Do Head-Mounted Display Stereo Deficiencies Affect 3D Pointing Tasks in AR and VR?

Anil Ufuk Batmaz[1], Mayra Donaji Barrera Machuca[2], Duc Minh Pham[3], Wolfgang Stuerzlinger[4]

School of Interactive Arts + Technology, Simon Fraser University

ABSTRACT

Most AR and VR headsets use stereoscopic displays to show virtual objects in 3D. However, the limitations of current stereo display systems affect depth perception through conflicting depth cues, which then also affect virtual hand interaction in peri-personal space, i.e., within arm's reach. We performed a Fitts' law experiment to better understand the impact of stereo display deficiencies of AR and VR headsets on pointing at close-by targets arranged laterally or along the line of sight. According to our results, the movement direction and the corresponding change in target depth affect pointing time and throughput; subjects' movements towards/away from their head were slower and less accurate than their lateral movements (left/right). However, even though subjects moved faster in AR, we did not observe a significant difference for pointing performance between AR and VR headsets, which means that previously identified differences in depth perception between these platforms seem to have no strong effect on interaction. Our results also help 3D user interface designers understand how changes in target depth affect users' performance in different movement directions in AR and VR.

Keywords: 3D pointing, virtual hand, selection, Fitts Law, AR, VR

Index Terms: Human-centered computing—Human-computer interaction (HCI)—Interaction paradigms—Virtual reality; Human-centered computing—Human-computer interaction (HCI)—Interaction paradigms—Mixed/Augmented Reality; Human-centered computing—Human-computer interaction (HCI)—Interaction techniques—Pointing; Human-centered computing—Interaction design;

1 INTRODUCTION

Most commercial virtual reality headsets (VR) and augmented reality see-through headsets (AR) use stereo displays to enable better spatial perception. To display 3D content, these headsets show two different images to the users' eyes from viewpoints that correspond to the eye positions in a human head. VR and AR headsets also allow users to directly manipulate virtual 3D objects using their hand or a controller and such interaction is easy to use and requires little training [6]. However, common stereo displays do not render spatial cues perfectly. Particularly for objects in peri-personal space, up to ~70 cm from the user, the human vision system can experience depth perception issues, including the vergence-accommodation conflict [15], diplopia [8], age-related near field vision problems [40,53] and personal stereo deficiencies.

Previous work by Barrera and Stuerzlinger [2] identified that the stereo display deficiencies of large 3D TVs affect virtual hand/wand pointing within arm's reach, but they did not investigate other types of stereo displays.

Thus, our primary goal in this work is to identify if the effect found by Barrera and Stuerzlinger [2] replicates with virtual hand/wand pointing in VR and AR headsets. We also hypothesize that the stereo display deficiencies affect interaction when using VR more than AR, because previous comparisons of depth perception in AR and VR headsets have found that VR depth perception is compressed relative to the real world [22,38] and that AR depth perception suffers less from overestimation [19,41,46]. Thus, targeting virtual objects in VR should be slower than in AR. A better understanding of how the depth perception issues of AR and VR headsets affect user interaction is also important because, as with real-world reaching movements, all targets in peri-personal space require users to accurately estimate the 3D position of the target for a rapid and successful selection. Knowing how stereo displays affect virtual hand pointing will also aid user interface designers to create 3D user interfaces with better performance and improve the user experience.

Barrera and Stuerzlinger [2] identified the change in target depth as the main predictor of movement time (MT) when using virtual hand/wand pointing in 3D TVs. Based on this observation, our experiment involves comparing VR and AR using four different movement directions, two with a strong change in visual depth and two without, in a Fitts' Law study. Fitts' law [13] predicts the movement time (MT), i.e., for how quickly people can point to a target. The Shannon formulation of Fitts' Law [42] is:

$$\mathrm{MT} = a + b * \log_2\left(\frac{D}{W} + 1\right) = a + b * \mathrm{ID} \qquad (1)$$

Where D and W are the target distance respectively size, while a and b are empirically derived via linear regression. The logarithmic term in Fitts' law is known as the index of difficulty (ID) and indicates the overall pointing task difficulty. Through comparing how stereo display deficiencies affect performance in AR and VR headsets our work extends the work of Barrera and Stuerzlinger [2]. We also verify if the differences of depth perception between AR and VR found by previous studies affect near distance pointing. For example, the work by Naceri et al. [31] found that participant performance diminished with HMDs compared to 3D TVs when comparing peri-personal depth perception, but they did not study augmented reality see-through headsets. Jones et al. [19] compared VR and AR headsets, but they only investigated distances between 3 and 7 m. In contrast, our experiment involves VR and AR headsets with targets within 70 cm of the user, which enables us to compare if these previously identified depth perception issues affect user interaction with virtual hand pointing across headsets.

2 RELATED WORK

Pointing at and selection of virtual 3D objects is a basic 3D user interaction method and a prerequisite for enabling users to manipulate such virtual objects [6]. Here, we focus on hand/wand pointing techniques, where the user needs to intersect the target

[1] e-mail: abatmaz@sfu.ca
[2] e-mail: mbarrera@sfu.ca
[3] e-mail: ducminhp@sfu.ca
[4] e-mail: w.s@sfu.ca

with their hand or wand in 3D space. This technique only applies to targets that are within arms' reach (in peri-personal space). In this section, we first review recent work on depth perception in VR and AR environments. Then we explore the pointing and hand movement literature for both display technologies.

2.1 VR and AR Depth Perception in Peri-personal Space

In the real world, the eyes focus at different distances in order to correctly "see" (visual) depth, whereas, in AR and VR headsets, the eyes must focus/accommodate on a 2D display at a fixed distance, which affects depth perception. For example, Hong and Kang [16] investigated stereoscopic fusion and found that accommodation was affected by the crossing points of the stereoscopic image only when these crossing points are close to the participant. Suryakumar et al. [44] also identified that the vergence angle affects the vergence time. Dutton et al. [11] and Durgin et al. [10] found that when there is a difference between vergence and focal distance, perceived depth is less accurate. Other work has also identified that depth perception is affected by depth cue conflicts [4,21,35]. Furthermore, previous work has shown that stereopsis, motion parallax, as well as convergence and accommodation all affect non-pictorial depth perception at less than 2 m distance [7,9,10,36,38].

