Distance Perception with a Video See-Through Head-Mounted Display

Kevin P. Pfeil
kevin.pfeil@knights.ucf.edu
University of Central Florida

Sina Masnadi
sina@knights.ucf.edu
University of Central Florida

Jacob Belga
jbelga1128@knights.ucf.edu
University of Central Florida

Jose-Valentin T. Sera-Josef
Jose.Valentin.serajosef@knights.ucf.edu
University of Central Florida

Joseph J. LaViola Jr.
jjl@cs.ucf.edu
University of Central Florida

ABSTRACT

In recent years, pass-through cameras have resurfaced as inclusions for virtual reality (VR) hardware. With modern cameras that now have increased resolution and frame rate, Video See-Through (VST) Head-Mounted Displays (HMD) can be used to provide an Augmented Reality (AR) experience. However, because users see their surroundings through video capture and HMD lenses, there is question surrounding how people perceive their environment with these devices. We conducted a user study with 26 participants to help understand if distance perception is altered when viewing surroundings with a VST HMD. Although previous work shows that distance estimation in VR with an HTC Vive is comparable to that in the real world, our results show that the inclusion of a ZED Mini pass-through camera causes a significant difference between normal, unrestricted viewing and that through a VST HMD.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; *User studies*; *Empirical studies in HCI*.

KEYWORDS

Video See-Through, Pass-Through Camera, Virtual Reality, Distance Perception, User Study

ACM Reference Format:
Kevin P. Pfeil, Sina Masnadi, Jacob Belga, Jose-Valentin T. Sera-Josef, and Joseph J. LaViola Jr.. 2021. Distance Perception with a Video See-Through Head-Mounted Display. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3411764.3445223

1 INTRODUCTION

In recent years, Video See-Through (VST) Augmented Reality (AR) devices have re-surged as a commonplace technology. To work, cameras are affixed to the front of a closed-view head-mounted display (HMD), and the live video feed is rendered in the display. This is unlike Optical See-Through (OST) displays, which allow the user to view the real world through a transparent panel (such as the Hololens[1]) [1, 22]. Common VST devices today include the Samsung Gear VR[2], which uses a smartphone's screen and embedded cameras to display the real world, and the ZED Mini[3], which is specifically designed to provide high-resolution, stereo views of the real world while a user wears an HMD. Although VST devices may be limited by the presence of a screen with lower resolution than that of natural human eyesight, and in some instances, the presence of depth quantization [16], Kruijff et al. note that VST devices do not suffer from the brightness and contrast hindrances that OSTs have, meaning that graphical overlays can be strongly seen in outdoor environments [18]. Other strengths of VST HMDs include ease of superimposing graphics into the view of the real world, a Field of View (FOV) based on the imaging device, and that these devices are capable of providing mixed reality experiences across the entire reality-virtuality continuum (see Milgram's definition [28]).

These HMDs naturally have the ability to render virtual environments (VEs). Using computer vision techniques, graphical overlays representing the real world can be added to the VE. Likewise, VST HMDs can naturally display the real world, and graphical overlays can be used to display AR cues [22]. Though originally conceived decades ago [1], we expect that these devices will increase in popularity, due to a variety of advantages. First, they provide users with the capability to switch between VEs and the real world without needing to take the headset off. Second, they can be used for both Virtual Reality (VR) and AR, instead of requiring a dedicated device for each modality. Third, they allow for unique AR experiences that OST displays can not provide. For instance, Jones et al. use VST HMDs to display the real world, but apply a graphics transformation to show users what it is like to have visual impairments [13]; and likewise, Masnadi et al. perform inverse transformations to *correct* eye-sight for visually impaired users [26]. In theory, these devices could be worn at all times, to correct more severe visual impairments during a user's daily life.

However, there is a large body of work that shows how users tend to perceive VEs differently from the real world. An HMD's screen resolution, FOV, and weight are all potential factors that degrade a user's perception of a VE [14, 43]. Although the displayed graphics are captures of the real world, and although modern technology is

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445223

---

[1] https://www.microsoft.com/en-us/hololens

[2] https://www.samsung.com/global/galaxy/gear-vr/

[3] https://www.stereolabs.com/zed-mini/

allowing human perception of VEs to become comparable to the real world [14], we question if users can properly perceive their surroundings while wearing a VST HMD. This would pose a problem for a variety of applications, including training, spatial awareness acquisition, and gaming. Though there is a corpus of literature that discusses distance underestimation in VR, there is little work that focuses on distance estimation using VST HMDs. Therefore, based on previous work, we conduct a 3x4 within-subjects study in order to measure user perception of a static environment with multiple viewing conditions, asking the following question:

RQ1: How do users perceive distance in the real world when wearing a Video See-Through Head-Mounted Display?

The results of our study show that people tend to underestimate distances when using our selected device combination, the HTC Vive with ZED Mini attachment. Though recent work suggests that viewing a VE with the Vive is comparable to real world viewing [14], our results show that using it as a VST HMD significantly compresses a user's perception of distance. Our results suggest that this is in part due to the reduced FOV that is provided by the device. We offer these findings to the SIGCHI community so to direct work towards better imaging devices, and to help explain potential perceptual disparities in current applications using VST HMDs.

2 RELATED WORK

Our work is related to a body of literature that tackles the problem of VR distance underestimation. In that corpus, authors typically compare how users perceive distances in real life to VR. In our study, our participants wear a VR head-mounted display (HMD) with a pass-through camera; they see the real world in stereo, but the quality of view - including FOV and resolution - is reduced due to the hardware. Therefore, in this section, we provide an overview of the relevant literature at the intersection of Distance Underestimation and VST HMDs.

