The Effect of the Vergence-Accommodation Conflict on Virtual Hand Pointing in Immersive Displays

Anil Ufuk Batmaz
aubatmaz@khas.edu.tr
Kadir Has University
Istanbul, Turkey

Mayra Donaji Barrera Machuca
mbarrera@dal.ca
Dalhousie University
Halifax, Canada

Junwei Sun
junwei.sun@huawei.com
Huawei HMI Lab
Markham, Canada

Wolfgang Stuerzlinger
w.s@sfu.ca
Simon Fraser University
Vancouver, Canada

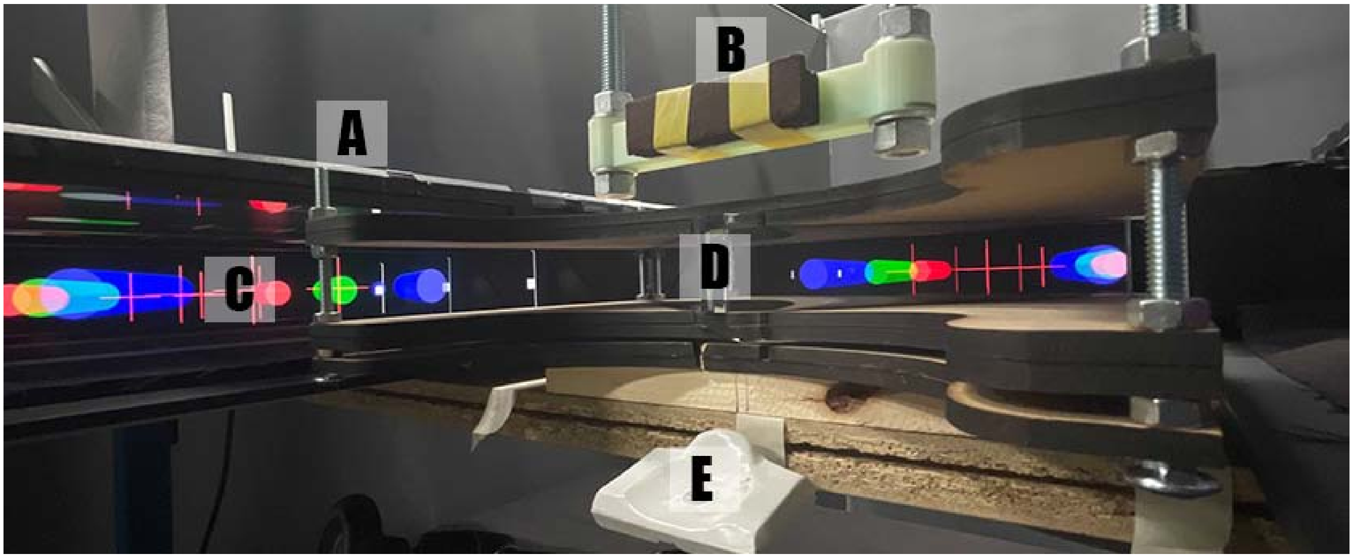


Figure 1: The Multifocal display. (A) 4K 32" monitor that renders the content. (B) Head support. (C) Beam splitters to reflect the three monitor images towards the central visor. (D) Visor with beam splitter to reflect the image towards the user. (E) Chin rest.

ABSTRACT

Previous work hypothesized that for Virtual Reality (VR) and Augmented Reality (AR) displays a mismatch between disparities and optical focus cues, known as the vergence and accommodation conflict (VAC), affects depth perception and thus limits user performance in 3D selection tasks within arm's reach (peri-personal space). To investigate this question, we built a multifocal stereo display, which can eliminate the influence of the VAC for pointing within the investigated distances. In a user study, participants performed a virtual hand 3D selection task with targets arranged laterally or along the line of sight, with and without a change in visual depth, in display conditions with and without the VAC. Our results show that the VAC influences 3D selection performance in common VR and AR stereo displays and that multifocal displays have a positive effect on 3D selection performance with a virtual hand.