Several researchers have studied depth perception limitations in VR, including Kenyon and Ellis [21], who stated that visual acuity and display resolution should match to provide an accurate image and that depth quantization should be avoided. In Renner et al.'s study [38], participants under-estimated virtual distances to be about 74% of the true distance in virtual environments. This difference was not connected to the VR display system, but it could be related to the individuals' vision system [14] or age [53]. In AR, Swan et al. [46] found that subjects overestimated the depth of objects in AR compared to their real target depth for targets close to the user (0.5 cm for targets at 34 cm and up to 2 cm for targets at 50 cm). Singh et al. [40] studied depth matching on near-field targets and found that both focal demand and brightness affect near-field AR depth matching.

2.2 Pointing in VR and AR

While stereo displays are beneficial for depth related tasks in the near field [24,28], stereoscopic views in virtual environments can compress distance perception in contrast to the real world [38]. Previous work on ray-pointing techniques in VR has found that depth cue conflicts affect pointing performance [18,47]. For 3D selection, virtual hand/wand techniques have been widely explored [5,8,37]. For example, Lin and Woldegiorgis [25] studied the effect of depth cue conflicts on virtual hand pointing performance and found that overestimation decreased with distance from the user. However, they only considered *distances beyond 65 cm*, i.e., outside peri-personal space. Barrera and Stuerzlinger [2] found that stereo display deficiencies affect pointing tasks with targets positions between 30 cm and 70 cm from the users on an 85" stereoscopic display. We are not aware of previous work that has studied how depth perception issues affect virtual hand selection in VR and AR headsets.

Moreover, in biomechanical studies of 3D pointing, researchers found that the plane of shoulder exertion affects the used muscles [1,27,43]. Further, hand movements that go over the mid-body line are more complicated than those that do not [32,39]. Lubos et al. [26] showed that when comparing motor actions and visual perception, visual perception has a more significant effect on subject performance [4], which is directly relevant for the mid-air pointing tasks without haptic feedback that we are exploring here.

3 Motivation

Previous work has found that vision impairments can affect visually guided motions [11] and that human lateral target discrimination is better than depth discrimination [49]. Both effects can affect virtual hand selection in peri-personal space in front of a large display, as found by Barrera and Stuerzlinger [2]. Based on their finding, our goal is to identify if stereo display deficiencies also affect pointing performance in AR and VR headsets. Understanding this effect is important because head-mounted stereo display systems are frequently used to display 3D scenes and are central to many VR and AR applications [23,30]. Most interactions in such systems let users directly interact with 3D objects, for example, to move one object from one place to another or to select an option in a menu. Thus, a better understanding of how stereo displays affect 3D pointing/selection is needed.

H1 – Effect of seeing the real world in target selection

We hypothesize that selection performance in peri-personal space is negatively impacted when users do not see the real world, i.e., a stylus and their hand, and rely on computer-controlled display systems. Based on our hypothesis, we expect that pointing tasks in AR headsets will exhibit better speed and throughput than the same tasks in VR headsets. The motivation for this hypothesis is that previous work has reported that depth perception is better in AR compared to VR when comparing 3D TVs and VR HMDIs [31], and for larger distances [19]. Ebrahimi et al. [12] also found that participants overestimated less when their participants saw the actual physical location of the stylus, but that overestimation grew with an offset. In our AR condition participants see the stylus at its actual physical location, but in the VR condition participants see only a 3D model of the wand. This can help participants establish a connection between the real and the virtual world in AR headsets [29]. Thus, we expect that this difference will affect 3D selection.

H2 – Effect of change in visual depth on AR and VR headsets on target selection

We hypothesize that selection performance in peri-personal space is negatively impacted when there is a change in depth between targets. Based on our hypothesis, we expect that pointing tasks with targets at different visual depths will exhibit different movement times and throughput compared to targets at the same visual depth. We base this hypothesis on previous work discussed above, and especially that such an effect was modeled by Barrera and Stuerzlinger [3] for large stereoscopic displays. On difference is that in a VR/AR headset, the screen is in front of the targets, while in a 3D TV the screen is behind the target. According to Kenyon and Ellis [21], this difference in display distance to the user leads to a different degree of accommodation changes. Thus, we believe also that VR and AR headsets might exhibit different pointing performance than 3D TVs. Since mid-air pointing interaction methods often do not provide any physical feedback (not even vibration), target selection is heavily based on visual feedback. Thus, in this work, we focus on the effect of visual depth perception in VR and AR.

4 User Study

The main goal of this study is to compare virtual hand pointing performance of users in VR and AR for different movement axes and to understand how target position affects user performance.

4.1 Methodology

Participants

We recruited 12 paid participants from the community (9 female). 6 of them were between 18 and 22 years old, 5 of them were 22-26 years, and one them was 30-34 years old. All participants measured normal when tested for stereo viewing capability and used their dominant hand for the task. All of them were familiar with interactive 3D content, typically through 3D video games, 10 of them played between 0-5 hours/week, and 2 of them 5-10 hours. 11 of them used 3D CAD systems between 0-5 hours/week, and one of them 5-10 hours. Among our participants, 10 had experienced VR and 6 had experienced AR, at least once before.

Apparatus

In this experiment, we used a PC with Intel Core i7-4790 with 16GB Ram and an NVIDIA GeForce GTX1080 graphics card running Windows 10. We chose commercial AR and VR headsets with (roughly) similar specifications, as follows:

Virtual Reality Headset (VR HMD): For the VR condition, we used an HTC Vive Pro headset (Figure 1a). For both eyes, the total display resolution for the headset is 2880 x1600 pixels with a refresh rate of 90 Hz. The HTC Vive Pro's field of view is 110° diagonally, and it weighs about 500 grams.