2.1 Distance Underestimation with Head-Mounted Displays

For over two decades, researchers have studied how users perceive distances in VR, and a common result is that users tend to significantly underestimate distances in VR compared to the real world [34]. There are a variety of factors that have been identified as contributors to a user's perception of their virtual environment. For instance, Mohler et al. investigated the inclusion of self-avatars (a virtual representation of the user), and found that these provided a significant frame of reference, resulting in more accurate distance judgements [30]. Likewise, Leyrer et al. show that camera height manipulations in a VE also affects a user's perception; for instance, by increasing the camera height in relation to the user's actual height, the authors found that users tend to underestimate distances [23]. Additionally, Kunz et al. show that the quality of computer graphics inside a VE affects judgements, such that lower-quality textures were more conducive to poorer distance estimations [19]. While these articles are pivotal to the VR distance underestimation literature, we note that these factors (presence of self-avatars, eye height, quality of graphics) are constants in the real world with VST displays. Therefore, these factors are not explored in our study. In our work, users can see their own body; the camera is mounted at their own natural eye height; and the graphics are live captures of the real world.

In the extant literature, researchers have utilized multiple procedures in order to measure distance estimations, including verbal judgments, blind walking, timed imagined walking, and blind throwing [34]. Verbal judgments have been used to draw out a user's estimation of depth with a simple procedure, but accuracy tends to decline as targets are located further away [19, 25, 34]. Perhaps the most popular method, blind walking consists of a participant viewing a target, becoming blindfolded, and then walking until they believe they reached the target. This procedure has been used in both real and virtual environments, and historically, participants viewing a VE have walked significantly shorter distances than those viewing the real world [7, 29, 38, 43]. One of the weaknesses of this technique, as reported by Jones et al., is that participants might be able to peek at the ground through the gap between the face and the HMD to use optical flow as feedback [10]. However, a more recent work by Jones et al. suggests that modern VR HMDs alleviate this problem, as wider FOVs fill the periphery of the viewer [12]. Similar to blind walking, timed imagined walking consists of a user viewing a target and becoming blindfolded, but instead of walking to the target, they imagine walking; then they tell the researcher when they believe they would have arrived, if they walked normally. Previous results have shown that this method elicits responses comparable to blind walking [8, 31]. Lastly, blind throwing measures a user's depth estimation while allowing them to remain stationary [33, 36]. Sahm et al. used blind throwing and found that responses were comparable to blind walking; they conclude that this procedure is a suitable measurement for when blind walking is not usable [36]. During a time when COVID-19 forces us to be diligent with safety precautions, we elect to use a blind throwing procedure for our study, in order to minimize the risk of spreading the virus; this allows us to remain socially distant from our participants, and eliminates the risk of participants tripping over HMD cables while their eyes are closed.

As noted in a recent literature survey by El Jamiy and Marsh, there is not as much work regarding AR distance estimation compared to that in VR [3], but there are some findings that have emerged in this area. Regarding device type, some researchers have studied the effects of wearing an OST AR display. For instance, Grechkin et al. [8], Jones et al. [10, 11], and Livingston et al. [24] used the NVIS nVisor ST device, which could be used as a VR or AR display. As the participants could see the real world normally through the lenses, the authors were able to directly study if the limited FOV or additional weight of a device caused a difference in depth estimation. Grechkin et al. found no statistical difference between wearing the HMD and normal viewing [8]; however, Jones did find a significant difference in depth estimation, such that users wearing the HMD underestimated distances [11]. This same result was confirmed in another experiment by Jones et al.; the authors found similar findings when using a within-subjects and a between-subjects study design [10]. Livingston et al. compared OST AR distance estimation between indoor and outdoor environments, in the presence of AR cues, and found that participants underestimated distance in an indoor hallway environment, but overestimated distances in an outdoor parking lot environment [24]. Swan et al. used a different OST AR device - the Sony Glasstron - and found that users underestimated distance with this device as well [37]. Although there is conflicting evidence, it seems that FOV may be a significant factor that affects a user's ability to judge distances. The diagonal FOV of our device (the ZED Mini) is 100°. Thus, we included three viewing conditions in our study, to help understand the effects of FOV with and without an HMD.

Fewer researchers still have studied how VST displays affect distance judgments of a physical environment. Vaziri et al. used a custom VST device and implemented software techniques to transform the view into a live, non-photo-realistic representation of the environment [40]. Here, the authors were able to study the effects of varying graphical representations of the real world, and found that the non-photo-realistic conditions were conducive to significantly different responses compared to viewing the real world through a VST; however, they did not make a direct comparison to real-world viewing without a VST device. In a separate study, Kyto et al. studied depth estimation while the user saw AR cues [20]. Instead of measuring how well a user perceived the physical environment through a VST HMD, they used an action-based measurement to understand perception of AR cue depth, with monoscopic and stereoscopic devices. Naturally, the authors found that stereoscopic rendering of AR cues was more conducive to understanding distance. In a recent study, Gagnon et al. used the HTC Vive with ZED Mini attachment in order to study user estimations of lengthier distances in a virtual environment [6]. Using a verbal reporting procedure, they found a trend where users moderately overestimated shorter distances (25m-200m), but then significantly underestimated larger distances (300m-500m). Although they used the same VST that we do, their setup involved the use of a green screen, to completely turn a physical environment into a virtual one. Our work differs in that the participants were able to view the real world instead of a virtual world, and we restricted our stimulus distances to 6m using a blind throwing task. While previous works utilize VST HMDs in various ways to measure distance estimation, to the best of our knowledge, ours is the first that directly compares perception of the real world when using and not using a VST HMD.

