places are more convenient and easier to align on the tape, but the presented distance is longer or shorter, respectively, than the intended distance for real-world participants. Though this is not detrimental for virtual experiments, as will be described next, the disparity between the presented and intended distances for real-world participants provides a poor comparison experiment.

The initial virtual position can be more forgiving if the software implements repositioning. Regardless of the true initial real position of the participant, the virtual environment can be translated such that the distance between the participant and the target is correct. Thus, the initial real position of virtual-world participants need not be as precise as for real-world participants. If repositioning is implemented, record the starting position from the tracking system for each trial so that the actual distance walked can be calculated. Even if repositioning is not implemented, recording the starting position can be helpful for multiple reasons. First, it allows experimenters to quantify and report how much variation and bias there was in putting people at the starting point. For example, without repositioning, if participants were placed ten centimeters too far forward on average, then a five meter target was actually 4.9 meters on average. Second, if multiple experimenters are present, placement performance can be compared. Third, if there are temporary and substantial tracking errors, they will be visible in the starting position data. Fourth, verifying that any bias did not change substantially over time or due to a recalibration of the tracking system is easier. For these reasons, we recommend always measuring the starting position, if possible, with a tracking system.

The VE (or just the target) could also be rotated around the location of the viewer so that the participant is always facing directly at the target regardless of how the experimenter oriented her at the starting point. However, this approach is problematic in tight spaces, and it is not easy to implement in an analogous real-world study. If space is limited, for example, and an experimenter points the participant ten degrees to the right, the participant may collide with or get close to a real-world obstacle during the walk. Since participants are instructed to walk toward the target, we encourage participants to orient themselves to directly face the target if necessary.

Since the virtual environment can be easily translated, we recommend using at least three different initial virtual positions separated by approximately one meter because it discourages the use of environmental objects as relative distance markers and forces the participant to pay attention to where he is in the world and what she is doing (Kuhl et al., 2009). For example, if four target distances are used, the participant might view a target and recall that a similar target was also at that location because it is near some landmark in the virtual space. Although three initial virtual positions may be used, it is possible to only use one real-world starting location. This can be implemented by sometimes translating the virtual world and the target forward or backward. All starting positions should be close enough to each other to be convincing to the user that the experiment might have also brought them back to the same starting position in the real world. If a real-world version of the experiment will be performed, then the experimenter may want to consider changing the real-world starting position randomly for both the virtual-world and real-world versions of the experiment so that they are consistent. In this case, the software would need to place the target at the appropriate distance based on the randomly selected starting position. Section 4.4 quantifies the importance of this random choice.

5.1.2 Measurement

The point in which a participant ceases walking is the final position, and the distance between the initial real position and the end of the walk represents the perception of distance of the participant for the current trial. Figure 20 illustrates two methods to measure the distance. The hypotenuse distance is the distance from the initial

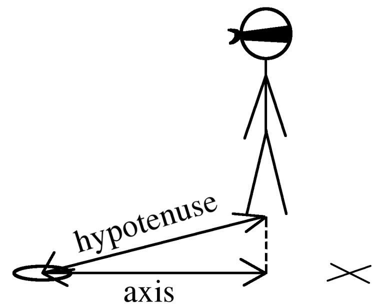


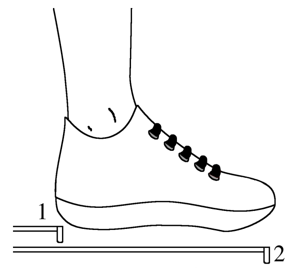
Figure 20: Starting from the initial position (left side), the hypotenuse distance is directed toward the final position of the participant while the axis distance is directed toward the target.

real position directly to the final position. The axis distance is the distance between the initial real position and the final position on the axis pointing directly to the target. The hypotenuse distance considers veer, is the more straight forward method if a measuring tape is being used, and may be a more accurate indicator of how far a participant walked. The axis distance does not consider veer, works well when a measuring tool is not being used and the distance walked is estimated by, say, counting floor tiles, and may be a more accurate indicator of how far participants believed they walked since people believe they walk straight even when blindfolded. Philbeck and Loomis, (1997) noted that participants do not veer markedly, so both measurement types will likely be similar. Using either measurement method, or both, is acceptable, but we recommend using the chosen method for all experiments for consistency and reporting the method used.

The method of measuring the walked distance depends on the experiment. Since virtual-world experiments use HMDs, which are tracked by a tracking system, and the location measurement can be programmed to consider the offset between the tracking dots and the eyes, the simplest measurement tool to measure the final position of virtual-world participants is the tracking system. It is worth mentioning that the tracking system can monitor other things, such as look direction, look duration, body sway, and more. Unless conditions dictate measuring excess motion, we recommend only measuring the final position for simplicity.

Some REs may not have a secondary tracking system, and real-world experiments often do not use any tracking system. In these cases, a traditional measuring tool, such as a measuring tape, wheel, or laser, that can accommodate through the safety meters in case of overestimating participants is easy to use. If applicable, counting floor tiles that are of known dimensions may be an even faster method of measuring the distance walked. Additionally, we recommend measuring to the same position on the foot as was aligned on the tape, either directly if using the hypotenuse distance or to the place it would be if using the axis distance. If the balls of the feet were aligned on the starting position tape, as per our earlier recommendation, measure to the balls of the feet, as shown in Figure 21a. If the heels or toes were aligned on the tape, then measure to the heels or toes, respectively, as illustrated in Figure 21b. Finally, some participants may stop with their feet offset by an inch or more. In this

(a) We recommend measuring to the balls of the feet.



(b) 1) Measure to the heels if the heels were on the tape. 2) Measure to the toes if the toes were aligned on the tape.

Figure 21: Measure to the same foot position as was aligned on the tape.

case, the eyes of the participant may not align with either foot. The experimenters should have a plan for how to handle this case or instruct participants to have their feet together when they try to stop at the target.

5.2 Participants

To perform a blind direct walking experiment, participants need to be able to walk unaided and unburdened without vision, hear and follow verbal instructions over the masking noise, and have normal or corrected-to-normal vision. Since participants need to walk, participants who use wheelchairs, crutches, canes, or otherwise have difficulty walking should not participate. High-risk individuals, such as the elderly or those with inner ear or other balance problems, may require additional walking and balance tests if not excluded from the study. Since blindfolded walking is generally unfamiliar, it may take some practice for unsteadiness to decrease. Furthermore, certain footwear, such as flip flops and high heels, may cause unsteadiness. Secured footwear, such as snugly laced or Velcro sneakers, is steadier and safer for use during experiments. Bare feet may also be permissible if the RE is clean and free of debris, but this presents an injury risk. In addition to communicating the walking requirements to the participant, the experimenter can also observe the participant walking during a blind walking practice session (see Section 6.5 for more) and end the experiment if the experimenter feels it is unsafe to continue. Similarly, if the experimenter discovers that the participant has trouble hearing instructions during the experiment, the experiment should also be halted for safety.

We use a Snellen Eye Chart Test to test eyesight. Testing eyesight is important to ensure that all participants can properly see the target and meets the baseline requirements of the experiment. The eye chart may be purchased online, which guarantees using the correct standard and opacity of paper. The test entails potential subjects correctly reading the 20/20 line, or 20/25 line if deemed acceptable, either with or without glasses or contacts when standing 20 feet away from the chart. The glasses of those who have them must fit under the HMD in order to avoid damaging either the glasses or the HMD. Completing the test under the same lighting conditions will help standardize the test across all potential subjects. Further, it may be necessary to test for stereoblindness, or the inability to fuse the separate images from each eye into one coherent image. We refer the reader to Richards, (1970) for a more thorough discussion about stereopsis and stereoblindness. The random-dot stereogram test, which entails a potential subject donning a pair of stereoscopic glasses, such as those pictured in Figure 22, and viewing a random-dot image, like the one pictured in Figure 23, works well when testing for stereoblindness. Creating a random dot image

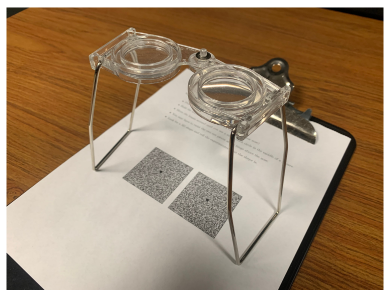


Figure 22: A pair of stereoscopic glasses positioned over a random dot stereogram.

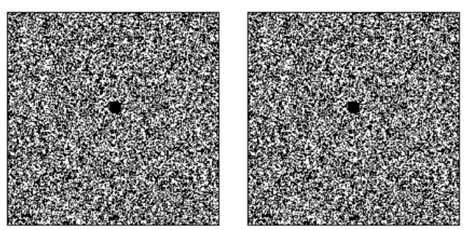


Figure 23: An example of a random dot stereogram.

is simple. First, create an image of random dots for the left eye. Next, duplicate that image for the right eye, isolate a particular section of the dots, and move them slightly to the right. We find that a plus sign works well since it is a shape that is easily identifiable but not one a person would generally guess, like a square, circle, or smiley face. Finally, fill the void area of the right-eye image with more random dots. In our image, we also added a black circle in the middle which the participant should see as a single dot when viewed correctly. When printed, the centers of the two images should be approximately 5.5 to six centimeters apart. Those who can see the three-dimensional image are not stereoblind while those who cannot are stereoblind. To our knowledge, no research has been done that studies the effects of stereoblindness on distance perception using blind direct walking. Visual tests other than those described may exist, but we find that these are simple and quick to perform.