CCS CONCEPTS

• **Human-centered computing → Pointing**.

KEYWORDS

3D pointing, virtual hand, selection, Fitts' Law, vergence-accommodation conflict

ACM Reference Format:
Anil Ufuk Batmaz, Mayra Donaji Barrera Machuca, Junwei Sun, and Wolfgang Stuerzlinger. 2022. The Effect of the Vergence-Accommodation Conflict

---

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3502067

on Virtual Hand Pointing in Immersive Displays. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3491102.3502067

1 INTRODUCTION

Recent technological advances have yielded head-mounted displays (HMDs) with low latency in rendering and tracking, and with wide field-of-views (FOVs) that allow for seamless viewing at up to 4k resolution [51]. To create a more realistic experience, modern HMDs also allow users to change the distance between the lenses to match their inter-pupillary distance (IPD), i.e., the distance between the center of the pupils of the eyes. Examples of such HMDs include the Varjo VR-3 [97] for Virtual Reality (VR) and the Canon MReal S1 [14] for Augmented Reality (AR). Regardless of these advances, current VR and AR HMDs still do not display all depth cues correctly, as identified by previous work [5, 60, 84, 88], which prevents users from correctly identifying an object's position in space. Some of the possible causes for depth perception problems are related to the human vision system and include age-related near field vision problems [103], personal stereo deficiencies [32], or diplopia [12], i.e., seeing two images of a single object. Other possible causes are related to the technology, like screen brightness [87] and the way an HMD displays the content.

In this paper, we focus on a specific depth perception problem that is due to the way VR and AR HMDs show content in stereo. To display 3D content, stereo displays show two different images to the users' eyes from viewpoints that correspond to the eye positions in a human head. Each image is displayed/projected at a *fixed* plane by the headset, typically on a 2D screen for VR or through a beam splitter for AR (potentially together with some lenses). Thus, when displaying 3D content that is not at the same depth as said fixed plane, a user's eye is exposed to a mismatch between focusing on the display plane (accommodation) and rotating the eyes to see the object at its correct visual depth (vergence). This problem is called the *vergence-accommodation conflict* (VAC) and does not happen with targets in the real world. One way to address the VAC is to correct the focal plane, e.g., by actively shifting the location of a single (or multiple) display planes based on the output of a gaze tracker, e.g., in varifocal displays [22]. Another way to address the VAC is by using more than one display planes, as demonstrated in some HMDs, e.g., the Magic Leap [55] and others [102]. However, these commercially available displays are limited to two focal planes, which is insufficient to cover the distances that are most important for 3D interaction. Most importantly, we were unable to find any investigation of how multifocal displays affect 3D virtual hand interaction in peri-personal space.

The goal of our work is to quantify the effect of the VAC on 3D selection performance when using stereo displays. We aim to identify if multifocal displays (which do not exhibit a VAC) cancel the negative effect of the depth perception issues that affect interaction in singlefocal stereo displays. Knowing about the effect of the VAC on interaction is important because stereo display systems are frequently used to display 3D scenes and are central to many VR and AR HMDs [59, 79, 84]. Therefore, identifying how this technology affects interaction with 3D objects can motivate technical innovation and lead to better 3D user interfaces.

We hypothesize that a stereo display without the VAC will not exhibit the reduction in performance for 3D target selection identified in previous work, e.g., by Barrera et al. [8] and Batmaz et al. [9]. Barrera et al. [8] identified that movements along the line of sight are 25% slower than movements in the lateral plane with stereo displays. Similarly, in a comparison between both movement directions in AR and VR headsets, Batmaz et al. [9] found a significant difference in throughput, a human performance measure that accounts for both speed and accuracy [89]. Ray-casting methods are also negatively affected by a change in target depth [30, 59]. The known differences in terms of depth perception between physical versus virtual targets [92] also indicate that interaction might be affected by the VAC. In these previous studies, the authors speculated that the most likely cause of the observed effects would be the VAC. Yet, none of them could not verify this as the display systems in their studies did not address the VAC.