2.2 Video See-Through Head-Mounted Displays

VST HMDs utilize forward-facing cameras affixed to the front of the device, allowing users to "see through" the hardware into the real world. Unlike OST AR displays that allow a user to see the real world normally, a VST HMD captures the real world and displays it on a screen that is typically used to show VEs. This subjects the user to some degree of latency, and it exposes the user to the HMD's screen resolution and FOV [35]. Although VSTs were conceptualized decades ago [1], we are seeing increased usage with modern technology. For instance, Kumaravel et al. utilized a suite of devices and visualization techniques to connect 2 remote users - one expert and one novice - and merge their environments [39]. Both users were able to see the first-person views of their own surroundings and that of their partner, by merging camera feeds. Similarly, Cao et al. used VST HMDs with AR to let a user see their surroundings while viewing an expert's movements, in a task akin to manufacturing / maintenance [2]. This allowed them to perform the task while simultaneously watching the expert's instructions. Walker et al. implemented a VST HMD to provide a remote collaborative AR experience to users [41]. Here, users saw the real world, with a graphical overlay of a virtual avatar that represented a live human partner. The authors studied the effects of using different sized avatars which were controlled by these remote users, and found that displaying to-scale avatars was more conducive towards an equitable interaction between both users.

Jones et al. discuss how to use computational approaches to simulate visual impairments [13]. They used a VR headset with eye tracking in order perform an appropriate transformation on the graphics, and then displayed the result in the HMD. By using a VST device, the users could see the real world with the distortion. Likewise, Masnadi et al. discuss a transformation that allows users with severe visual impairments, such as age-related macular degeneration, to see with better acuity [26]. It is implied that by using a VST HMD equipped with eye tracking, a visually impaired person could see the world normally during their daily lives. In a more general use case, Rabbi et al. deployed VST HMDs to users engaging in a weight training circuit [32]. Here, virtual feedback was shown to the users as they lifted the weights, to correct their form. Then, the "see-through" mode was engaged to allow them to navigate to the next station without removing the HMD. Although this example doesn't utilize the VST mode for intense usage, it still highlights the practicality of such a device in daily life, allowing the user to switch modalities without removing the HMD. We suspect that VST devices will see a surge of personal use, even if just for allowing users to see their physical space in-between VR sessions. As such, it is important to understand how people perceive their surroundings while using these devices.

3 METHODS

We conducted a user study to help understand distance perception in real world environments when wearing a VST HMD compared to normal viewing. The following sections describe our study details.

3.1 Study Design

We conducted a 3x4 within-subjects study with two independent variables - *Headgear* and *Distance*. The *Headgear* variable had three levels - "Nothing", where the users did not wear any headgear; "VST", where the users wore a VST HMD; and "Shell", where the users wore just a plastic casing from a stripped-down HMD, effectively emulating the reduced field of view in the VST HMD condition. The *Distance* variable had four levels - 3m, 4m, 5m, and 6m. This resulted in 12 unique conditions, and the participants performed a blind throwing task 3 times each, for a total of 36 trials per participant; these trials were randomized for each individual. The task and Distance conditions were influenced by previous work by Sahm et al. [36]. While typical distance perception studies use blind-walking as the primary measure [34], we elect to use blind throwing, as it has been shown to elicit responses comparable to blind-walking [36], and it allows the researchers and participants to remain socially distant during times of COVID-19.

Figure 1: Left: Video See-Through HMD, the HTC Vive with ZED Mini attachment. Right: Shell of an HMD.

3.2 Subjects

We used the G*Power software package to perform a power analysis [4]. With a medium effect size and 12 conditions in a within-subjects study, our target N was 18 subjects. We recruited 28 participants from the University of Central Florida to participate in our study, but 2 were dismissed for not passing an eye test. Our final participant pool consisted of 26 individuals (24 male, 2 female). Their ages ranged from 18 to 29 (M = 19.46, SD = 2.45). We screened all participants for vision acuity using a Snellen eye chart, and all could see better than 20/32 in each eye. If a participant wore corrective lenses, they were required to wear them for all conditions during the study. We asked participants to rate their experience with VR, on a scale of 1 to 5 where 1 means little and 5 means much. The mean response was 2.30 (SD = 1.03).

3.3 Apparatus

Our conditions called for participants to wear either no headgear, a VST HMD, or a plastic shell of a headset. For the VST conditions, participants wore an HTC Vive[4] that was equipped with a ZED Mini[5] pass-through camera mounted on the front. This pass-through camera operated at 60FPS, with a resolution of 2560x720, providing a stereo video feed which resulted in a 3D view. Although the HTC Vive display has an approximate field of view (FOV) of 110°, inclusion of the pass-through camera reduced the FOV to 90° vertically, and 60° horizontally. The rest of the display was filled with black pixels. The weight of the VST HMD was 550g. We used an out-of-the-box Unity3D program which was provided by the creators of ZED Mini – this program can be found on their website[6]. We only modified the scene such that pressing the space bar would turn the camera on or off; thus, we did not develop any additional software, and only used a commercial-off-the-shelf solution. For the Shell conditions, participants wore a stripped-down Oculus Rift SDK2. Here, there was no display; all that remained were the plastic casing and the adjustable straps. Users could see the real world, though their FOV was limited by the plastic casing. We added black cardboard to the top and bottom of the plastic, to further restrict the FOV so that it more accurately matched that of the VST condition. See Figure 1 for illustration of the VST display and the shell device. The weight of the Shell device was 150g. For all trials, participants threw beanbags that weighed approximately 450g and were square, measuring 15cm on each side. The target that users were asked to throw the beanbag towards was a circle approximately 16cm in diameter.