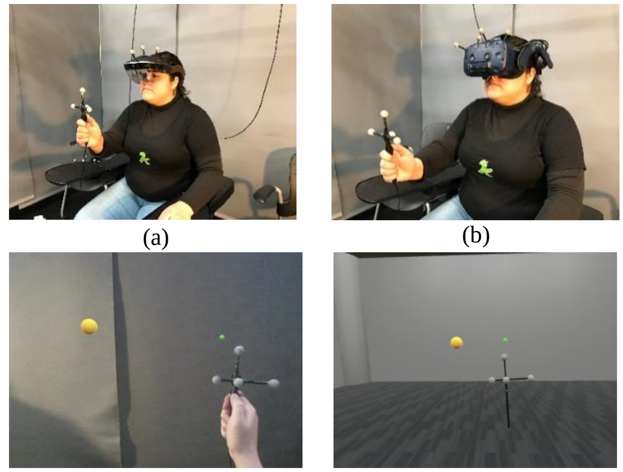
Augmented Reality See-through Headset (AR HMD): For the AR condition, we used a Meta 2 headset as shown in Figure 1b. The Meta AR headset has a field of view of 90°, which is large compared to many other current AR headsets. The total resolution of the headset is 2560x1440 with a refresh rate of 60 Hz. The Meta2 weighs about 550 grams.

Optical Tracking System: Both VR and AR headsets have their own tracking systems but use different technologies. The Meta2 uses RGB cameras and IR-based sensors to track the headset with inside-out positional tracking. The HTC Vive Pro uses two base station emitters to track the system with outside-in positional tracking. These different methods might cause differences in latency and accuracy. Thus, we decided to use a different tracking system to keep the quality tracking consistently between the AR and VR conditions. We used an OptiTrack tracking system, which is an external outside-in optical tracking system. We placed eight OptiTrack S250e, 250 Hz IR cameras on a 2.4 m x 1.8 m rectangular metal frame, mounted 2.1 m above the study setup area. Four optical markers were placed on each headset with different configurations to ensure reliable tracking (Figure 1). We used a dedicated tracking computer to collect and process all the OptiTrack camera data and send the tracking results to the main experimental system over the network. This configuration practically eliminated the potential for dropped frames on the main experimental system.

Calibration of Virtual Object Display: Our VR and AR headsets used different technologies to show the virtual object in space. To compensate for any potential differences in terms of display accuracy, we manually calibrated the display of both headsets to render the virtual targets at the correct physical distance from the user. For this, we measured the distance and angle between each target and each headset in the real-world through tape measurements and in the virtual world by the recording the transformation data provided by Unity3D. We then manually adjusted the output from the optical tracking system by multiplication with an adjustment transformation until the real-world and virtual word transformation data were the same.

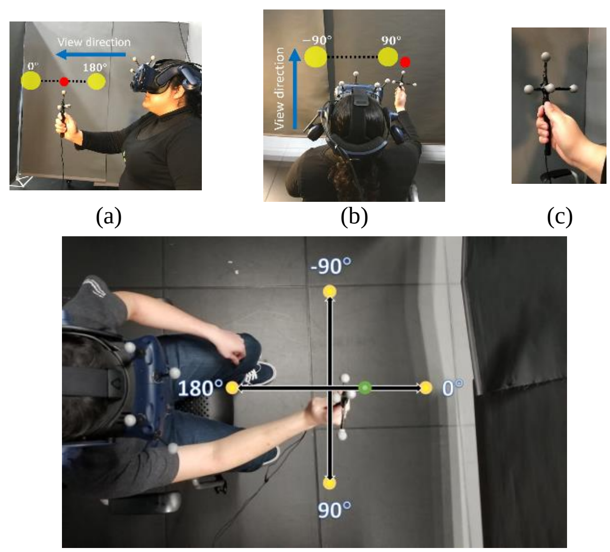
For movements in the view direction, target distances were between 40 and 70 cm from the user. For the lateral direction movements, all targets were displayed at 55 cm from the user, i.e., the midpoint of the range for movements in the view direction.

Wand: We designed a wand-like object as the pointing device (Figure 2c). A push button, similar to the click button on a regular mouse, was placed on the wand. IR markers were placed on the wand to enable tracking of the user's movements. The wand's dimensions were 21 cm x 23 cm x 5cm and it weighed 60 grams.



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

Figure 1 – Experimental setup with optical tracking markers. A user with the wand in the (a) AR headset and (b) VR headset. Participants' view for (c) AR and (d) VR.



(a) (b) (c)

(d)

Figure 2 – Movement Directions: (a) view direction, (b) lateral direction, (c) wand used by the participant, and (d) top view illustrating all movement directions.

Procedure

The procedure study was approved by the Ethics Board of Simon Fraser University. First, we screened participants to see if they could merge stereo targets correctly. After that participants signed an informed consent form and answered a demographics survey. Then participants were seated at the middle of the tracking area in a chair that does not allow for rotations (no swivel) to keep their body facing into a consistent direction during the study. Participants used their dominant hand to perform the task. In both AR and VR, the centers of each pair of targets were placed to be always at their eye level, for both lateral and depth movements (Figure 2a and 2b).

Participants were instructed only to move their arm while keeping their head and body in (approximately) the same position, to keep the view direction mostly constant. As mentioned above, and to eliminate the effect of vertical disparity, we matched the visual target height to each participants' eye-level. We did not adjust for individual interocular distance (IPD) in our user study but did take motion parallax through head tracking into account. We considered the field of view (FOV) for both headsets, where the VIVE Pro headset had a 20° larger (diagonal) FOV than the Meta2 headset and made sure that all the targets in our conditions were positioned so our participants could see both alternating targets without rotating their head.

After putting the headset on, the participants were instructed on the task and encouraged to practice until they felt comfortable with it. The task was a variant of the ISO 9241-411 task [17], where the targets are positioned along a single axis, either laterally or in depth (Figure 2a and 2b). The task involves selecting the 3D targets reciprocally as fast and as accurate as possible. In the VR headset, subjects were looking at a virtual gray wall, and a yellow (target) sphere was placed in front of them at eye-level. A green sphere was displayed 3 cm above the top of the virtual model of the wand and acted as the virtual cursor. In AR, we covered the experimental setup area with gray curtains and subjects saw the target, virtual cursor, and the real wand.