Obviously, if a potential subject does not meet the above conditions, the subject may not participate in the study. However, some potential subjects apply to participate on the basis of class credit. Depending on institutional policies, there are several avenues that may allow the rewarding of class credit. One option is to have the disqualified subject complete an alternate task. For example, the disqualified subject could read a short related paper and answer some questions. Another option is to allow the disqualified subject to participate in the experiment, with accommodations as necessary, but throw out the data. This method allows the experimenter to test the experiment and make adjustments if needed. Other institutionally permissible actions could be taken instead. Furthermore, disqualified subjects who did not apply on the basis of class credit may also expect the agreed upon compensation for their time.

Other factors to consider include various group differences, such as sex, age, effort, and more. Though, these differences are rarely considered. The most commonly considered group difference is sex, but even that goes largely ignored. From the same 30 paper survey as before, only three papers accounted for sex-based differences in their results, though none of them found it to be a significant contributor. However, across all 30 papers, only half stated the sex breakdown in their experiments. Of those, 55% of all participants were male, and the majority of experiments used significantly more male than female subjects. This could be due to male-dominated interest in the topic or to the institution having a larger male subject pool. Either way, sex-based differences, even though they are the most commonly considered group difference, are clearly not considered often. This is especially troubling based on the results of the sex-based differences study done by Norman et al., (2018). In their real-world egocentric bisection study, they found that males judged distances more accurately than females across all of the real environments used. They noted that differences between the sexes have been difficult to analyze since one sex is often more significantly represented in experiments than the other, which may degrade the validity of results because sex-based differences may be more significant than previously thought. Balancing experiments for sex would better represent the total population anyway.

Another group difference is age. It is widely known that visual and stereo acuity decrease with age, so there may be significant differences in distance perception between older and younger adults. Unfortunately, the results of past studies (Bian and Andersen, 2013; Norman et al., 2018; Schubert et al., 2016) have reached no conclusion. Bian and Andersen, (2013) found that though older and younger adults use various cues similarly, older adults made more accurate distance judgments than younger adults in a real-world experiment using verbal reports and blind rope pulling. Contrarily, Norman et al., (2018) found that younger adults performed a real-world bisection task more accurately than older adults. Contrarily to both of these, in virtual alignment and distance focusing tasks on a three-dimensional display monitor, Schubert et al., (2016) found that both younger and older adults perform similarly. The reasons for the stark disparity in these results could range from the differences in tasks performed to luck of the draw from the participant pool. However, to our knowledge, no experiments have studied age differences when either blind direct walking or a HMD were employed, nor did any of the 30 surveyed papers account for age in their analyses even though some of them used subjects between the ages of 20 to 60 or older. Since most of these types of experiments occur at universities, the participant pool is dominated by students between the ages of 18 and 35. Further, based on popular online entertainers and the target audiences of certain entertainment advertisements, the main consumers of VR technology seem to be older children and younger adults. Because of the university population, we cap our participant age to 35. However, if older adults are allowed to participate, Bian and Andersen, (2013) recommend testing their cognitive and physical abilities to ensure they are fit to participate.

Another difference is the amount of effort perceived to complete a task. This could be physical effort, such as from carrying a heavy load, climbing a hill, being overweight, in pain, or unwell, or mental effort, such as the result of being tired, stressed, or depressed. The effort hypothesis, or the apparent elongating of distances when under some burden, has been extensively tested in terms of physical burden (Hutchison and Loomis, 2006a; Hutchison and Loomis, 2006b; Proffitt et al., 2003; Proffitt et al., 2006a; Proffitt et al., 2006b; Witt et al., 2004; Woods et al., 2009). Specifically, Proffitt et al., (2003) discovered that subjects wearing a heavy backpack viewed distances as longer than unencumbered subjects, and Witt et al., (2004) added that this elongation effect is seen only if the next intended action followed from how the distance was to be judged. Hutchison and Loomis, (2006a) contrarily found no effect of effort when reexamining Proffitt et al., (2003). After arguing over methodological differences (Hutchison and Loomis, 2006b; Proffitt et al., 2006a; Proffitt et al., 2006b), Woods et al., (2009) carefully replicated the studies by Proffitt et al., (2003) and Witt et al., (2004) and found no effect of effort on blind walking or blind throwing except when instructions to the participant specifically mentioned to use nonvisual factors when judging distances. This suggests that the instructions are crucial to an experiment. Firestone, (2013) vehemently opposed the effort hypothesis citing significant inconsistencies and disproportionalities between perception and physical ability, the size of the manipulation and the size of distance judgments, affirmation convenience and real-life experiences, and cognitive influences and perception. Clearly, the debate over the effort hypothesis has led to few conclusions regarding physical burden and no conclusions regarding mental burden. We err on the side of caution and exclude those who, say, ran to the experiment, but we leave it to the experimenter to consider the physical and mental burdens of participants.

Other group or individual differences that exist but, to our knowledge, have not yet been examined in the context of blind direct walking include familiarity with distance judgment experiments and VR in general, mood, illness, impairments, interests, or nearly any other characteristic that an individual might have. Some research on inherent traits, such as personality and childhood, have been studied pertaining to distance perception in general, and we refer the reader to Renner et al., (2013) for a summary. Indeed, it is impossible to standardize or consider all of these factors, but a large enough sample size may help account for some of them. Unless the above factors are being specifically studied, any eligible person who would like to voluntarily participate during the testing phase should be allowed to take the qualification tests, and ideally, the experiment would be advertised in a way that attracts a diverse population.

More than just accounting for individual and group differences, a large enough sample size aids in studying trends and improving statistical power. The average sample size from the 30 paper study is 30 participants per paper. Some papers used as few as six to ten participants, which may not have enough statistical power. Other studies used 100 or more participants, which would aid in trend analysis and have good statistical power, but it would take a long time to complete an experiment with this many participants. There are many ways to determine a good sample size. We use around 15 subjects per condition, or 30 subjects if two things are being compared, since its a common number and we find it to hold enough statistical power. Ultimately, it is up to the experimenter to determine the needed number of subjects.

5.3 Safety and Cleanliness Procedures

The safety and well-being of participants is above all else. The safety and cleanliness of equipment is also important to subject safety and the overall success of an experiment. Even so, these details are missing entirely from research papers as they have little to do with the research itself. Here, we present clear guidelines to keeping participants and equipment safe and clean.

Perhaps the most obvious concern is the physical safety of participants, including fall prevention, mitigating the risk of simulator sickness, and reducing the chances of spreading germs. Walking while blindfolded increases the risk of tripping and falling since a participant cannot see obstacles or other hazards. The easiest and safest way to reduce the risk of tripping is to remove all obstacles, such as tables, chairs, hardware pieces, and other equipment and objects, from the walking area and ensure the floor is free of debris and, in the case of hard floors, dry. Further, walking next to the participant allows the experimenter to catch the participant by the shoulders if necessary. We find that this is helpful in other areas, as will be explained shortly.

Additionally, recall that it is important to prevent participants from accidentally brushing against surfaces as that could impact the confidence of the participant and skew results; see Section 4.1 for more. These simple steps greatly increase participant safety.

Currently, most contemporary HMDs are wired, which presents another tripping hazard. If the wires snag on something, it may be the instinct of the participant to pull on the cables, which could damage them. Wireless HMDs may become the norm in the near future, but for now, the wires must be dealt with to protect the subject and the HMD. There are several methods for cable management. One option is to put the computer on a wheeled cart that can be pulled alongside the participant, but this is cumbersome. Another option is to connect the HMD to a laptop placed inside of a backpack along with any cable slack. However, if the backpack is held by the experimenter, participant movement may be impaired, but if the backpack were held by the participant and the effort hypothesis has merit, the backpack would add an undesirable variable into the research. Further, keeping the laptop charged while inside the backpack may be difficult. Instead, we find that walking next to the participant and holding the cables, as illustrated in Figure 24, is a good way to mitigate risks to the participant and the HMD. By holding the cables and allowing

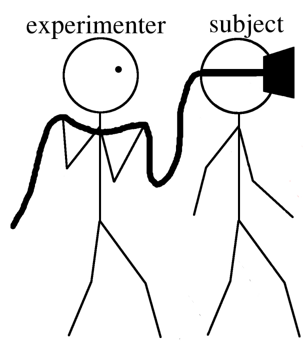


Figure 24: Walk next to the participant while holding the HMD cables out of the way to both aid the participant if necessary and protect the cables.

some slack, the cables will not pull on the HMD, so participants will not be tempted to yank on them. Moreover, the experimenter can monitor the trailing cables and more gently release snags. As mentioned, this also puts the experimenter in a good position to stabilize the subject should the need arise and prevents the subject from using the voice of the experimenter as a localization cue.