The study was conducted in our closed laboratory. While typical distance estimation studies are conducted in an empty hallway, or lab was filled with desks, chairs, television monitors, and miscellany, resulting in a visually richer environment that is perhaps more representative of real use cases. Our space was large enough to accommodate the furthest distance of our study, 6m. There was approximately 1.8m of buffer from the 6m target to the closest non-study object, and there was at least 1.3m on either side of the targets. The ceiling was approximately 3m high. We did not move any of the objects in our lab until all participants completed data collection. Thus, any objects in the environment which could have been used as reference frames were constant across all participants. See Figure 2 for illustration of our environment.

---

[4] https://www.vive.com

[5] https://www.stereolabs.com/zed-mini/

[6] https://www.stereolabs.com/docs/unity

Figure 2: Environment used for our study. The target distances were 3m, 4m, 5m, and 6m.

3.4 Procedure

Upon arrival, recruited participants were asked to review an informed consent document, but we did not collect signatures. We introduced users to the different headset configurations and conditions that we would be testing throughout the session, and then had each participant perform an eye examination with a Snellen chart. We recorded their vision acuity. Next, we collected the remainder of user demographics. We then gave the participants an overview of our study. We explained that the objective was to toss beanbags at targets with varying distances, while wearing one (or none) of our headgear. Users were asked to first view the target, and then after a time that they were comfortable with, close their eyes and attempt to hit the target as close as they could with the beanbag. We noted to the participants that we would only count the initial contact with the ground, and not bounces or slides of the beanbag. Prior to starting the trials, we had the user practice throwing 4 bean bags at 4 different targets without closing their eyes (16 practice throws in total). These targets used the same distances as the actual trials, but were not the same as the ones used during data collection.

We then proceeded to collect data from our participants. During the trials, we instructed the user to which headgear (if any) they would need to put on. We encouraged participants to tell us if the VST conditions were too blurry, and showed them how to adjust the interpupillary distance when necessary. Before each toss, we instructed the user to close their eyes and to keep them closed after they tossed. The researcher ensured that the participants' eyes were closed during this time; for the VST conditions, we blacked out the screen, to verify that the user could not see. Then the user threw the beanbag, and a researcher used a tape measure to log the distance from the target to the spot where the beanbag first made contact with the ground. After each measurement, we removed the bean bag from the ground before telling the participant that they could open their eyes. This process repeated for a total of 36 times. After data collection was complete, participants were given 5USD in cash. The time to complete the study was approximately 30 minutes.

3.5 COVID-19 Considerations

Due to the ongoing COVID-19 pandemic, we wanted to ensure safety for the participants and researchers. Following our institutions guidelines, all individuals were required to wear face masks at all times. Between each participant, we sanitized all devices and surfaces that the participants and researchers would be in contact with, to ensure safety during the study. Furthermore, all users were required to wear a face mask in order to participate in the study, but we provided each individual face masks, hand sanitizer, cleaning wipes, latex gloves, and single-use VR Eye mask coverings, to reduce risk of contracting the disease. Though we cleaned all surfaces between participants, we allowed each individual to clean devices as desired.

3.6 Hypotheses

While findings have historically shown how people tend to significantly underestimate distances when wearing a VR HMD [34], recent work has shown that the gap of perception between the real world and virtual representations is shrinking [14]. However, when using a VST HMD to view the real world, there is an additional layer of graphics that might affect how users perceive distance. Lastly, we note through previous work that a reduced FOV alone is not a cause for distance underestimation [34]. Therefore, we hypothesize the following for our study:

H1: Participants will under-throw the beanbag in the VST condition compared to the Nothing and Shell conditions.
H2: Participants will make comparable errors when throwing the beanbag in the Nothing and Shell conditions.
H3: Participants will make larger errors when throwing the beanbag to further distances.

3.7 Data Analysis Approach

The participants tossed the bean-bag 3 times for each headgear-distance pair, and the error of their toss in relation to the target (magnitude and direction) was recorded. If the bag was tossed to the side, we transposed the landing spot such that it aligned with the line from the participant to the target. Positive values indicate overthrowing the bag, and negative values indicate under-throwing. For each condition, we averaged the error of 3 tosses together to form one data point. Thus, though each participant threw the bag 36 times, each individual had 12 associated data points. We performed a Shapiro-Wilks test for normality on the data, and found that the responses were not normally distributed ($p < .01$). We therefore transformed our data using ARTool [44], so that we could run a repeated-measures ANOVA on our data. In the event of a significant omnibus test, we planned on conducting pair-wise t-tests on the main effects. Since we had multiple comparisons, we controlled for Type I errors using Holm's Sequential Bonferroni Adjustment [9].

4 RESULTS

In this section, we report the findings of our statistical analyses. First we describe the descriptive statistics, followed by omnibus and post-hoc tests.

4.1 Descriptive Statistics

Table 1 details the descriptive statistics for our data. Expectedly, our participants made greater errors with further targets; but, generally, participants threw the beanbags more accurately when not wearing any headgear.

4.2 Repeated Measures ANOVA Results

After transforming our data using the ARTool [44], we conducted a 3x4 repeated-measures ANOVA. We found a significant main effect of Headgear, and a significant main effect of Distance; but we did not find an interaction effect between these two variables. Table 2 depicts the results of our omnibus test. Having found statistical significance, we proceeded to conduct post-hoc tests.