Participants pushed the button on the wand to select a target. For a valid selection, the virtual cursor had to be inside the target when the participant pushed the button. If the participant missed the target, the color of the target was changed to black to give visual feedback, and an error sound was played from the speakers of the PC. Then, subjects continued to the next target. Participants selected 11 targets in each set. Target size and distance between targets changed randomly for each new set. This configuration applied to both movement directions (lateral movement and depth movement). Movement direction and AR/VR headsets were counterbalanced across all subjects using a Latin square design. Participants were permitted to take a break of up to 60 seconds between the movement types and up to 3 minutes between the headset conditions. When participants got tired, they were allowed to take a break until they felt rested. The duration of such breaks did not exceed 2 minutes.

4.2 Experimental Design

The experiment consisted of 11 trials per experimental condition for 12 subjects with 72 experimental conditions: subjects executed the task with four different Target Positions (TP_4: -90°, 90°, and 0°, 180°), three Target Separations (TS_3: 10, 20, and 25 cm), and three different Object Sizes (OS_3: 1.5, 2.5, and 3.5 cm) for both Headsets (HS_2: AR and VR). We used these conditions in the study in a TP_4 x TS_3 x OS_3 x HS_2 within-subjects design. We measured movement time (milliseconds, ms), error rate (percentage of missed targets), throughput (bits per seconds, bps), and cursor speed (cm/s) as dependent variables. We also recorded path movement of the cursor for the movement analysis. We calculated nine distinct IDs (ID_9) by using Eq. (1) with three different object sizes (OS_3) and three different target separations (TS_3) which resulted in a range of IDs from 1.94 to 4.39. Each participant completed three repetitions for each ID for each movement direction (MD_2: lateral movement and depth movement). Thus, each subject completed MD_2 x HS_2 x ID_9 x 11 trials x 3 repetitions = 1188 trials. In total, we collected 14256 data points for each dependent variable.

4.3 Results

The results were analyzed using (RM) repeated measures ANOVA with $\alpha = 0.05$. When analyzing the data, we found that the distribution had a long tail and that the data was not normal (even after a log-transform). We believe this distribution shape might be a side effect of the stereo display deficiencies, especially because most of the data outside a 3σ criterion was for movements in the view direction. Based on this, we decided to exclude only double-clicks (0.8% of the data) and transformed the data with an Aligned Rank Transform (ART) [51] before the ANOVA. We used the interaction test as recommended by ART to compare the means.

At the start of the experimental data analysis, we identified that one of the participants had an abnormally high error rate with the VR headset (19%), with more than a dozen sets with 100% error rate. We believe that this participant stopped doing the task at some point towards the experiment. Consequently, we deleted this subject's data and analyzed the rest of the participant's results. Only significant results are described in detail here. Statistical results are reported in Table 1 and Table 2, with *** for $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, and N.S. not significant.

4.4 Single factor analyses

The general results for one-way effects are shown in Table 1 and Figure 4.

Table 1 – Single factor analysis results

	Mov. Direction		ID		Device	
	F (3,30)	p	F (8,88)	P	F (1,10)	p
Movement Time	27.7	***	128.7	***	0.16	N.S.
Error Rate	11.88	***	2.8	N.S.	1.12	N.S.
Throughput	44.5	***	15.2	**	0.024	N.S.
Target Re-Entry	4.3	**	17.5	**	0.05	N.S.
Speed	12.47	***	696.5	***	50.8	***
Ballistic Time	39.2	***	110	***	0.3	N.S.
Correction Time	9.01	***	169.3	***	2.15	N.S.

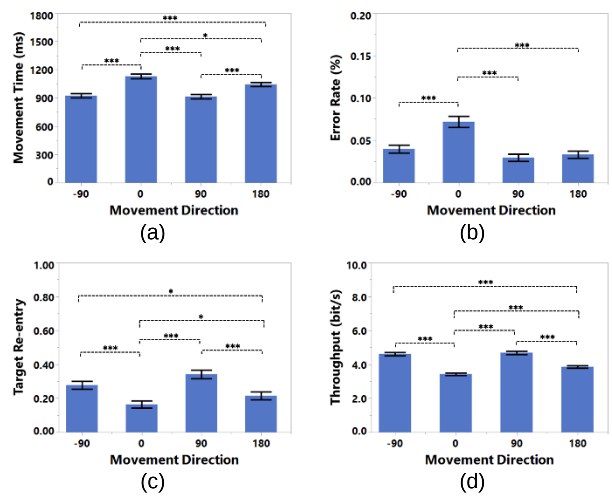


Figure 4 – Single factor analysis results for movement direction. (a) movement time, (b) error rate, (c) target re-entry and (d) throughput.

Movement Time: Overall, there was a significant main effect of direction on movement time (MT), (Figure 4a). Average MT in the lateral direction was significantly faster than the view direction. For movements in the view direction, the movements towards the user were faster than the movements away from the user. A post-hoc test identified the following groups, ordered from lower to higher speed: (90°, -90°), (180°), and (0°). There was no significant difference between devices for MT.

Error Rate: There was a significant main effect of movement direction on error rate (Figure 4b). The error rate for movements away from the user in the view direction was significantly higher than for lateral and towards the user movements. A post-hoc test found the following groups, ordered from lower to higher error rates: (-90°, 90° and 180°) and (0°). There was no significant difference between devices for error rate.

Throughput: (Effective) throughput (THP) was computed using the ISO 9241-411 method adapted for 3D motions. There was a significant effect of movement direction on throughput, see Figure 4d. THP for movements in the lateral direction was significantly faster than for the view direction. For movements in the view direction, the movements towards the user had higher THP than movements away from the user. A post-hoc test found the following groups from higher to lower THP: (90°, -90°), (180°), and (0°). There was no significant difference between devices for THP.