To investigate the effect of the VAC, we used a Fitts' law task along two different movement directions, one with a pronounced change in visual depth and one without, using a custom-made stereo display that offers both multifocal and singlefocal display in VR and AR (Figure 1). Fitts' law [27] predicts the movement time (MT), i.e., how quickly people can point to a target. The most used formulation [61] of Fitts' law is:

$$MT = a + b * log_2(D/W + 1) = a + b * ID \tag{1}$$

In the above equation 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. Fitts' law holds not only in one dimension, but also in 2D, e.g., [62], and 3D, e.g., [93, 94].

By comparing pointing to targets at different depths using a stereo display with and without the VAC, we can identify if the VAC is the cause of the interaction slowdowns in current stereo displays. Quantifying the effect of the VAC on user performance in peri-personal space, up to ≈70 cm from the user, is important as in many consumer-level VR/AR applications users interact with objects inside this range. Our work extends previous work on depth perception that identified that stereo display deficiencies affect interaction [8, 9]. It also extends work that investigates the VAC [33, 48, 53]. Our contributions are:

- **A study of the effect of the vergence-accommodation conflict (VAC) on interaction:** We evaluate virtual 3D target selection in peri-personal space using our custom-made stereo display system, which shows either single- or multifocal images (See Figure 7). In our study, we evaluated user performance when selecting two pairs of targets, one with no change in depth and the other along the line of sight with a strong change in depth. We identified that the VAC significantly affects 3D selection performance, and adversely affects user performance for virtual hand interaction for targets at different visual depths.
- **A study of interaction with multifocal displays:** We describe the first virtual hand selection study involving a multifocal stereo display to evaluate user 3D selection performance. Our results confirm that virtual hand selection with

multifocal displays improves user performance compared to singlefocal displays. Compared to VR multifocal displays, AR multifocal displays are also less affected by target depth changes. Finally, we also verify that Fitts' Law applies to virtual hand selection with multifocal displays.

- **A custom VR/AR display apparatus for mid-air user interaction:** Inspired by Akeley et al.'s [2] prototype, we built a VR/AR multifocal stereo display apparatus (Figure 1) that eliminates the VAC at several distances relevant to our work on virtual hand pointing in peri-personal space. Our goal was to create a controllable, yet flexible stereo display device for our experiment that uses three different mirrors positioned at specific distances for displaying the targets used. The three mirrors support three linearly distributed depth planes, where virtual objects can be displayed without suffering from the VAC. This display enables interaction mid-air objects, as required for our user study.
- **A novel design for 3D calibration markers for stereo display systems:** Inspired by the ubiquitous checkerboard calibration pattern, we present a novel double-conical marker that affords accurate calibration of multiple depth planes in stereo display systems.

2 RELATED WORK

In this section, we first review recent work on depth perception in virtual environments (VE) with stereo displays. Then, and again for stereo displays, we summarize the literature on virtual hand selection of 3D objects, i.e., pointing movements that involve putting a 3D cursor inside a target before pressing a button to select it. These sections also demonstrate an absence of work on the effect of the VAC on user performance for 3D target selection in peri-personal space. Finally, we discuss previous work on stereo display systems that prevent the VAC.

2.1 Depth Perception in Peri-personal Space

The human visual system bases depth perception on pictorial and non-pictorial cues [19]. Examples of pictorial depth cues are occlusion and relative size [19]. In this paper, we focus on non-pictorial depth cues that do not depend on the shapes in the scene [84]. For distances less than 2 meters, non-pictorial depth perception is based on stereopsis, motion parallax, (con)vergence, and accommodation [11, 23, 81, 84]. Stereopsis is the perception of depth based on retinal disparity, i.e., the horizontal distance of the image seen by each eye [81]. Motion parallax is a continuous sampling of two-dimensional perspectives in the form of relative motions that support depth perception [23]. Vergence is the simultaneous (inward & outward) rotational movement of the eyes when there is a change of the target distance, while accommodation is the change in the (eye) lens curvature to focus on objects at various distances [84]. For real targets, the human visual system couples vergence and accommodation.