Table 1: Descriptive statistics of average user error, by Headgear and Target Distance

Headgear	Target Distance	Error (centimeters)
Nothing	3 meters	M = -5.24, SD = 27.3
	4 meters	M = -17.1, SD = 31.8
	5 meters	M = -23.0, SD = 29.9
	6 meters	M = -44.8, SD = 50.5
Shell	3 meters	M = -12.0, SD = 27.1
	4 meters	M = -25.5, SD = 34.2
	5 meters	M = -36.1, SD = 35.8
	6 meters	M = -58.5, SD = 43.0
VST	3 meters	M = -14.0, SD = 29.2
	4 meters	M = -19.4, SD = 43.4
	5 meters	M = -40.3, SD = 37.8
	6 meters	M = -66.7, SD = 53.3

Table 2: Repeated Measures ANOVA results

Effect on Error	ANOVA Result
Main Effects	
Headgear	$F(2, 50) = 5.902, p < .01, \eta_p^2 = .191$
Distance	$F(3, 75) = 25.62, p < .001, \eta_p^2 = .506$
Interaction Effect	
Headgear * Distance	$F(6, 150) = 1.563, p = .162, \eta_p^2 = .059$

4.3 Post-hoc Test Results

Using the transformed data, we conducted pair-wise t-tests on the Headgear conditions; see Table 3 for statistical test results, and Figure 3 for illustration. Testing for the effect of Headgear, we found significant differences when comparing Nothing with Shell, and Nothing with VST. However, we did not find a significant difference between the Shell and VST conditions. Generally, participants performed the task with less error when viewing the environment normally. Regardless of device, viewing the room with a reduced field of view induced more error.

We also performed pair-wise t-tests on the Distance conditions; see Table 3. We found statistical significance when comparing each of the target distances. Generally, as the targets increased in distance, the participants performed the task with greater error.

5 DISCUSSION

The results of our study provide insight into distance perception when using a VST HMD. The following section details the implications of our findings.

5.1 Distances are Underestimated with VST Devices

The results of our study showed a significant difference in participants' ability to estimate distances when they used the VST

Table 3: Post-hoc t-tests results

Condition A	M	SD	Condition B	M	SD	t-test Result
Effect of Headgear						
Nothing	-8.86	14.78	VST	-13.65	18.13	$t(103) = 2.667, p < .01^*$
Nothing	-8.86	14.78	Shell	-12.96	15.45	$t(103) = 2.626, p < .05^*$
Shell	-12.96	15.45	VST	-13.65	18.13	$t(103) = 0.261, p = .80$
Effect of Distance						
3m	-4.27	10.26	4m	-7.88	14.42	$t(77) = 2.920, p < .01^*$
3m	-4.27	10.26	5m	-12.82	13.81	$t(77) = 5.384, p < .001^*$
3m	-4.27	10.26	6m	-22.35	19.46	$t(77) = 9.271, p < .001^*$
4m	-7.88	14.42	5m	-12.82	13.81	$t(77) = 2.581, p < .05^*$
4m	-7.88	14.42	6m	-22.35	19.46	$t(77) = 6.142, p < .001^*$
5m	-12.82	13.81	6m	-22.35	19.46	$t(77) = 4.054, p < .001^*$

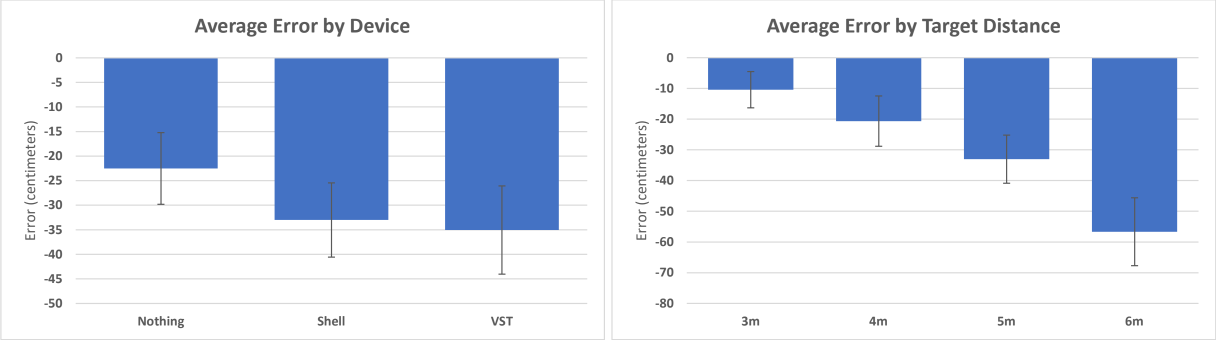

Figure 3: Left: Average error by device with 95% confidence. Right: Average error by Distance with 95% confidence. Generally, participants under-threw the beanbag to the targets, but were less accurate when wearing a device on their head; further, they were less accurate as the targets distance increased.

device, compared to real-world viewing (H1). Though prior work has shown that viewing a VE with an HTC Vive is comparable to that of normal viewing of the real world [14], we found that the inclusion of the ZED Mini pass-through camera worsened distance estimation. There are a few factors which changed as a result of adding the ZED Mini. Device weight (and therefore weight distribution) immediately changed, as the ZED Mini is 63g, accounting for 11% of the total device weight which was attached to the frontmost part of the headset. Only when using the HMD as a VST device does the FOV reduce. The normal FOV of the Vive is approximately 110° (both horizontal and vertical), but the ZED Mini reduced this to 90° horizontally and 60° vertically. The ZED Mini's pass-through mode operates at a high resolution (2560x720), so we do not suspect resolution to be a significant factor for the present study.