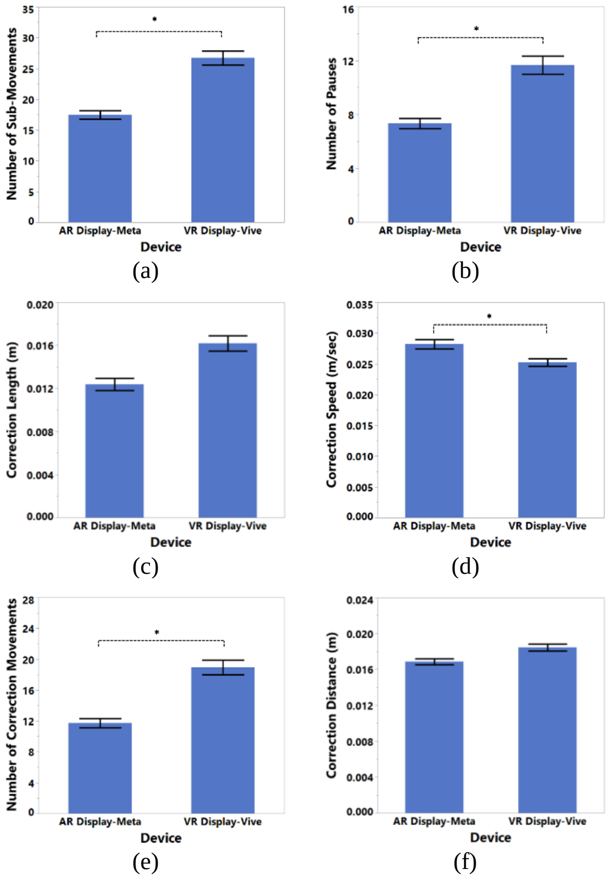


Figure 5 – Movement path analysis for both AR and VR devices (a) movement path length, (b) average number of submovements, (c) average number of pauses (d) correction distance (e) correction speed and (f) correction length.

Movement Path: We analyzed the movement paths using target re-entry events (Figure 4b), speed, ballistic and correction times. Ballistic and correction times were calculated with Nieuwenhuizen et al.'s method [33]. There was a significant main effect of movement direction on all measures (Table 1). Both movement directions are significantly different for target re-entry, speed and correction time, with lateral movements (90°, -90°) being "superior" in all measures compared to the view direction (0°, 180°). For ballistic times, all movement directions were different.

For the device, there was only a significant main difference in speed, where pointing in the AR HMD was faster than in the VR HMD. We also further analyzed the movement trajectories for both devices to better identify their differences regarding the number of sub-movements, number of pauses between sub-movements, path length, ballistic and correction path length as well as speed, and the corresponding numbers of sub-movements (Figure 5). All these values were again calculated using Nieuwenhuizen et al.'s method [33]. We found a statistically significant main difference between devices on the number of sub-movements ($F_{1, 10} = 8.78$, $p < 0.05$) and pauses ($F_{1, 10} = 6.76$, $p < 0.05$), where the VR HMD had more sub-movements and pauses than the AR HMD. For ballistic ($F_{1, 10} = 31.9$, $p < 0.001$) and correction speed ($F_{1, 10} = 8.29$, $p < 0.05$), both speeds where higher in the AR HMD than in the VR HMD. We also found a statistically significant main difference between devices for the number of ballistic sub-movements ($F_{1, 10} = 6.8$, $p < 0.05$) and corrective sub-movements ($F_{1, 10} = 5.6$, $p < 0.05$), here the VR HMD had more sub-movements than the AR HMD.

4.5 Interactions and interaction contrasts

We show the general results for two-way effects in Table 2 and Figure 6. We will only discuss the interaction between movement direction and device.

Table 2 – Interactions

	Mov. Direction x ID		Device x ID		Mov. Direction x Device	
	F (24.240)	p	F (8,88)	p	F (3,33)	p
Movement Time	7.07	*	0.48	N.S.	6.08	**.
Error Rate	1.25	N.S.	1.68	N.S.	3.35	*
Throughput	42.2	***	1.6	N.S.	5.3	***
Target Re-Entry	1.16	N.S.	6.2	*	0.9	N.S.
Speed	1.48	N.S.	3.6	N.S.	1.5	N.S.
Ballistic Time	11.02	***	5.4	*	4.7	**
Correction Time	2.6	N.S.	2.6	N.S.	5.9	**

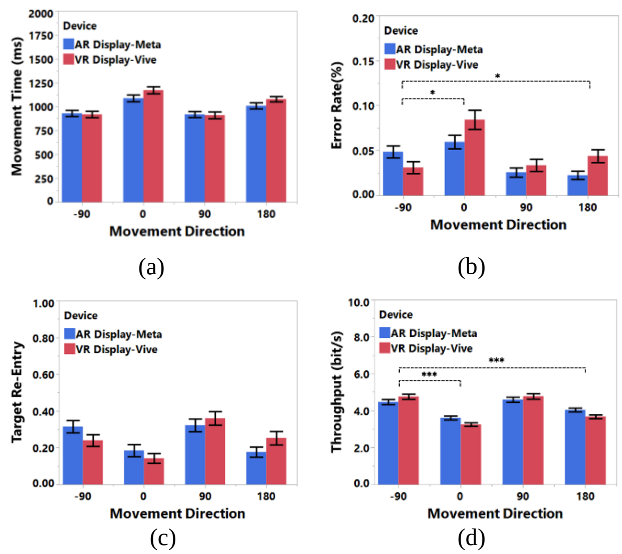


Figure 6 – Movement Direction and device interactions for (a) movement time, (b) error rate, (c) target re-entry and (d) throughput.

Movement Time: There was a significant interaction between movement direction and device for MT (Figure 6a). According to the results, within each movement direction, there was no significant difference between AR and VR headsets.

Error Rate: There was a significant interaction between movement direction and device for error rate (Figure 6b). In general, subjects had higher error rates with the VR headset compared to the AR headset in the view direction.

Target Re-entry: There was no significant interaction between movement direction and device for target re-entry (Figure 6c).

Throughput: There was a significant effect between movement direction and device for throughput (Figure 6d). In general, subjects had lower THP values for the VR headset in the view direction than for the AR condition and the lateral direction.

5 Discussion and Limitations

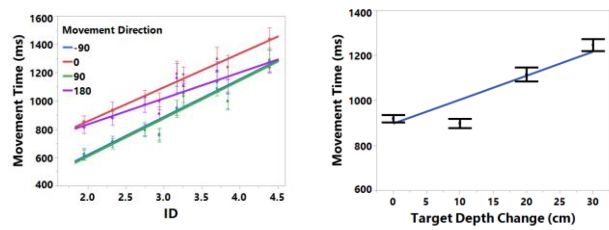
In this work, we explored how target depth and two different headsets affect human movement performance for a 3D pointing task with a wand.