The union of pictorial and non-pictorial cues makes depth perception with real targets accurate, with typically less than 1.4 mm of error for objects in peri-personal space [92]. However, previous work has identified that common stereo displays suffer from depth perception issues. For example, Renner et al. [84] reviewed previous work on human depth perception in VR systems and identified a mean underestimation, 74% of the actual distance, independent of the VR display system used. Similarly, Swan et al. [92] examined how humans perceive and estimate target depth in VEs in comparison to physical ones, and found that users overestimate distances in AR. In other words, when using stereo displays, people believe that objects may be closer than they are displayed as. The specific stereo display deficiencies that cause this effect have not been identified [58], but previous work has postulated the VAC and the display resolution as potential causes.

2.1.1 Vergence-Accommodation Conflict (VAC). The presence of a VAC causes several problems in the human ocular system. First, the VAC affects depth perception. For example, Dutton et al. [24] and Durgin et al. [23] found that when there is a difference between vergence and focal distance, perceived depth is less accurate. Another problem caused by the VAC is visual fatigue. For instance, Hong and Kang [34] investigated stereoscopic fusion for objects at different visual depths, i.e., distances from the screen. They found that for close locations, the view direction of both eyes intersects outside the range of the depth-of-field, which affects accommodation. They concluded that this effect implies potential fatigue due to the VAC. Hoffman et al. [33] also identified visual fatigue due to the reduced stereo-acuity caused by the differences between focal and vergence distances. Similar fatigue effects of the VAC were found by Banks et al. [7], Bando et al. [6], Oh and Lee [77], and Iskander et al. [39]. The VAC also affects the performance of the visual system. For example, it lowers the speed of binocular fusion and increases the vergence latency [98], making people overshoot their vergence movements for non-blurred images [28]. It also causes the eyes to converge closer than required [37, 39]. Finally, the VAC also affects the cognitive load of the user [20]. One limitation of all previous research mentioned here is that it focused only on visual perception, but not on the consequences of the VAC on interaction with the displayed content.

2.1.2 Display Resolution. Kenyon and Ellis [45] stated that visual acuity and display resolution should match or exceed human limits to provide an accurate image. They also noted that depth quantization should be avoided; as quantization in digital systems, like stereo displays, effectively reduces the depth "resolution" of a system. Finally, Singh et al. [87] studied depth matching of near-field targets and found that both focal demand and brightness affect near-field AR depth matching. Foveated displays address this problem by matching the resolution characteristics of a human eye both within the fovea and at the periphery [47], but such systems are outside the scope of our work.

2.2 3D Selection in Stereo Display Systems in VR and AR

Previous work [59, 69] found that stereo displays are beneficial for depth-related tasks in the near-field with virtual ray techniques [50, 94]. However, pointing throughput is typically well below what users can achieve in 2D tasks [89, 93, 94].

One likely explanation for this lower performance is that targets are arranged at varying depths. For example, for virtual ray pointing, Teather and Stuerzlinger [94] showed that varying target depth affects performance. Janzen et al. [42] found that pointing for targets at depths between 110 and 330 cm is affected and identified an effect of the user's distance to the screen. For 3D virtual hand/wand pointing, Barrera and Stuerzlinger [8] found that lateral and depth movements were different when selecting targets displayed via a large stereo display. Batmaz et al. [9] verified that the same effect exists in current AR and VR headsets. Chen et al. [17] studied moving target selection, including object movement in depth. Finally, Lin et al. [57] compared the 3D pointing accuracy between HMDs and stereoscopic widescreen displays but found no difference between them. In all these studies, the authors were not able to identify the cause of their results. One common denominator of these studies is that they used commercial stereo display systems, all of which suffer from the VAC.