Though previous works suggested that FOV is not by itself a main factor that affects distance estimation [34, 43], our work suggests that this very well may be the case (H2); however, our present study is unable to isolate the effects of weight on distance compression. Our participants exhibited statistically comparable responses between the Shell and VST conditions, although we note that the Shell elicited slightly more accurate responses. This slight error reduction might be attributed to the fact that the Shell weighed approximately 25% of the VST device's weight. Previously, Willemsen et al. concluded that reduced FOV *combined* with the weight of an HMD caused this difference; they used a Shell of the nVisor SX device which weighed approximately 1kg[7] [43], but our Shell only weighed a fraction of that (150g) and its mass is closer to the head, which produces near-negligible torque on the head. This reasonably suggests that FOV is a main factor that affects distance estimation, even with modern hardware. However, isolated investigation is required to evaluate the effect of HMD weight, since 150g weight of the headset might produce enough torque to affect head movements and subsequently affect the distance judgement.

5.2 Implications for VST HMD Design

Although recent research has shown that distance perception between VR and the real world is comparable with modern VR HMDs such as the HTC Vive [14], our study demonstrates that when adding a VST device to the setup, the reduction of FOV, additional weight of the imaging camera, or a combination of both, significantly hinders user ability to perceive depth in an action-based

---

[7] https://est-kl.com/manufacturer/nvis/nvisor-sx.html

context. Current VST devices have reduced imaging FOVs compared to their displays (common devices such as the HTC Vive and Oculus Rift boast approximately 110° for the display) - looking to VST exemplars, the ZED Mini has 90° by 60°; Varjo XR-1 has 87°; and the HTC Vive SRWorks has 96° by 80° (at a reduced resolution [21]). Therefore, our study implies that VST HMD designers should work towards improved imaging devices which will offer a FOV closer to 110°, and work towards miniaturizing the equipment such that weight can be reduced. Since FOV reductions cause users to rotate their heads more during visual acquisition, and since additional weight of an HMD can cause damage to the neck [17, 42], it is important to consider expanding FOVs.

However, we must be aware that an increased FOV might have a significant trade-off concerning simulator sickness. While our paper did not assess this (e.g. through use of the Simulator Sickness Questionnaire [15]), previous research indicates that a reduced FOV helps to prevent users from experiencing discomfort when navigating an environment [5]. Since some envisioned use cases involve prolonged wearing of a VST HMD (e.g. for correcting visual impairments during everyday life [26], we ponder, then, if increasing the FOV would actually cause users to stop wearing the device. More work is necessary in order to isolate this problem.

5.3 Limitations and Future Work

We acknowledge the limitations of our study. Our work is grounded in the distance compression literature, in which the primary measurement is blind walking. Due to the COVID-19 pandemic, we chose to utilize the blind throwing method, to allow both participant and experimenter to remain socially distant from each other, and to eliminate risk of tripping over HMD cables. Here, although blind throwing has been shown to be comparable to blind walking [36], we are unable to make direct comparisons with previous literature. Although blind throwing has been shown to be an appropriate alternative [36], this technique is not well documented. Other, recent articles that employ throwing do so while the participants are not blindfolded, and do not compare against walking [27, 45]. When the threat of COVID-19 is reduced, we plan on conducting a similar experiment with the blind walking protocol; but in general, future work should consider investigations into alternatives to blind walking, as walking might not always be possible for participants. In addition, we also note that both blind throwing and blind walking procedures ultimately restrict the distances that we can use in a study, due to physiological limitations and safety concerns. As such, the findings of our present study cannot be generalized to great distances such as those used by Gagnon et al. [6], and verbal estimations should be employed in parallel to other data collection methods.

We also note that our sample is male dominated, and we were unable to test for gender differences; typically our lab recruits at a 2:1 M:F ratio, but we were unable to remotely approach that mark; we suspect that COVID-19 caused this disparity in recruitment. Thus, we cannot yet generalize our results for all genders. Further, though our results show that distance is compressed with VST devices, we acknowledge that we were unable to explore the effects of one of the limitations of VST HMDs - latency. Our participants were exposed to the latency of the camera, but ultimately, due to the blind nature of the experiment, we do not expect it to have affected our results. We do anticipate it to affect more action-based tasks that involve visual search and hand-eye coordination, such as catching a ball; we plan on conducting more in-situ experiments to understand how users perceive their surroundings when using VST HMDs during more intensive scenarios. Lastly, what is puzzling is that our device's FOV (90° x 60°) is wider than that of previous work (48° x 40°) [8, 11, 43], yet our Shell device elicited a significantly different response than unrestricted viewing. We are motivated to reproduce this study using an even wider FOV, generated by a device such as the Pimax HMD[8] and pass-through cameras that would fill up more of the screen.

6 CONCLUSION

In this paper, we demonstrated with an action-based protocol that user perception of distances in a real-world environment is compressed with a reduced FOV and increased device weight. Modern VST HMDs currently exhibit reduced FOVs, and we thus question if these devices can be used continuously in daily life, in their current state. We anticipate future devices will provide wider FOVs, but we plan on pursuing a line of work that studies how people can safely use these devices until the next wave of VST HMDs are developed.

ACKNOWLEDGMENTS

This work is supported in part by NSF Award IIS-1917728.

REFERENCES

[1] Ronald T. Azuma. 1997. A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments* 6, 4 (1997), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

[2] Yuanzhi Cao, Xun Qian, Tianyi Wang, Rachel Lee, Ke Huo, and Karthik Ramani. 2020. An Exploratory Study of Augmented Reality Presence for Tutoring Machine Tasks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376688

[3] F. El Jamiy and R. Marsh. 2019. Survey on depth perception in head mounted displays: distance estimation in virtual reality, augmented reality, and mixed reality. *IET Image Processing* 13, 5 (2019), 707–712. https://doi.org/10.1049/iet-ipr.2018.5920

[4] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods* 41, 4 (2009), 1149–1160.