If the experimenter is holding the cables, then the cables are uncontained during the restarting phase, reintroducing a tripping and wrapping hazard. We find that laying the cables to one side and always having the participant turn away from them, as illustrated in Figure 25, is a good way to reduce this risk. For example, say the cables are placed to the left side of the participant after completing a walk. In this case, the

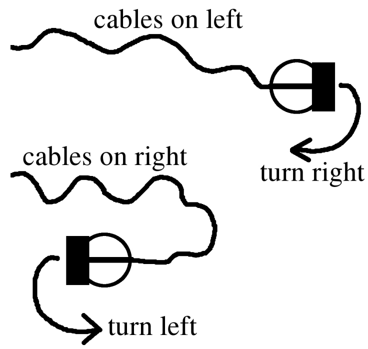


Figure 25: To avoid tripping hazards, have participants always turn away from the cables.

participant would turn right to avoid the cables. Once the participant returns to the initial real position, the cables would be on the right side of the participant. In this case, the participant would turn left to avoid the cables. The subject would now be on the starting position and not be wrapped by the cables.

Being in a HMD for any length of time comes with the risk of simulator sickness, motion sickness, headaches, dizziness, and nausea (Jones et al., 2011). A good way to monitor how a participant is feeling is to ask him. Asking suggests to a participant that the experimenter is genuinely concerned about her well-being and that it is acceptable to feel unwell. The participant may therefore feel more comfortable answering the question honestly or stating that he does not feel well without being asked. If a participant is not asked, she may feel less inclined to speak up about any discomfort and pressured to finish an experiment even if he feels unwell. This may lead to the subject rushing through an experiment in order to take off the HMD sooner or working more timidly due to feeling sick, both of which will bias the data of the participant. However, asking a participant about her state of being does not guarantee an honest response. He may lie in an effort to receive compensation or because she really wants to complete the experiment. Luckily, participants often cannot hide the symptoms of simulator sickness, such as excessive sweating, heavy breathing, stumbling, shaking, and gagging. It is up to the experimenter to monitor participants for such symptoms and decide what to do with the participant and the data.

Note that there are several instances where the HMD screen may flash, including during blanking and turning on the display, during glitches, or as a feature of certain environments. Participants, especially those with a history of epileptic seizure, may want to know this during the informed consent process. If epileptic participants feel comfortable carrying out the experiment and it is allowable under IRB guidelines, it is alright to include them, but it may be a good idea to monitor such subjects for signs of discomfort.

At the time of writing, the world is in the middle of the COVID-19 pandemic, so illness is at the forefront of concern. There are many options to slow the spread of illness. Since all virtual participants will be wearing the same HMD, our laboratory offers the participant a disposable plastic shower cap to wear over his hair and under the HMD. This is a cheap and effective way to prevent the spread of lice. Further, all participants will likely be touching the same surfaces. We find that offering participants hand sanitizer or disposable latex-free, in case of latex allergy, gloves are good ways to prevent the spread of germs. Disposable face-covers designed for HMDs are also available from online retailers. These provide a paper surface between the area around the eyes and the HMD. Additionally, as the pandemic is ongoing, our institution requires that all personnel and visitors fill out a symptom questionnaire before coming to campus and that everyone wear a mask at all times while on campus. There are many other effective alternatives to preventing the spread of disease, but we find that these work well.

In addition to offering participants items to help prevent the spread of disease, we clean all relevant equipment between uses. Cleaning instructions can often be found in the user guide or other documents for a particular piece of equipment. Unfortunately, some products may not come with specific care and cleaning instructions. Depending on the product, carefully cleaning it with a dry cloth or non-abrasive cleaning solution may be acceptable. If the product could easily be damaged, however, contacting the manufacturer is the fastest way to find out the recommended cleaning procedures. To give a few examples of our cleaning methods, we use camera or eye glass cleaning solution on the HMD lenses as they are made for lenses with a coating and are safe for the HMD lenses. For hard plastics, such as the outer portion of the HMD, controllers, keyboards, and the outer portions of headphones, we use a soft cloth dampened with isopropyl alcohol. Soft wearable items, such as body suits, can be laundered with hypo-allergenic laundry detergent. Finally, large items that cannot be laundered, such as chairs, are wiped with some cleaning wipe or vacuumed. Other things that may need to be cleaned include blindfolds, harnesses, clipboards, writing utensils, HMD straps, rails, hand-held items, and anything else that was used during an experiment.

6 Pre-Experiment Procedures

6.1 Consent Forms

Participants must give their informed consent by signing the requisite forms before any pre-screening or experimentation may begin. These forms typically outline minimum requirements for participation, potential physical and psychological dangers and side effects as a result of participating, general expectations, compensation, a statement of anonymity, and a statement informing subjects that they may quit at any time. For the purpose of anonymity, these forms generally are the only things linking participant names to corresponding identification numbers and are stored in a safe, locked, or encrypted location until they can be destroyed.

6.2 Instructions

After the consent forms have been signed, perform pre-screening procedures and present instructions to the participant. We present instructions in both written and verbal format with a demonstration. The dual-format presentation seems to help overall understanding of instructions and offers subjects the opportunity to ask questions. Though the experimental requirements will determine relevant instructions, egocentric distance judgment experiments have some general commonalities, including:

1) What the participant will be doing - walking without vision to a target on the floor.

2) An explanation of what the target(s) look like.

3) How to look a the environment, such as by swiveling the head or taking note of where the subject is in the environment.

4) How long to look at the environment. We typically allow the participant to look as long as she needs to obtain a "good image" of the environment, or until he can "see" the environment even with her eyes closed.

5) What not to do, such as shuffle the feet, bend at the waist, use cognitive factors like counting the number of steps taken, counting floor squares, or performing mathematical calculations, and pulling on the HMD cables.

6) Alert the experimenter when ready to walk, and have the participant either close his eyes or put on the blindfold, if applicable. Virtual participants should also be made aware that the HMD will be blanked.

7) How to walk. We generally use strong and descriptive language like "naturally" or "decisively" when describing how to walk to discourage participants from acting timidly or cautiously.

8) Stop walking when the participant believes she is standing on top of the center of the target and note that he will not feel the target under her feet.

9) How to return to start. We normally tell the participant to keep his eyes closed and blindfold on, if applicable, and lead the subject back to the starting position by the shoulders. Other restarting methods are discussed in Section 7.3.

Sometimes different experiment-specific instructions must be given or different treatment conditions are given slightly differing instructions, term definitions, or language and wording (Alekseev et al., 2017; Silva and Santos, 1984; Woods et al., 2009). For example, the term "good image" in the above list could mean different things to different people, so we specifically defined it for participants to mean "the ability to see the environment and target clearly in your mind with your eyes closed." Another example is the term "distance." Though this term seems simple enough, when coupled with a slight manipulation in the effort of a task and non-neutral experimenter-participant interactions, the definition of "distance" was found to have significantly changed the results of a blind throwing experiment when "distance" was defined as one of "objective" distance, "apparent" distance, or the distance as determined by use of "non-visual factors" (Woods et al., 2009). Because of this, Woods et al., (2009) recommend that experimenter-participant interactions be limited and instructions be presented as neutrally as possible. Contrarily, when the definition of "distance" was manipulated but the details of the real-world verbal report task were not, the actual definition made no difference (Silva and Santos, 1984). Readers may be interested to know the specific instructions or definitions given to better understand procedures or do their own comparison experiments, so including differing instructions or definitions in papers would be helpful. Additionally, past (non-VR) researchers have recommended that instructions be clear and orderly, standardized with enough context as to not be misinterpreted, use appropriate language, be presented in multiple different formats, allow for practice, be detailed but succinct, delivered authoritatively and neutrally, and possibly be pre-tested before put into practice (Alekseev et al., 2017; Baggett, 1983; Bigoni and Dragone, 2012; Rashotte et al., 2005).

We use the verbal format and demonstration as an opportunity to give subjects a better idea of what they need to do. However, we are careful to give a simplified version of the procedures as to not bias participant actions. For example, when explaining how to look at the environment by only rotating the head, the experimenter could stand in place and rotate only her head. As another example, when explaining the target and walking procedures, the experimenter could stare at an indeterminate spot on the floor, close his eyes when giving the ready signal, and take a few confident strides toward the imagined target. This demonstration further offers participants a chance to ask questions and gain clarification about anything they find confusing.

6.3 Pre-Questionnaires

The use of pre-questionnaires, or a set of questions the participant answers before beginning an experiment, is entirely up to the experimenter. Obviously, the questions asked are experiment-specific and will change depending on the research, if any are needed at all. Generally, questioning prior to beginning an experiment is focused on health and safety, which is completed during the informed consent and instructions procedures.