In our first hypothesis, **H1**, we expected better speed and throughput with the AR headset because subjects were able to see the real environment including their hands and the wand. As stated in the motivation, previous studies found that participants' depth perception is different in VR and AR [19,31] and we hypothesized that this difference was also going to affect pointing performance. However, when we look at the results, movement time, error rate, and throughput were not significantly different between devices (Table 1). Still, we found a significant difference in speed between devices, where the VR HMD had a lower speed than the AR HMD. This higher speed with the AR HMD is coupled with a significantly larger number of sub-movements (Figure 5a) and pauses (Figure 5b) in the VR HMD. These results confirm the value of previous work that has found that speed and sub-movements are a good way to analyze how sure are users of their hand movements [33].

Based on these results we decided to investigate the movement trajectories of the users in more depth. Especially for targeting in peri-personal space, where the accommodation-vergence conflict, dipoplia, and stereopsis can influence depth perception more [9], one would expect to see more (unintended) movement corrections that occur after the initial (ballistic) movement. And indeed, we found differences between the speed (Figure 5d) and the number of sub-movements (Figure 5e), where the correction phase was less efficient in VR than when using AR. A similar difference was also found between devices when comparing ballistic speed and number of sub-movements in the ballistic phase. We also found a trend between correction path length and correction distance, i.e., the distance from the last ballistic phase position to the target (Figures 5f and 5d). However, this reduced efficiency is not translated to a difference in the whole movement time and throughput, as the length of corrective movements is small compared to the whole hand movement, on average 1.6 cm in VR and 1 cm in the AR condition. Another reason could be the number of movements without a correction phase, 14% for the AR HMD and 12% for the VR HMD.

These results partially agree with our **H1**, that seeing the real world had a positive effect on target selection. Our findings thus partially contradict and partially provide additional detail for the results from previous work [31]. However, the differences we found here are not large enough to affect overall user performance, even if both ballistic and correction phases are less efficient in VR. Other explanations for our results could be the additional depth perception cues afforded by AR headsets since any imperfection in AR headset displays can cause conflicts between real and virtual depth cues [45]. During the pointing task, the user typically focuses on the target, but can get additional cues from the environment. In VR, we only rendered the wand, which is similar to the Vive VR controllers. The pictorial cues, texture, lighting, shading and even the surfaces of the real world could enable the user to perceive depth more accurately [48,52]. The results could also be affected by the virtual representation of the user. Recent work by Tran et al. [50] showed that variations in virtual hand representation can affect user performance for selection tasks in VR. In our AR condition, users also could see their upper limb, and this might provide additional visual information about the wand position and movement as well as better motion parallax. Additionally, users might feel more comfortable when seeing (parts of) their real body, which could also increase their feeling of presence and their confidence during the task execution. However, all such speculations need to be verified in future work.

In our second hypothesis **H2**, we anticipated that a depth change would have a negative impact on selection performance. Like in the previous work of Barrera and Stuerzlinger [2], we were expecting to observe worse user performance with an increase in target depth change. Our results indeed show that subjects get slower when the change in target depth was larger. We observed that in both headsets lateral movements had noticeable better performance than those in the view direction. When analyzing the slope of each view direction for MT vs. ID, lateral movements show similar behaviors, while for movements in the view direction are different (Figure 7a). This effect is also visible in the throughput measure, where movements in the view direction have consistently about 22% fewer bps than lateral ones. When analyzing the effect of depth change in pointing (Figure 7b), we found a similar linear relationship (r^2=0.86) between an increase in target depth change and time as did Barrera and Stuerzlinger [3]. These findings support hypothesis **H2**, as selecting targets with similar IDs but in different movement directions exhibit different performance, and the gap increases with higher IDs.



(a) (b)

Figure 7 – (a) Fitts' law model for movement direction and (b) movement time vs target depth change.

However, our analysis of movement paths did not identify a similar pattern as Barrera and Stuerzlinger [3], where participants did a second ballistic movement instead of a correction movement in the view conditions. For depth movements, only 5% of the data (2.9% for AR HMD and 4.3% for the VR HMD) had a high correction phase speed (more than 20% of maximum). These values are similar to the ones for lateral movements, where only 3% of the data (2.5% for AR HMD and 2.7 for the VR HMD) showed this pattern. In contrast to our results, Barrera and Stuerzlinger [3] found that 12% of their data in the depth movements showed this pattern. Additionally, by comparing their results with ours, we can observe that VR and AR headsets exhibit different pointing performance than 3D TVs.

A limitation of our work is that while we identify that a target depth change affects user performance in 3D pointing in AR and VR headsets, we did not adjust for the interpupillary distance (IPD) for each individual in our experiment. One of the motivations was that we have seen few users "in the wild" adjust their IPD correctly. Moreover, we observed performance variations between further away and close targets, and the participant's performance was worse for further away targets. If the IPD was a major user performance issue, one could expect to observe that the performance would decrease for closer targets, which is exactly the opposite of our observations. Thus, we believe that any potential issue around the IPD had a limited impact on our results. For similar reasons, we did not limit the VR HMD FOV to match the AR HMD, like previous work by Nilsson et al. [34] and Jones et al. [20] has done, as we wanted to evaluate the capabilities of current headsets. Also, our task did not rely on peripheral vision, as we made sure that all the targets were visible, even the ones with the largest separation, in both headsets and the participants' visual focus needed to be on the targets to perform the task efficiently. Further, since our work aimed to extend the previous findings of Barrera and Stuerzlinger [2,3], we wanted to replicate the exact target distances they applied in their pointing task. This was also a reason for us to use headsets which provide an FOV large enough to display targets at the same distances in both conditions.

In the experiments, we also did not use the internal head-tracking of the AR and VR headsets. Both systems use IMU data to improve the latency of the visual rendering, but as mentioned above, we did not use this data in our experiment. While we increased the latency by ignoring such input, this helped us to perform a fair(er) comparison between the two headsets that is unaffected by the peculiarities of each tracking system.