[5] A. S. Fernandes and S. K. Feiner. 2016. Combating VR sickness through subtle dynamic field-of-view modification. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, Greenville, SC, 201–210. https://doi.org/10.1109/3DUI.2016.7460053

[6] Holly C. Gagnon, Lauren Buck, Taylor N. Smith, Gayathri Narasimham, Jeanine Stefanucci, Sarah H. Creem-Regehr, and Bobby Bodenheimer. 2020. Far Distance Estimation in Mixed Reality. In *ACM Symposium on Applied Perception 2020* (Virtual Event, USA) *(SAP '20)*. Association for Computing Machinery, New York, NY, USA, Article 9, 8 pages. https://doi.org/10.1145/3385955.3407933

[7] Michael Geuss, Jeanine Stefanucci, Sarah Creem-Regehr, and William B. Thompson. 2010. Can I Pass? Using Affordances to Measure Perceived Size in Virtual Environments. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization* (Los Angeles, California) *(APGV '10)*. Association for Computing Machinery, New York, NY, USA, 61–64. https://doi.org/10.1145/1836248.1836259

[8] Timofey Y. Grechkin, Tien Dat Nguyen, Jodie M. Plumert, James F. Cremer, and Joseph K. Kearney. 2010. How Does Presentation Method and Measurement Protocol Affect Distance Estimation in Real and Virtual Environments? *ACM Trans. Appl. Percept.* 7, 4, Article 26 (July 2010), 18 pages. https://doi.org/10.1145/1823738.1823744

[9] Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* 6, 2 (1979), 65–70.

[10] J. Adam Jones, J. Edward Swan, Gurjot Singh, and Stephen R. Ellis. 2011. Peripheral Visual Information and Its Effect on Distance Judgments in Virtual

---

[8] https://www.pimax.com

and Augmented Environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization* (Toulouse, France) *(APGV '11)*. Association for Computing Machinery, New York, NY, USA, 29–36. https://doi.org/10.1145/2077451.2077457

[11] J. Adam Jones, J. Edward Swan, Gurjot Singh, Eric Kolstad, and Stephen R. Ellis. 2008. The Effects of Virtual Reality, Augmented Reality, and Motion Parallax on Egocentric Depth Perception. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization* (Los Angeles, California) *(APGV '08)*. Association for Computing Machinery, New York, NY, USA, 9–14. https://doi.org/10.1145/1394281.1394283

[12] J. A. Jones, J. E. Swan II, and M. Bolas. 2013. Peripheral Stimulation and its Effect on Perceived Spatial Scale in Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics* 19, 4 (2013), 701–710. https://doi.org/10.1109/TVCG.2013.37

[13] P. R. Jones and G. Ometto. 2018. Degraded Reality: Using VR/AR to simulate visual impairments. In *2018 IEEE Workshop on Augmented and Virtual Realities for Good (VAR4Good)*. IEEE, Reutlingen, 1–4.

[14] Jonathan W. Kelly, Lucia A. Cherep, and Zachary D. Siegel. 2017. Perceived Space in the HTC Vive. *ACM Trans. Appl. Percept.* 15, 1, Article 2 (July 2017), 16 pages. https://doi.org/10.1145/3106155

[15] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3 (1993), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

[16] Robert V. Kenyon and Stephen R. Ellis. 2014. *Vision, Perception, and Object Manipulation in Virtual Environments*. Springer New York, New York, NY, 47–70. https://doi.org/10.1007/978-1-4939-0968-1_4

[17] James F. Knight and Chris Baber. 2004. Neck Muscle Activity and Perceived Pain and Discomfort Due to Variations of Head Load and Posture. *Aviation, Space, and Environmental Medicine* 75, 2 (2004), 123–131. https://www.ingentaconnect.com/content/asma/asem/2004/00000075/00000002/art00005

[18] E. Kruijff, J. E. Swan, and S. Feiner. 2010. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*. IEEE, Seoul, 3–12. https://doi.org/10.1109/ISMAR.2010.5643530

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

[20] Mikko Kytö, Aleksi Mäkinen, Timo Tossavainen, and Pirkko T Oittinen. 2014. Stereoscopic depth perception in video see-through augmented reality within action space. *Journal of Electronic Imaging* 23, 1 (2014), 011006.

[21] Stereo Labs. 2018. *Vive Pro AR (SRWorks) vs. ZED Mini*. Stereo Labs. Retrieved 2020-12-22 from https://www.stereolabs.com/blog/vive-pro-ar-zed-mini/

[22] Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional, Boston, MA, USA.

[23] Markus Leyrer, Sally A. Linkenauger, Heinrich H. Bülthoff, Uwe Kloos, and Betty Mohler. 2011. The Influence of Eye Height and Avatars on Egocentric Distance Estimates in Immersive Virtual Environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization* (Toulouse, France) *(APGV '11)*. Association for Computing Machinery, New York, NY, USA, 67–74. https://doi.org/10.1145/2077451.2077464

[24] M. A. Livingston, Z. Ai, J. E. Swan, and H. S. Smallman. 2009. Indoor vs. Outdoor Depth Perception for Mobile Augmented Reality. In *2009 IEEE Virtual Reality Conference*. IEEE, Lafayette, LA, 55–62. https://doi.org/10.1109/VR.2009.4810999

[25] Jack M Loomis and John W Philbeck. 2008. Measuring spatial perception with spatial updating and action. In *Carnegie Symposium on Cognition, 2006, Pittsburgh, PA, US*. Psychology Press, Psychology Press, Pittsburgh, PA, US, 1–43.