6.4 Putting on Wearables

The time for participants to put on any wearable items is after the subject has read and filled out all requisite forms and is deemed able to participate but before any practice or trials begin. Wearable items include gloves, shower caps, body suits, motion capture gear, and anything else that is wearable except the HMD, which is part of the actual virtual experiment and not for practice or information purposes. Further, real-world participants will never use a HMD. We generally let the participant handle this step herself and assist only when necessary. Additionally, we hand all participants a blindfold at this stage because our next step is allowing the participant to practice walking while blindfolded. We generally use a sleeping mask called the Mindfold because it is comfortable and provides near complete darkness, which is important because light peeking through gaps in the blindfold or HMD may aid in distance perception (Jones et al., 2013). If, on the other hand, the participant is wearing glasses, the Mindfold cannot be used. Instead, we use an old pair of chemistry laboratory goggles that were roughed up with sandpaper and spray painted black. Covering them with thick black construction paper, fabric, or other opaque material also works. This is not perfectly opaque, but it should suffice for walking practice and entering the laboratory. Other blindfolding methods exist, but the general consensus is that the blindfold should block out as much light as possible.

6.5 Practice Blind Walking

We recommend allowing the participant to practice walking while blindfolded. Most people are used to walking with vision, so walking without vision is foreign and intimidating. Allowing subjects to practice will help build confidence so that they walk as naturally and decisively as possible during the experiment and trust with the experimenter to assuage concerns about colliding with obstacles. Before this practice begins, we explain to the participant to walk naturally as if he is walking with vision. This is the most desirable form of walking, so we want to encourage participants to walk in this manner. We also explain that the participant should start and stop walking only when instructed and that we will touch her shoulders when it is time to turn. We explicitly state that there will be several stops during the practice. Recall that any indication of doing something wrong may negatively effect the participant during the experiment. When walking without vision, people tend to veer along a curve rather than walk a straight path. If practice is taking place in a narrow hallway, this tendency to veer means participants may collide with the wall or have to be forcefully stopped more often, which could result in the aforementioned side effects. However, the expectation of stopping many times provides an excellent cover story and avoids the negative consequences of correction. Even if the practice will take place in an open area, it may still be a good idea to warn the participant about stopping multiple times just in case. In this light, we follow the participant as closely as if holding the HMD cables, as illustrated in Figure 26. When given the

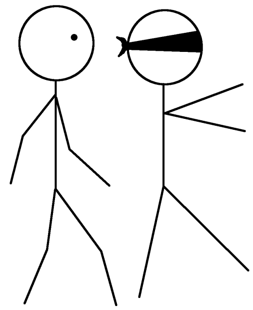


Figure 26: Follow the participant as closely as if holding the cables out of the way while practicing walking without vision.

stop instruction, it may take a second for the instruction to register with the subject. If given too late, the participant may not stop in time and collide with an obstacle, so walking this closely with the participant puts the experimenter in prime position to grab the shoulders of the subject for an emergency stop.

When walking without vision for the first time, people are typically timid. Participants may walk slowly and with their arms out in front of them as protection against collisions. Although this is a perfectly natural response, this clearly does not constitute "natural walking," so participants may need a gentle reminder to walk naturally. An example of the difference between natural and timid walking is shown in Figure 27.

(a) Timid walking. **(b)** Natural walking.

Figure 27: An example of the difference in appearance between timid and natural walking.

Though the general consensus is that allowing participants to practice walking without vision is good, there is no agreement on the duration. Very little research has been conducted on blind walking practice. Jones et al., (2008) conducted a pilot study showing that participants felt comfortable walking without vision after five trials of walking in a hallway. However, they do not define what "felt comfortable" meant. They also do not define what a "trial" of blind walking is nor indicate the length of the hallway or the time it took to complete the trials. Philbeck et al., (2008), on the other hand, studied the effects of blind walking practice. In their first experiment, participants practiced either with or without vision for three minutes before completing a real-world blind direct walking task. Those that practiced without vision overestimated distances more than those who practiced with vision underestimated the distances, but few trials were done to avoid adaptation. In their second experiment, participants performed several rounds of blind marching, blind direct walking, and visually guided walking. The first round of blind direct walking included 37 trials to encourage adaptation. They found that participants did eventually adapt to blind walking and started overestimating distances, though increased walking pace lengths, velocities, and task familiarity were not contributing factors. Participants indicated in a post-questionnaire that they started focusing on different parts of the task or attempted to compensate for earlier trials. Further, the authors speculate that decreased idle times between trials as the experimenters became more efficient at running the experiments may have caused some bias as evidenced by later participants more significantly overestimating distances than earlier participants.

Clearly, allowing some adaptation to blind walking helps familiarize subjects with the task, but too much adaptation or loss of participant focus may be detrimental to the experiment. In terms of the duration of practice, Philbeck et al., (2008) recommends three minutes with periodic breaks during the experiment to avoid overadaptation. We normally allow subjects to walk down and back between two hallways once or twice, sometimes more if a participant seems particularly hesitant. Future work on time-based, distance-based, and walking speed-based blind walking practice may shed more light on what the ideal amount of practice is.

Once blind walking practice is completed, some experiments (see Philbeck and Loomis, (1997), Proffitt et al., (2003), and Steinicke et al., (2009) for examples) conducted practice trials either in the same location as the pre-experimental procedures or in the main laboratory space in such a way that it was obvious to participants that these were practice trials. There is no reason for participants to know about practice trials or for the practice trials to be completed during the pre-experimental procedures. Instead, we guide the participant to the experiment location with the blindfold on. When the experiment takes place in an indoor laboratory, we dim the lights to prevent participants from seeing the laboratory and to block light from leaking into the HMD.

Once the participant is inside of the laboratory, we tell her to close his eyes, take off the blindfold, and put on the HMD, if applicable, since virtual-world participants should never see the laboratory before the experiment. This is the appropriate time for virtual-world participants to open their eyes and to complete any last hardware calibrations. Obviously, real-world subjects do not use a HMD, so they may simply open their eyes. We then play the masking noise through the appropriate headphones and music player, and we are ready to begin.

7 During Experiment Procedures

7.1 Other Helpful Procedures

During the experiment, there are a lot of things the experimenter has to keep track of, such as checking the well-being of the participant, managing cables, taking measurements, making sure all hardware is working properly, and communicating with the participant. In this section, we suggest some things that should make these tasks easier. Also, let the term "phase" imply the current place within a trial or experiment, such as viewing the targets, closing the eyes, walking, stopping, measuring, and returning back to start.

Recall that even though subjects receive instructions, some may still do strange things during the first few trials of an experiment, such as holding onto the HMD, walking timidly, stutter stepping, or forgetting something due to possible misunderstandings or from not having paid attention during the instructions phase. Thus, as explained in Section 4.4, we recommend having the first two or three trials be practice and simply throwing out the data for those trials. We also use the very first trial as a reminder of the instructions and walk the participant through it step-by-step. Additionally, after each phase during each trial, we say something along the lines of "Good," "Alright," "Okay, next," or some other phase-appropriate statement in a neutral-pleasant tone. This is a good way to tell the participant that the current phase or trial is over and the next phase or trial is about to begin. Further, stating these transitional phrases in a neutral-pleasant tone avoids presenting any implicit feedback about how the participant did.

Since the experimenter will likely be following the subject, constantly having to run over to the keyboard to go to the next phase in a trial is inconvenient. There are many options to make this easier. One option is to have two experimenters in the room, one to follow the participant around and the other to press keyboard buttons. The disadvantage to this is scheduling experimentation times that work for both experimenters. Another option and the one we use is utilizing a slide show clicker where buttons are mapped to keyboard buttons, as shown in Figure 28. This is convenient since only one experimenter need be present, but it is only applicable to virtual-world participants.

Regardless if the keyboard or a clicker will be used, the buttons must obviously correctly change the phase of a trial. There are many ways to accomplish this. One option is to have one button do everything. This is very easy to program and does not require a lot of thinking during the trials. The danger lies in double tapping the button; there is no going back to the correct phase. In this light, it may be advantageous to have at least one additional button that allows the experimenter to go back in case of an accidental double tap. Another option is to map one button to

Figure 28: This example clicker has three buttons - the "next task" button goes to the next phase in the experiment, the "go back" button returns to the previous phase in the experiment, and the "action" button does an action, such as take a position measurement or blank or unblank the screen, depending on the trial phase.

one or two functions. For example, "B" could be to blank and unblank the HMD, "M" could be used to take a measurement, and "Enter" could go to the next trial. If each key were programmed to work only when in the correct trial phase, this would almost, though not entirely, eliminate the double tapping danger. For example, double tapping the blank and unblank button may cause the screen to flash or be unblanked (blanked) when it should be blanked (unblanked), which could give the participant unintended feedback if he did not close her eyes; though, this risk can be mitigated by showing what the subject sees on a computer screen visible to the experimenter. Another example is if the measurement button is pressed multiple times, the measurement will be taken multiple times or overwritten. Sometimes it is necessary to take multiple measurements if the participant stutter steps, and we take the most recently measured position as the final position for a trial, so this is not a big deal anyway. Other reasonable steps may be necessary to mitigate the double tapping danger.