6 Conclusion

We conducted a user study to investigate the effect of VR and AR headsets on 3D virtual hand pointing. We identify that both headsets had similar performance, but that depth perception issues had a more significant impact on VR than AR. We also identify that the change in target depth affects virtual hand interaction in peri-personal space with VR and AR headsets. Overall, our results match Barrera and Stuerzlinger [3]'s results for 3D TVs and generalize their work towards current VR and AR headsets. The results also provide support the need for better depth display functionality in VR and AR headsets, since interaction is negatively affected by current VR/AR headsets compared to the real world. Finally, we also verified Nieuwenhuizen et al.'s [33] findings that it is important to investigate the movement trajectories to better understand any differences, as not all effects are visible in time and throughput measures.

References

[1] N. T. Antony and P. J. Keir, "Effects of posture, movement and hand load on shoulder muscle activity". *J. Electromyogr. and Kinesiol.* vol. 20, no. 2, pp.191–198. Apr. 2010.

[2] M. D. Barrera Machuca and W. Stuerzlinger, "Do Stereo Display Deficiencies Affect 3D Pointing?" *Ext. Abstr. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI '18)*, pp. 1–6, 2018.

[3] M. D. Barrera Machuca and W. Stuerzlinger, "The Effect of Stereo Display Deficiencies on Virtual Hand Pointing," in *Proceedings of SIGCHI Conf. Hum. Factors Comput. Syst. (CHI '19)*, 2019.

[4] G. P. Bingham, A. Bradley, M. Bailey, and R. Vinner, "Accommodation, occlusion, and disparity matching are used to guide reaching: A comparison of actual versus virtual environments," *J. Exp.l Psychol. Hum. Percept. Perform., vol.* 27, no. 6: pp. 1314–1334, 2001.

[5] J. Boritz and K. S. Booth. "A study of interactive 3D point location in a computer simulated virtual environment," in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '97)*, 1997, pp. 181–187.

[6] D. A. Bowman, E. Kruijff, J. J. LaViola Jr., and I. Poupyrev, *3D User Interfaces: Theory and Practice*, 1st ed. Addison-Wesley, 2004.

[7] J. M. Brown and N. Weisstein, "A spatial frequency effect on perceived depth," *Percept. Psychophys., vol.* 44, no. 2, pp. 157–166, Mar. 1988.

[8] G. Bruder, F. Steinicke, and W. Stuerzlinger, "Effects of visual conflicts on 3D selection task performance in stereoscopic display environments," in *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI '13)*, 2013, pp. 115–118.

[9] J. E. Cutting and P. M. Vishton, "Perceiving Layout and Knowing Distances: The Integration, Relative Potency and Contextual Use of Different Information about Depth," in *Handbook of perception and cognition, Vol 5; Perception of space and motion*, San Diego, California, USA: Academic Press, 1995, pp. 69–117.

[10] F. H. Durgin, D. R. Proffitt, T. J. Olson, and K. S. Reinke, "Comparing depth from motion with depth from binocular disparity," *J. Exp. Psychol. Hum. Percept. Perform., vol.* 21, no. 3, pp. 679–699, 1995.

[11] G. N. Dutton *et al*., "Association of binocular lower visual field impairment, impaired simultaneous perception, disordered visually guided motion and inaccurate saccades in children with cerebral visual dysfunction—a retrospective observational study," *Eye, vol.* 18, no. 1, pp. 27-34, Jan. 2004.

[12] E. Ebrahimi, *et al*., "Effects of visual and proprioceptive information in visuo-motor calibration during a closed-loop physical reach task in immersive virtual environments" in P*roceedings of the ACM Symposium on Applied Perception (SAP'14)*: pp. 103–110, 2014.

[13] P. M. Fitts and J. R. Peterson, "Information capacity of discrete motor responses," *J. Exp. Psychol., vol.* 67, no. 2, pp. 103–112, 1964.

[14] R. F. Hess, L. To, J. Zhou, G. Wang, and J. R. Cooperstock, "Stereo Vision: The Haves and Have-Nots," *I-Perception, vol.* 6, no. 3, pp. 1–5, Jun. 2015.

[15] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *J. Vis., vol.* 8, no. 3, pp. 33.1-30, Mar. 2008.

[16] H. Hong and S. H. Kang, "Measurement of the lens accommodation in viewing stereoscopic displays," *J. Soc. Inf. Disp.*, vol. 23, no. 1, pp. 19–26, Jan. 2015.

[17] ISO, "ISO 9241-411:2012 Ergonomics of human-system interaction - Part 411: Evaluation methods for the design of physical input devices" 2012.

[18] I. Janzen, V. K. Rajendran, and K. S. Booth, "Modeling the Impact of Depth on Pointing Performance," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*, 2016, pp. 188–199.

[19] A. Jones, J. E. Swan, G. Singh, and E. Kolstad, "The Effects of Virtual Reality, Augmented Reality, and Motion Parallax on Egocentric Depth Perception," in *2008 IEEE Virtual Reality Conference*, 2008, pp. 267–268.

[20] J. A. Jones, J. E. Swan, and M. Bolas, "Peripheral stimulation and its effect on perceived spatial scale in virtual environments," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 4, pp. 701–710, 2013.

[21] R. V. Kenyon and S. R. Ellis, "Vision, Perception, and Object Manipulation in Virtual Environments," in *Virtual Reality for Physical and Motor Rehabilitation*, P. L. Weiss, E. A. Keshner, and M. F. Levin, Eds. New York, NY: Springer New York, 2014, pp. 47–70.

[22] M. Kleiber and C. Winkelholz, "Distortion of depth perception in virtual environments using stereoscopic displays: quantitative assessment and corrective measures," in *Proc. SPIE 6803, Stereoscopic Displays and Applications XIX*, 2008, vol. 6803, p. 68030C.

[23] Leap Motion, "Orion." 2016.

[24] C. J. Lin and B. H. Woldegiorgis, "Interaction and visual performance in stereoscopic displays: A review," *J. Soc. Inf. Disp.*, vol. 23, no. 7, pp. 319–332, Jul. 2015.

[25] C. J. Lin and B. H. Woldegiorgis, "Egocentric distance perception and performance of direct pointing in stereoscopic displays," *Appl. Ergon.*, vol. 64, pp. 66–74, Oct. 2017.