[26] S. Masnadi, B. Williamson, A. N. V. González, and J. J. LaViola. 2020. VRiAssist: An Eye-Tracked Virtual Reality Low Vision Assistance Tool. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, Atlanta, GA, USA, 808–809.

[27] Helen C Miles, Serban R Pop, Simon J Watt, Gavin P Lawrence, Nigel W John, Vincent Perrot, Pierre Mallet, Daniel R Mestre, and Kenton Morgan. 2014. Efficacy of a virtual environment for training ball passing skills in rugby. In *Transactions on Computational Science XXIII*. Springer, Berlin, Heidelberg, 98–117.

[28] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: a class of displays on the reality-virtuality continuum. In *Telemanipulator and Telepresence Technologies*, Hari Das (Ed.), Vol. 2351. International Society for Optics and Photonics, SPIE, Boston, MA, 282 – 292. https://doi.org/10.1117/12.197321

[29] Betty J. Mohler, Sarah H. Creem-Regehr, and William B. Thompson. 2006. The Influence of Feedback on Egocentric Distance Judgments in Real and Virtual Environments. In *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization* (Boston, Massachusetts, USA) *(APGV '06)*. Association for Computing Machinery, New York, NY, USA, 9–14. https://doi.org/10.1145/1140491.1140493

[30] Betty J Mohler, Sarah H Creem-Regehr, William B Thompson, and Heinrich H Bülthoff. 2010. The Effect of Viewing a Self-Avatar on Distance Judgments in an HMD-Based Virtual Environment. *Presence: Teleoperators and Virtual Environments* 19, 3 (2010), 230–242. https://doi.org/10.1162/pres.19.3.230

[31] Jodie M. Plumert, Joseph K. Kearney, James F. Cremer, and Kara Recker. 2005. Distance Perception in Real and Virtual Environments. *ACM Trans. Appl. Percept.* 2, 3 (July 2005), 216–233. https://doi.org/10.1145/1077399.1077402

[32] Fazlay Rabbi, Taiwoo Park, Biyi Fang, Mi Zhang, and Youngki Lee. 2018. When Virtual Reality Meets Internet of Things in the Gym: Enabling Immersive Interactive Machine Exercises. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 78 (July 2018), 21 pages. https://doi.org/10.1145/3214281

[33] E. D. Ragan, C. Wilkes, Y. Cao, and D. A. Bowman. 2012. The effects of virtual character animation on spatial judgments. In *2012 IEEE Virtual Reality Workshops (VRW)*. IEEE, Costa Mesa, CA, USA, 141–142.

[34] Rebekka S. Renner, Boris M. Velichkovsky, and Jens R. Helmert. 2013. The Perception of Egocentric Distances in Virtual Environments - A Review. *ACM Comput. Surv.* 46, 2, Article 23 (Dec. 2013), 40 pages. https://doi.org/10.1145/2543581.2543590

[35] Jannick P Rolland and Henry Fuchs. 2000. Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators & Virtual Environments* 9, 3 (2000), 287–309.

[36] Cynthia S. Sahm, Sarah H. Creem-Regehr, William B. Thompson, and Peter Willemsen. 2005. Throwing versus Walking as Indicators of Distance Perception in Similar Real and Virtual Environments. *ACM Trans. Appl. Percept.* 2, 1 (Jan. 2005), 35–45. https://doi.org/10.1145/1048687.1048690

[37] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman. 2007. Egocentric depth judgments in optical, see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics* 13, 3 (2007), 429–442.

[38] William B. Thompson, Peter Willemsen, Amy A. Gooch, Sarah H. Creem-Regehr, Jack M. Loomis, and Andrew C. Beall. 2004. Does the Quality of the Computer Graphics Matter when Judging Distances in Visually Immersive Environments? *Presence: Teleoperators and Virtual Environments* 13, 5 (2004), 560–571. https://doi.org/10.1162/1054746042545292

[39] Balasaravanan Thoravi Kumaravel, Fraser Anderson, George Fitzmaurice, Bjoern Hartmann, and Tovi Grossman. 2019. Loki: Facilitating Remote Instruction of Physical Tasks Using Bi-Directional Mixed-Reality Telepresence. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 161–174. https://doi.org/10.1145/3332165.3347872

[40] Koorosh Vaziri, Peng Liu, Sahar Aseeri, and Victoria Interrante. 2017. Impact of Visual and Experiential Realism on Distance Perception in VR Using a Custom Video See-through System. In *Proceedings of the ACM Symposium on Applied Perception* (Cottbus, Germany) *(SAP '17)*. Association for Computing Machinery, New York, NY, USA, Article 8, 8 pages. https://doi.org/10.1145/3119881.3119892

[41] M. E. Walker, D. Szafir, and I. Rae. 2019. The Influence of Size in Augmented Reality Telepresence Avatars. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Osaka, Japan, Japan, 538–546.

[42] Matthias Wille, Lars Adolph, Britta Grauel, Sascha Wischniewski, Sabine Theis, and Thomas Alexander. 2014. Prolonged Work with Head Mounted Displays. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program* (Seattle, Washington) *(ISWC '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 221–224. https://doi.org/10.1145/2641248.2645638

[43] Peter Willemsen, Mark B. Colton, Sarah H. Creem-Regehr, and William B. Thompson. 2009. The Effects of Head-Mounted Display Mechanical Properties and Field of View on Distance Judgments in Virtual Environments. *ACM Trans. Appl. Percept.* 6, 2, Article 8 (March 2009), 14 pages. https://doi.org/10.1145/1498700.1498702

[44] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[45] Tim Zindulka, Myroslav Bachynskyi, and Jörg Müller. 2020. Performance and Experience of Throwing in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3313831.3376639