Another step that could be taken includes sounding a soft ding when switching phases. This alerts the experimenter that the trial phase was changed, which allows him to go back if necessary, and provides a way for the experimenter to notice if a key press was not registered by the computer. This audio feedback cannot be heard by the participant over the masking noise. Further, other soft sounds could be employed, such as one indicating that the experiment is halfway done. However, one sound the experimenter does not want to hear is someone banging on the laboratory door, if applicable, during an experiment. To mitigate this, we hang the sign shown in Figure 29 on the outside of the door.

7.2 Real-World Comparison

Most of the above explanations pertain to specifically virtual-world experiments. Although real-world experiments are largely the same, there are some important differ-

Figure 29: The door sign that warns passersby to be quiet.

ences. Here, we summarize these similarities and differences.

First, we will discuss the similarities. Just as with virtual experiments, real-world participants should not view the laboratory space before trials begin as to not gain any unintended feedback. Thus, we recommend conducting pre-experimental procedures in a separate area. During the instructions phase, most of the instructions should be largely the same, which the exception of HMD-specific items. For consistency and if able, it may be a good idea to block windows so weather does not impact lighting conditions. Outside lights may also add to glare on shiny floors, which would make the targets hard to see. During the experiments, participants would listen to the same masking noise file as virtual-world participants for consistency. Just as for virtual-world experiments, the experimenter would follow the participants and turn them as if there were cables present, even though none are, and ask participants periodically if they feel alright, even though simulator sickness is not of concern. This is done to keep procedures between the two worlds as similar as possible. When measuring the distance walked, measure using the same method as for virtual-world participants - hypotenuse, axis, or both - for consistency. During the return trip, we lead the still blindfolded real-world participants back by their shoulders, just as with virtual-world participants, because it is easy and fast.

Now, we will discuss the differences. Virtual-world participants will never see the laboratory space before or during the experiments, but real-world participants must see the laboratory space during experiments. Using only the laboratory lights will maximize consistency, but it may not eliminate glare. If glare is too strong, turning the experiment around generally helps. Virtual-world participants will use a HMD while real-world participants use only a blindfold, so any instructions pertaining to the HMD are not relevant for real-world participants and may be skipped. Further, the same blindfold as was used during the pre-experimental procedures will suffice. Real-world participants will generally not be connected to any sort of tracking system as no HMD will be used. As such, targets must be manually chosen, measured, and placed. Some ways to choose the targets include rolling a die or randomly picking a target out of an opaque box. One way to place targets is to measure the target distance before each trial, but this will add significant idle time. Instead, we mark all possible target placements with tiny marks that are not visible to the participant. The targets must also be removed before the participant might step on them. One experimenter can do this after the participant puts on the blindfold, such as in Figure 30, or if

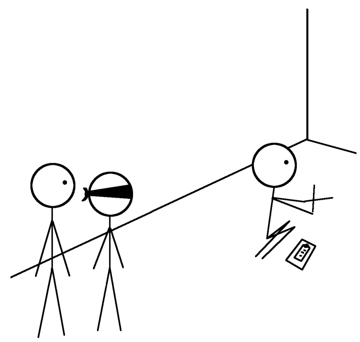


Figure 30: One experimenter aids the participant while the other sets up and removes targets.

fishing line is attached to the target and looped through a straw or hoop anchored to the floor or wall, it is possible to create a system where the experimenter can quickly pull the line to move the target out of the way, such as in Figure 31. For consistency

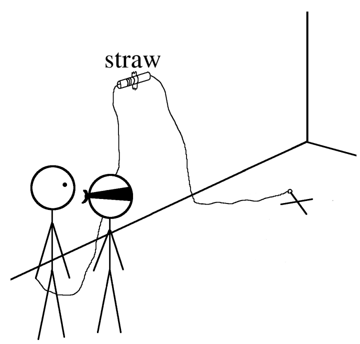


Figure 31: Some creativity is necessary if only one experimenter is present. To reduce idle time between viewing and walking, the experimenter attached the target to a rope that moves the target when pulled through a straw taped to the wall.

with HMD studies, the experimenters may wish to ensure that participants never see a person in front of them while performing the experiment. Additionally, the distance walked must be manually measured with a measuring tape, wheel, laser, or some other method. If known distances are marked in a way that is not visible to the participant, the marks can also be used as landmarks to measure the walked distance from. Real-world experiments are often done with two experimenters so that the second experimenter can aid in target placement and distance measurement while the first experimenter can focus on communicating with and monitoring the participant. This provides efficient transitions between viewing and walking, as well as reduces idle time between measuring, recording, and restarting. It is possible for one experimenter to conduct trials alone with a creative touch to reduce some idle time, such as the string through the straw technique described in Figure 31. However, the idle time between restarting and viewing is limited to how fast the experimenter can change targets.

7.3 Return to the Initial Position

There are several methods of returning a participant back to the initial real position, including guiding her by the shoulders, verbal instructions, allowing full vision (Jones et al., 2011), allowing the participant to view the virtual space and initial virtual position (Ries et al., 2008), or by providing markers in an otherwise blank scene (Steinicke et al., 2009). We guide participants back by the shoulders with a blank screen and eyes closed because it is fast and efficient. Verbal instructions are less efficient and could be confusing. Allowing full vision provides unnecessary feedback. Allowing participants to view the virtual space and initial virtual position also provides unnecessary feedback, and Ries et al., (2008) strongly recommends against this method. Finally, providing markers in an otherwise blank scene is not applicable to the real world.

Recall that it is safest to turn participants away from the cables during the return trip, and we suggest doing this for real-world participants for consistency. We also suggest not turning participants 180 degrees and walking straight back. Though Figure 32 is an exaggerated representation, walking along a curve prevents participants from accurately counting their steps if they are doing so. Additionally, we suggest slightly

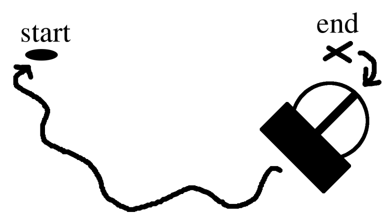


Figure 32: This is an exaggerated return curve the participant could take on the return trip. The actual return trip does not need to be this squiggly.

overshooting the initial real position, as illustrated in Figure 33, to further prevent feedback and to make fine-tuning the position easier.

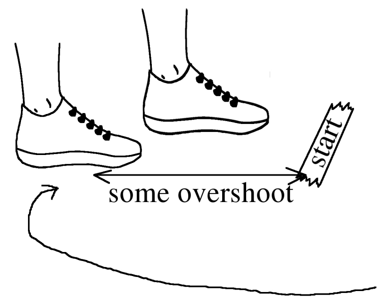


Figure 33: We suggest slightly overshooting the starting position and making adjustments as necessary.

8 Feedback

Much research has gone into studying the effects explicit feedback has on distance judgments. Employed feedback methods include some diagram comparing the estimate to the actual distance shown to the participant after a walk (terminal diagram) (Richardson and Waller, 2005), walking while viewing the VE (continuous visual) (Kelly et al., 2018; Kelly et al., 2017; Mohler et al., 2006; Steinicke et al., 2009), viewing the environment at the end of the walk (terminal visual) (Mohler et al., 2006; Nguyen et al., 2008), instructing the participant to stop once the target is reached (verbal) (Mohler et al., 2006), completing both RE and VE trials with RE trials occurring before VE trials with participants memorizing distances (ordering) (Ziemer et al., 2009), previewing the RE and then showing a replica of the RE as the VE (preview) (Kelly et al., 2018), and a combination of preview feedback and continuous visual feedback (Kelly et al., 2018). All forms of feedback did significantly improve blind direct walking estimates (Kelly et al., 2018; Kelly et al., 2017; Mohler et al., 2006; Nguyen et al., 2008; Richardson and Waller, 2005). In some cases, the improvements allowed near veridical estimates (Mohler et al., 2006) while other cases did not show quite as significant improvements (Kelly et al., 2018).

It is unclear if feedback influences cognition, perception, action, or some combination thereof (Kelly et al., 2018; Kelly et al., 2017; Mohler et al., 2006; Richardson and Waller, 2005). What is clear is that feedback comes in many forms (Kelly et al., 2018; Kelly et al., 2017; Mohler et al., 2006; Richardson and Waller, 2005; Steinicke et al., 2009; Ziemer et al., 2009), and it significantly improves distance estimates, though possibly not to the same extent depending on the manner of estimating distances (Kelly et al., 2018). Implicit forms of feedback, such as adaptation to the virtual-world and natural improvement over time (Jones et al., 2011; Jones et al., 2009; McManus et al., 2011), the size of the pre-experimental area as compared to the size of the experimental area (Kelly et al., 2018), stating how well a participant did after a trial, implying performance through tone of voice, facial expression, or changing transitional phrases, stating that people generally underestimate distances in VR before experiments begin, among other similar things, though they have not yet been tested, may supply this same bias. For this reason, it is important to communicate with participants using the same language and tone of voice for the whole experiment. Since unintentional feedback may skew results, it may be important to identify potential forms of feedback and minimize or eliminate them.