[26] P. Lubos, G. Bruder, and F. Steinicke, “Analysis of direct selection in head-mounted display environments,” *Proc. IEEE Symp. 3D User Interfaces*, pp. 11–18, 2014.

[27] A. McDonald, B. R. Picco, A. L. Belbeck, A. Y. Chow, and C. R. Dickerson, “Spatial dependency of shoulder muscle demands in horizontal pushing and pulling,” *Appl. Ergon.*, vol. 43, no. 6, pp. 971–978, Nov. 2012.

[28] J. P. McIntire, P. R. Havig, and E. E. Geiselman, “Stereoscopic 3D displays and human performance: A comprehensive review,” *Displays*, vol. 35, no. 1, pp. 18–26, Jan. 2014.

[29] D. Medeiros, M. Sousa, D. Mendes, A. Raposo, and J. Jorge, “Perceiving Depth: Optical versus Video See-through,” in *Proceedings of the Conference on Virtual Reality Software and Technology (VRST ’16)*, 2016, pp. 237–240.

[30] MetaVision, “MetaVision desktop.” 2012.

[31] A. Naceri, R. Chellali, F. Dionnet, and S. Toma, “Depth Perception Within Virtual Environments: Comparison Between two Display Technologies,” *Int. J. Adv. Intell. Syst.*, vol. 3, pp. 51–64, Nov. 2010.

[32] K. Newell C., *The Slow Learner in the Classroom*. Columbus: C. E. Merrill Books, 1962.

[33] K. Nieuwenhuizen, J.-B. Martens, L. Liu, and R. van Liere, “Insights from Dividing 3D Goal-Directed Movements into Meaningful Phases,” *IEEE Comput. Graph. Appl.*, vol. 29, no. 6, pp. 44–53, Nov. 2009.

[34] N. C. Nilsson, S. Serafin, and R. Nordahl, “Establishing the range of perceptually natural visual walking speeds for virtual walking-in-place locomotion,” *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 4, pp. 569–578, 2014.

[35] R. Patterson, “Human Factors of 3-D Displays,” *J. Soc. Inf. Disp.*, vol. 15, no. 11, pp. 861–871, 2007.

[36] R. Patterson and W. L. Martin, “Human stereopsis.,” *Hum. Factors J. Hum. Factors Ergon. Soc.*, vol. 34, no. 6, pp. 669–92, Dec. 1992.

[37] M. Pfeiffer and W. Stuerzlinger, “3D virtual hand pointing with EMS and vibration feedback,” in *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI’15)*, 2015, pp. 117–120.

[38] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert, “The perception of egocentric distances in virtual environments - A review,” *ACM Comput. Surv.*, vol. 46, no. 2, pp. 1–40, Nov. 2013.

[39] W. N. Schofield, “Do children find movements which cross the body midline difficult?,” *Q. J. Exp. Psychol.*, vol. 28, no. 4, pp. 571–582, 1976.

[40] G. Singh, S. R. Ellis, and J. E. Swan, “The Effect of Focal Distance, Age, and Brightness on Near-Field Augmented Reality Depth Matching,” *IEEE Trans. Vis. Comput. Graph.*, vol. PP, no. c, p. 1, 2018.

[41] G. Singh, J. E. Swan, J. A. Jones, and S. R. Ellis, “Depth judgment measures and occluding surfaces in near-field augmented reality,” in *Proceedings of the Symposium on Applied Perception in Graphics and Visualization (APGV ’10)*, 2010, pp. 149–156.

[42] R. W. Soukoreff and I. S. MacKenzie, “Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts’ law research in HCI,” *Int. J. Hum. - Comput. Stud.*, vol. 61, no. 6, pp. 751–789, Dec. 2004.

[43] H. Strasser and K.W. Müller, “Favorable movements of the hand-arm system in the horizontal plane assessed by electromyographic investigations and subjective rating,” *Int. J. Ind. Ergon.*, vol. 23, no. 4, pp. 339–347, Mar. 1999.

[44] R. Suryakumar, J. P. Meyers, E. L. Irving, and W. R. Bobier, “Vergence accommodation and monocular closed loop blur accommodation have similar dynamic characteristics,” *Vision Res.*, vol. 47, no. 3, pp. 327–337, 2007.

[45] J. E. Swan, A. Jones, E. Kolstad, M. A. Livingston, and H. S. Smallman, “Egocentric depth judgments in optical, see-through augmented reality,” *IEEE Trans. Vis. Comput. Graph.*, vol. 13, no. 3, pp. 429–442, 2007.

[46] J. E. Swan, G. Singh, and S. R. Ellis, “Matching and Reaching Depth Judgments with Real and Augmented Reality Targets,” *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 11, pp. 1289–1298, Nov. 2015.

[47] R. J. Teather and W. Stuerzlinger, “Pointing at 3d target projections with one-eyed and stereo cursors,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’13)*, 2013, p. 159.

[48] J. T. Todd and E. Mingolla, “Perception of surface curvature and direction of illumination from patterns of shading,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 9, no. 4, pp. 583–595, 1983.

[49] J. J. Tramper and S. Gielen, “Visuomotor coordination is different for different directions in three-dimensional space.,” *J. Neurosci.*, vol. 31, no. 21, pp. 7857–7866, 2011.

[50] T. Q. Tran, H. Shin, W. Stuerzlinger, and J. Han, “Effects of virtual arm representations on interaction in virtual environments,” in *Proceedings of the Symposium on Virtual Reality Software and Technology (VRST ’17)*, 2017, pp. 1–9.

[51] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins, “The aligned rank transform for nonparametric factorial analyses using only anova procedures,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’11)*, 2011, p. 143

[52] M. J. Young, M. S. Landy, and L. T. Maloney, “A perturbation analysis of depth perception from combinations of texture and motion cues,” *Vision Res.*, vol. 33, no. 18, pp. 2685–2696, 1993.

[53] C. M. Zaroff, M. Knutelska, and T. E. Frumkes, “Variation in Stereoacuity: Normative Description, Fixation Disparity, and the Roles of Aging and Gender,” *Invest. Ophthalmol. Vis. Sci.*, vol. 44, no. 2, p. 891, Feb. 2003.