9 After Experiment Procedures

9.1 Post-Questionnaires

It is sometimes necessary to ask participants post-experimental questions in order to obtain more information about experimental experiences. Some famous post-questionnaires include the simulator sickness questionnaire (SSQ) (Kennedy et al., 1993) and the presence questionnaire (PQ) (Witmer and Singer, 1998). However, the reliability of PQs has been called into question. Usoh et al., (2000) tested two different PQs after a real and a virtual experiment. They stated that if the PQs were truly reliable, PQ scores for real-world participants should be higher than for virtual-world participants. They instead found that PQ scores were not significantly different between real-world and virtual-world participants, speculating that the nature of the questions and the overall context may have caused real-world participants to reinterpret the meaning of the questions in a way that made sense to them. This suggests that what is asked, how it is asked, and the context for which to base answers on is very important. Further, Slater, (2004) believes that, though the concept of presence itself is useful, PQs are simply ridiculous. He exemplifies this point by detailing a meaningless study on the colorfulness of a person's yesterday to show that experimenters may ask and analyze what they wish and receive the desired result. It is up to the experimenter to decide whether the SSQ and PQ are appropriate for their experiments. Often, these questionnaires are performed before and after the experiment to measure how responses changed for each individual.

Other experiment-specific post-questionnaires may be necessary. Unlike SSQs and PQs, experiment-specific post-questionnaires can be thought of as a means to both identify problems in the experiment and disqualify participants. We recommend asking the following three questions to all participants:

1) Did you notice anything strange during the experiment?

2) Were any of the experimental instructions confusing?

3) What strategies did you use to judge distances?

Sometimes participants believe certain malfunctions, including glitches in the VE or HMD, tracking errors, or ceasing of the masking noise, are features rather than problems and do not say anything during the experiment. Therefore, the first question acts as an alert to experimenters about possible hardware issues. Severe problems may bias the data of a participant and warrant significant troubleshooting. Note that though this question is specific to virtual-world participants, real-world participants should also be asked this, and all other questions, for consistency. The second question helps identify areas for improvement if a similar experiment were to be conducted again in the future. The third question can often identify when participants use strategies that were not allowed in the experiment, such as counting steps. The data of those who admit to using cognitive influences or otherwise biasing the data must be disqualified and thrown out. Other questions can be added as necessary, but we recommend using the post-questionnaire as an opportunity to identify problems with the experiment and to identify participants who did not follow the experimental instructions.

There are two main ways to present the post-questionnaire to participants. First, the experimenter could ask the questions aloud to the participant, who will then verbally answer the questions while the experimenter writes the responses. This allows for a more natural discussion where the participant may feel more comfortable asking clarifying questions and the experimenter can ask the participant to expand his answers. However, this is slow and the experimenter cannot set the next experiment up during post-questionnaire procedures. The second option is to give the questions to the participant and have her write his own responses. This is much faster and allows the experimenter to set up the experiment for the next participant. However, the participant may be less likely to ask clarifying questions, and the experimenter is unable to ask for more information from the participant. It is up to the experimenter to decide how to present post-questionnaires to participants.

9.2 Transfer Effects and Debriefing

IRB guidelines require that participants be debriefed about any intentionally misleading information after the conclusion of experiments. Some participants may ask about the research specifically at this point. Revealing too much information, such as how well the participant did, the average performance of other participants, and information about the conditions of the research, may compromise the research if the participant passes the information along, so obviously use discretion when divulging information. Additionally, during the post-questionnaire and debriefing stages, monitor participants for transfer effects and simulator sickness for safety purposes, if applicable. If any signs do show, they normally dissipate quickly, but some participants may need to wait until simulator sickness symptoms subside before leaving to reduce the risk that their symptoms interfere with other tasks, like operating a vehicle.

9.3 Reporting Results

Clearly, the last step in any research endeavor is analyzing the data. Since many distance judgment experiments in VR have been done, there are lots of examples in the literature for how to complete the data analysis. Most of the studies cited in the paper use some ANOVA, MANOVA, or t-test analysis method, depending on the design structure. Furthermore, as discussed above there are many instances where data must be thrown out. Few researchers devote much space to discussing the discarding of data because of outliers or other experimental reasons. We believe that it is a crucial part of the data analysis because it alerts readers and other researchers about potential issues or design flaws, things to look out for in their own research, and reasons why certain analyses could be skewed. When data is discarded, the researchers might want to analyze the data twice, with and without the removed data, and briefly mention how excluding the results did or did not change the overall results.

10 Conclusions

In this paper, we provide an overview of issues that those conducting blind direct walking studies should consider for both real-world and virtual-world experiments. By describing the process and making suggestions when appropriate, we hope to make it easier for more people to conduct similar studies. There are many things to consider when designing the experiments, including where experiments should take place, the virtual environment to model, which HMD, tracking system, music player, headphones, and communication method to use, how to properly calibrate the hardware, where to start the participant in both the real and virtual worlds, what kind of avatar to use, if any, how to keep participants and equipment safe and clean, where to complete pre-experimental procedures, which forms are required, what questions to ask participants, how to conduct instruction procedures, safety procedures, debriefing, and various other things. We attempted to make the list of procedural aspects as comprehensive as possible to help shorten procedure sections in research articles, but significant deviations or future developments in the field may require detailed explanations in research papers.

The field of VR continues to grow in both size and complexity. Distance judgment experiments continue to be a popular area of research as we continue to understand the perceptual mechanisms that drive it and how virtual reality technologies might prevent performance from matching real-world performance. As new developments arise, so too must new investigations to aid in our understanding of perception and how to effectively use this technology. This article documents and suggests procedures which standardize this research.

11 References

Achenbach, J., Waltemate, T., Latoschik, M. E., & Botsch, M. (2017). Fast generation of realistic virtual humans. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. VRST '17. Gothenburg, Sweden: Association for Computing Machinery.

Alekseev, A., Charness, G., & Gneezy, U. (2017). Experimental methods: When and why contextual instructions are important. *Journal of Economic Behavior & Organization*, *134*, 48–59.

Baggett, P. (1983). Four principles for designing instructions. *IEEE Transactions on Professional Communication*, *PC-26*(3), 99–105.

Bian, Z. & Andersen, G. J. (2013). Aging and the perception of egocentric distance. *Psychology and Aging*, *28*(3), 813–825.

Bigoni, M. & Dragone, D. (2012). Effective and efficient experimental instructions. *Economics Letters*, *117*(2), 460–463.

Buck, L. E., Young, M. K., & Bodenheimer, B. (2018). A comparison of distance estimation in HMD-based virtual environments with different HMD-based conditions. *ACM Transactions on Applied Perception*, *15*(3).

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8.

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

Cutting, J. E. & Vishton, P. M. (1995). Perceiving layout and knowing distance: The integration, relative potency and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Perception of Space and Motion* (pp. 69–117). New York: Academic Press.

Ellis, S. R., Adelstein, B. D., Baumeler, S., Jense, G. J., & Jacoby, R. H. (1999). Sensor spatial distortion, visual latency, and update rate effects on 3D tracking in virtual environments. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)* (pp. 218–221).

Erlebacher, A. (1977). Design and analysis of experiments contrasting the within- and between-subjects manipulation of the independent variable. *Psychological Bulletin*, *84*(2), 212–219.

Firestone, C. (2013). How "paternalistic" is spatial perception? Why wearing a heavy backpack doesn't—and couldn't—make hills look steeper. *Perspectives on Psychological Science*, *8*(4), 455–473.

Freeman, G., Zamanifard, S., Maloney, D., & Adkins, A. (2020). My body, my avatar: How people perceive their avatars in social virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–8). CHI EA '20. Honolulu, HI, USA: Association for Computing Machinery.

Gaines, D. & Kuhl, S. (2020). Methods for evaluating depth perception in a large-screen immersive display. In *Symposium on Spatial User Interaction*. New York, NY, USA: Association for Computing Machinery.

Gonçalves, G., Monteiro, P., Melo, M., Vasconcelos-Raposo, J., & Bessa, M. (2020). A comparative study between wired and wireless virtual reality setups. *IEEE Access*, *8*, 29249–29258.

Grechkin, T. Y., Nguyen, T. D., Plumert, J. M., Cremer, J. F., & Kearney, J. K. (2010). How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Transactions on Applied Perception*, *7*(4).

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin*, *83*(2), 314–320.

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, *8*(3), 33–33.

Hutchison, J. J. & Loomis, J. M. (2006a). Does energy expenditure affect the perception of egocentric distance? A failure to replicate experiment 1 of Proffitt, Stefanucci, Banton, and Epstein (2003). *The Spanish Journal of Psychology*, *9*(2), 332–339.

Hutchison, J. J. & Loomis, J. M. (2006b). Reply to Proffitt, Stefanucci, Banton, and Epstein. *The Spanish Journal of Psychology*, *9*(2), 343–345.

Jo, D., Kim, K., Welch, G. F., Jeon, W., Kim, Y., Kim, K.-H., & Kim, G. J. (2017). The impact of avatar-owner visual similarity on body ownership in immersive virtual reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. VRST '17. Gothenburg, Sweden: Association for Computing Machinery.

Jones, J. A., II, J. E. S., & Bolas, M. (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. A., Swan II, J. E., Singh, G., Kolstad, E., & Ellis, S. R. (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* (pp. 9–14).

Jones, J. A., Swan, J. E., Singh, G., & Ellis, S. R. (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* (pp. 29–36). APGV '11. Toulouse, France: Association for Computing Machinery.

Jones, J. A., Swan, J. E., Singh, G., Franck, J., & Ellis, S. R. (2009). The effects of continued exposure to medium field augmented and virtual reality on the perception of egocentric depth. In *Proceedings of the 6th Symposium on Applied*

Perception in Graphics and Visualization (p. 138). APGV '09. Chania, Crete, Greece: Association for Computing Machinery.

Kelly, J. W., Cherep, L. A., Klesel, B., Siegel, Z. D., & George, S. (2018). Comparison of two methods for improving distance perception in virtual reality. *ACM Transactions on Applied Perception*, *15*(2).

Kelly, J. W., Cherep, L. A., & Siegel, Z. D. (2017). Perceived space in the HTC vive. *ACM Transactions on Applied Perception*, *15*(1).

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, *3*(3), 203–220.

Kramida, G. (2016). Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, *22*(7), 1912–1931.

Kuhl, S. A., Thompson, W. B., & Creem-Regehr, S. H. (2006). Minification influences spatial judgments in virtual environments. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization* (pp. 15–19). APGV '06. Boston, Massachusetts, USA: Association for Computing Machinery.

Kuhl, S. A., Thompson, W. B., & Creem-Regehr, S. H. (2009). HMD calibration and its effects on distance judgments. *ACM Transactions on Applied Perception*, *6*(3).

Kunz, B. R., Wouters, L., Smith, D., Thompson, W. B., & Creem-Regehr, S. (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.

Lambooij, M., Fortuin, M., Heynderickx, I., & IJsselsteijn, W. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology*, *53*(3), 30201-1–30201-14.

Lappin, J. S., Shelton, A. L., & Rieser, J. J. (2006). Environmental context influences visually perceived distance. *Attention, Perception, & Psychophysics*, *68*(4), 571–581.

Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., & Botsch, M. (2017). The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. VRST '17. Gothenburg, Sweden: Association for Computing Machinery.

Leyrer, M., Linkenauger, S. A., Bülthoff, H. H., Kloos, U., & Mohler, B. (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* (pp. 67–74). APGV '11. Toulouse, France: Association for Computing Machinery.

Leyrer, M., Linkenauger, S. A., Bülthoff, H. H., & Mohler, B. J. (2015). Eye height manipulations: A possible solution to reduce underestimation of egocentric distances in head-mounted displays. *ACM Transactions on Applied Perception, 12*(1).

Li, B., Nordman, A., Walker, J., & Kuhl, S. A. (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* (pp. 53–56). SAP '16. Anaheim, California: Association for Computing Machinery.

Li, B., Walker, J., & Kuhl, S. A. (2018). The effects of peripheral vision and light stimulation on distance judgments through HMDs. *ACM Transactions on Applied Perception, 15*(2).

Li, B., Zhang, R., & Kuhl, S. (2014). Minication affects action-based distance judgments in Oculus Rift HMDs. In *Proceedings of the ACM Symposium on Applied Perception* (pp. 91–94). SAP '14. Vancouver, British Columbia, Canada: Association for Computing Machinery.

Li, B., Zhang, R., Nordman, A., & Kuhl, S. A. (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* (pp. 55–58). SAP '15. Tübingen, Germany: Association for Computing Machinery.

Lin, Q., Xie, X., Erdemir, A., Narasimham, G., McNamara, T. P., Rieser, J., & Bodenheimer, B. (2011a). Egocentric distance perception in HMD-based virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization* (p. 123). APGV '11. Toulouse, France: Association for Computing Machinery.

Lin, Q., Xie, X., Erdemir, A., Narasimham, G., McNamara, T. P., Rieser, J., & Bodenheimer, B. (2011b). 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* (pp. 75–82). APGV '11. Toulouse, France: Association for Computing Machinery.

Lugrin, J.-L., Landeck, M., & Latoschik, M. E. (2015a). Avatar embodiment realism and virtual fitness training. In *2015 IEEE Virtual Reality (VR)* (pp. 225–226).

Lugrin, J.-L., Latt, J., & Latoschik, M. E. (2015b). Avatar anthropomorphism and illusion of body ownership in VR. In *2015 IEEE Virtual Reality (VR)* (pp. 229–230).

Mania, K., Adelstein, B. D., Ellis, S. R., & Hill, M. I. (2004). Perceptual sensitivity to head tracking latency in virtual environments with varying degrees of scene complexity. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization* (pp. 39–47). APGV '04. Los Angeles, California, USA: Association for Computing Machinery.

McManus, E. A., Bodenheimer, B., Streuber, S., de la Rosa, S., Bülthoff, H. H., & Mohler, B. J. (2011). The influence of avatar (self and character) animations on distance estimation, object interaction and locomotion in immersive virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization* (pp. 37–44). APGV '11. Toulouse, France: Association for Computing Machinery.

Meehan, M., Razzaque, S., Whitton, M. C., & Brooks, F. P. (2003). Effect of latency on presence in stressful virtual environments. In *Proceedings of the IEEE Virtual Reality 2003* (p. 141). VR '03. Washington, DC, USA: IEEE Computer Society.

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

Mohler, B. J., Bülthoff, H. H., Thompson, W. B., & Creem-Regehr, S. H. (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* (p. 194). APGV '08. Los Angeles, California: Association for Computing Machinery.

Mohler, B. J., Creem-Regehr, S. H., & Thompson, W. B. (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* (pp. 9–14). APGV '06. Boston, Massachusetts, USA: Association for Computing Machinery.

Mohler, B. J., Creem-Regehr, S. H., Thompson, W. B., & Bülthoff, H. H. (2010). The effect of viewing a self-avatar on distance judgments in an HMD-based virtual environment. *Presence: Teleoperators and Virtual Environments*, *19*(3), 230–242.

Nguyen, T. D., Cremer, J. F., Kearney, J. K., & Plumert, J. M. (2011). Effects of scene density and richness on traveled distance estimation in virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization* (pp. 83–86). APGV '11. Toulouse, France: Association for Computing Machinery.

Nguyen, T. D., Ziemer, C. J., Grechkin, T., Chihak, B., Plumert, J. M., Cremer, J. F., & Kearney, J. K. (2008). Effects of scale change on distance perception in virtual environments. *ACM Transactions on Applied Perception*, *8*(4).

Norman, F., Dowell, C. J., Higginbotham, A. J., Fedorka, N. W., & Norman, H. F. (2018). Sex and age modulate the visual perception of distance. *Attention, Perception, & Psychophysics*, *80*(8), 2022–2032.

Philbeck, J. W. & Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(1), 72–85.

Philbeck, J. W., Woods, A. J., Arthur, J., & Todd, J. (2008). Progressive locomotor recalibration during blind walking. *Perception & psychophysics*, *70*(8), 1459–1470.

Phillips, L., Ries, B., Kaeding, M., & Interrante, V. (2010). Avatar self-embodiment enhances distance perception accuracy in non-photorealistic immersive virtual environments. In *2010 IEEE Virtual Reality Conference (VR)* (pp. 115–1148).

Proffitt, D. R., Stefanucci, J., Banton, T., & Epstein, W. (2003). The role of effort in perceiving distance. *Psychological Science*, *14*(2), 106–112.

Proffitt, D. R., Stefanucci, J., Banton, T., & Epstein, W. (2006a). A final reply to Hutchison and Loomis. *Spanish Journal of Psychology*, *9*(2), 346–348.

Proffitt, D. R., Stefanucci, J., Banton, T., & Epstein, W. (2006b). Reply to Hutchison and Loomis. *Spanish Journal of Psychology*, *9*(2), 340–342.

Rashotte, L. S., Jr., M. W., & Whitmeyer, J. M. (2005). Pretesting experimental instructions. *Sociological Methodology*, *35*(1), 151–175.

Razzaque, S., Kohn, Z., & Whitton, M. C. (2001). Redirected Walking. In *Eurographics 2001 - short presentations*. Eurographics Association.

Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013). The perception of egocentric distances in virtual environments - A review. *ACM Computing Surveys*, *46*(2).

Richards, W. (1970). Stereopsis and stereoblindness. *Experimental Brain Research*, *10*(4), 380–388.

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

Ries, B., Interrante, V., Kaeding, M., & Anderson, L. (2008). The effect of self-embodiment on distance perception in immersive virtual environments. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology* (pp. 167–170). VRST '08. Bordeaux, France: Association for Computing Machinery.

Ries, B., Interrante, V., Kaeding, M., & Phillips, L. (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* (pp. 59–66). VRST '09. Kyoto, Japan: Association for Computing Machinery.

Sahm, C. S., Creem-Regehr, S. H., Thompson, W. B., & Willemsen, P. (2005). Throwing versus walking as indicators of distance perception in similar real and virtual environments. *ACM Transactions on Applied Perception*, *2*(1), 35–45.

Schubert, R. S., Hartwig, J., Müller, M., Groh, R., & Pannasch, S. (2016). Are age differences missing in relative and absolute distance perception of stereoscopically presented virtual objects? In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (pp. 307–308). VRST '16. Munich, Germany: Association for Computing Machinery.

Silva, J. A. D. & Santos, R. A. D. (1984). The effects of instructinos on scales for perceived egocentric distance in a large open field. *Bulletin of the Psychonomic Society*, *22*, 189–192.

Slater, M. (2004). How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *13*(4), 484–493.

Slater, M., Khanna, P., & Mortensen, J. (2009). Visual realism enhances realistic response in an immersive virtual environment. *IEEE Computer Graphics and Applications*, *29*(3), 76–84.

Steinicke, F. & Bruder, G. (2014). A self-experimentation report about long-term use of fully-immersive technology. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction* (pp. 66–69). SUI '14. Honolulu, Hawaii, USA: Association for Computing Machinery.

Steinicke, F., Bruder, G., Hinrichs, K., Lappe, M., Ries, B., & Interrante, V. (2009). Transitional environments enhance distance perception in immersive virtual reality systems. In *Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization* (pp. 19–26). APGV '09. Chania, Crete, Greece: Association for Computing Machinery.

Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S. H., Loomis, J. M., & Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments*, *13*(5), 560–571.

Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, *9*(5), 497–503.

Waltemate, T., Gall, D., Roth, D., Botsch, M., & Latoschik, M. E. (2018). The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, *24*(4), 1643–1652.

Willemsen, P., Colton, M. B., Creem-Regehr, S. H., & Thompson, W. B. (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* (pp. 35–38). APGV '04. Los Angeles, California, USA: Association for Computing Machinery.

Willemsen, P., Colton, M. B., Creem-Regehr, S. H., & Thompson, W. B. (2009). The effects of head-mounted display mechanical properties and field of view on distance judgments in virtual environments. *ACM Transactions on Applied Perception*, *6*(2), 1–14.

Witmer, B. G. & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240.

Witt, J. K., Proffitt, D. R., & Epstein, W. (2004). Perceiving distance: A role of effort and intent. *Perception*, *33*(5), 577–590.

Witt, J. K., Stefanucci, J. K., Riener, C. R., & Proffitt, D. R. (2007). Seeing beyond the target: Environmental context affects distance perception. *Perception*, *36*(12), 1752–1768.

Woods, A. J., Philbeck, J. W., & Danoff, J. V. (2009). The various perceptions of distance: An alternative view of how effort affects distance judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1104–1117.

Zhang, R., Nordman, A., Walker, J., & Kuhl, S. A. (2012). Minification affects verbal- and action-based distance judgments differently in head-mounted displays. *ACM Transactions on Applied Perception*, *9*(3).

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

Zimmons, P. & Panter, A. (2003). The influence of rendering quality on presence and task performance in a virtual environment. In *IEEE Virtual Reality, 2003. Proceedings.* (Pp. 293–294).

Appendix A. Checklist

- Experimental Conditions (Section 3.1)
  - ☐ Have you thought about the specific thing you want to test?
- Design Structure (Section 3.2)
  - ☐ Have you considered the design structure - within-subjects, between-subjects, or combo - of you experiments?
- Experimental Location (Section 4.1)
  - Real
    - Have you found a location that is:
      - ☐ Long enough for all desired distances + initial real position + safety meters?
      - ☐ Wide enough?
      - ☐ Supports necessary equipment?
      - ☐ Has few or manageable distractions?
  - Virtual
    - ☐ Have you considered the kind of VE you want and how to make or obtain it?
    - is the VE:
      - ☐ Long enough for all desired distances + initial virtual position + safety meters?
      - ☐ Wide enough?
    - ☐ Have you considered what pictorial depth cues you would like to include?
    - ☐ Have you thought about the nonpictorial depth cues and how to mitigate their side effects?
    - ☐ Have you thought about the environmental context of the chosen VE?

- Pre-Experiment
    - ☐ Have you considered where to complete any pre-experimental procedures?
- Hardware (Section 4.2)
  - HMD
    - ☐ Have you thought about which HMD to use for virtual-world experiments? Note that the manufacturer of popular, name-brand, contemporary HMDs will have completed major calibration tasks.
    - ☐ Have you thought about what calibration items may need to be double checked and how to do it?
    - ☐ Have you considered how to block outside light from entering the gaps between the HMD and the user's face?
  - Tracking System
    - ☐ Have you considered a secondary tracking system that is large enough for the room?
    - ☐ Have you thought about how to calibrate the secondary tracking system?
  - Masking Noise
    - ☐ Have you considered using on-ear or over-ear headphones if headphones are not built into the HMD?
    - ☐ Have you considered which portable music player to use, if applicable?
    - ☐ Have you thought about the masking noise audio file?
  - Communication
    - ☐ Have you thought about how to communicate with a participant during the experiment?
- Avatars (Section 4.3)
  - ☐ Have you considered avatar use?
  - ☐ If you will use an avatar, have you considered how to make or obtain it?

- Targets and Distances (Section 4.4)
  - ☐ Have you considered the targets you will use?
  - ☐ Have you thought about the distances you will use?
  - ☐ Have you considered the practice distances you will use?
- Positioning and Measurement (Section 5.1)
  - ☐ Have you thought about where to start participants in the real world?
  - ☐ Have you thought about where to start participants in the virtual world?
  - ☐ Have you thought about how you will measure the distance walked?
- Participants (Section 5.2)
  - Have you thought about how to test potential participants':
    - ☐ walking ability
    - ☐ hearing
    - ☐ eyesight - both vision and stereoblindness
  - ☐ Have you thought about compensation?
  - ☐ Have you considered what to do with those who fail pre-screening tests and are participating on the basis of class credit?
  - ☐ Have you thought about possible group and individual differences, such as sex, age, alcohol consumption, sleep deprivation, etc., whether or not to allow certain conditions, and how to factor them into the data analysis?
  - ☐ Have you thought about how many participants you will need?
- Safety and Cleanliness Procedures (Section 5.3)
  - ☐ Have you considered how to reduce tripping hazards?
  - ☐ Have you thought about how to manage HMD cables, if applicable?
  - ☐ Have you thought about other ways to keep participants safe?
  - ☐ Have you thought about what to do with participants who become sick or

wish to stop the experiment and with any data from such participants?

- ☐ Have you thought of ways to prevent the spread of disease?
- ☐ Have you considered how to safely clean and maintain any equipment?

- Consent Forms (Section 6.1)
  - ☐ Have you thought about how to organize, present, and store consent forms?
- Instructions (Section 6.2)
  - ☐ Have you thought about what instructions you will use?
  - ☐ Have you thought about how to present the instructions to participants?
  - ☐ Have you considered which terms may be confusing and how to define them?
  - ☐ Have you considered demonstrating the experimental procedures?
- Putting on Wearables (Section 6.4)
  - ☐ Have you thought about what equipment other than the HMD the participant must wear and when to put them on?
  - ☐ Have you considered the blindfold you will use for practice and/or real-world experiments?
- Practice Blind Walking (Section 6.5)
  - ☐ Have you considered where, how, and for how long to allow participants to practice walking without vision?
  - ☐ Have you thought about how to accomplish the events after practicing blind walking, such as entering the lab and conducting any last calibrations?
- Other Helpful Procedures (Section 7.1)
  - ☐ Have you thought about how to move between phases within a trial?
  - ☐ Have you considered the tone and wording of transitional phrases when communicating with participants?
  - ☐ Have you considered hanging a sign to tell passersby to be quiet?

- Real-World Comparison (Section 7.2)

  ☐ If you are doing an analogous real-world experiment, have you considered the differences between conducting real and virtual experiments?

- Return to the Initial Position (Section 7.3)

  ☐ Have you considered a way to guide participants safely back to the initial real position?

- Feedback (Section 8)

  ☐ Have you thought about what unintentional forms of feedback may exist, and do you have a plan to minimize or eliminate them?

- Post-Questionnaires (Section 9.1)

  ☐ Have you thought about what kinds of questions to ask participants after the conclusion of the experiment? We recommend asking at least the following three:

    ∗ Did you notice anything strange during the experiment?

    ∗ Were any of the experimental instructions confusing?

    ∗ What strategies did you use to judge distances?

- Transfer Effects and Debriefing (Section 9.2)

  ☐ Is there any misleading information or anything else participants should know after the conclusion of experiments?

- Reporting Results (Section 9.3)

  ☐ Have you thought of appropriate methods to analyze the data?

  ☐ Have you considered how to explain any data that was thrown out?