2.3 Volumetric Displays

While volumetric displays support stereo vision and focal depth cues, they *typically* do not allow the user to reach into the display volume. Current state-of-the-art 3D volumetric displays allow direct interaction on the surface of the display, such as [90], and thus still suffer from the VAC during interaction when content is displayed away from that surface. While there are volumetric displays that allow users to interact with the virtual content directly, e.g., [46], such systems are still based on a single (slowly actuated) display plane, which again introduces a VAC when content appears in more than a single focal plane.

2.4 Stereo Displays without the VAC

Kouliers et at. [52] recommended that stereo displays should drive accommodation by presenting focal cues correctly to avoid the VAC. Many technologies to address this problem have been presented, including multifocal, varifocal, light-field, holographic, and volumetric displays [102].

Multifocal displays, e.g., [2, 63, 73, 83, 85], present multiple focal image planes simultaneously, allowing the viewer to refocus among the available planes. Varifocal displays, e.g., [1, 49, 78, 79, 99, 100], use eye-tracking to determine at which distance the user is looking in depth, by determining the vergence angle of the eyes. Then, the focal distance of the image is changed to match the accommodation. Previous work showed that multifocal [21, 64, 65] and varifocal [31, 41, 66] displays are capable of removing the vergence-accommodation conflict. Light-field displays, e.g., [36, 54, 71, 96], approximate the correct light field by multiple directional views, each of which offers correct monocular focus cues. Finally, true holographic displays, e.g., [67], reconstruct the full wavefront, which enables them to display all depth cues correctly. Our work uses a multifocal display, as such displays are (relatively) easier to calibrate with high precision. It also allows us to build on a well-known approach to build multifocal displays [2].

3 MULTIFOCAL DISPLAY APPARATUS DESIGN

Most existing VR or AR headsets do not display objects at different focal distances, which would eliminate the VAC. Instead, current VR headsets, such as HTC Vive Pro 2 [35] and Oculus Quest 2 [76], use a fixed distance from the user's eye to where the screen appears, typically at 1 m or beyond. Similarly, the Microsoft Hololens 2 [70], the Varjo XR-3 [97], and other AR headsets also display the virtual content at a fixed focal distance. One exception is the Magic Leap [55], which supports two focal layers at 50 cm and 1.5 m [44], but this is not the correct range for virtual hand interaction, as only one of the focal planes is within arm's reach. LightSpace Technologies is currently developing multifocal VR and AR HMDs [2, 56], but their headsets are currently not available to the public. Based on these challenges, we built a multifocal stereo-display apparatus that directly matches the requirements for our Fitts' law study.

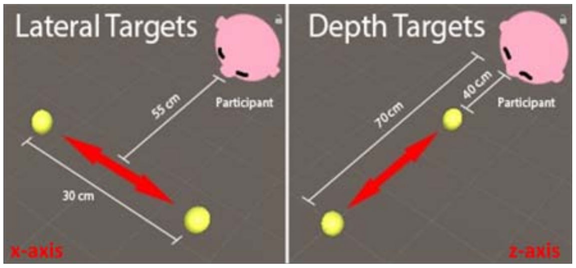


Figure 2: Diagram of the used 3D selection task. A) A pair of lateral targets and B) depth targets.

In our work, we aimed to replicate Barrera and Stuerzlinger's [8] and Batmaz et al. [9] experimental setup and 3D selection task within a multifocal (and singlefocal) display system so we can compare our results directly with this previous work. They used a variant of the ISO 9241-411 selection task [40], where a pair of targets are positioned along a single axis, arranged either laterally or along the line of sight in depth (Figure 2). The experiment involves pointing at targets at three different distances from the user: namely 40, 55, and 70 cm. More precisely, one task involves a lateral motion between two targets, which are 30 cm apart (centered around the view direction), both at 55 cm from the user. This task does not involve a change in target depth (Figure 2 A). The other task asks users to point alternately at two straight-ahead targets positioned at 40 and 70 cm from the user. This pointing movement involves only a change in target depth (Figure 2B). By comparing these two movements, the authors of previous work [8, 9] identified that the presence of target depth changes affects a pointing task negatively. This task design also allows us to investigate our research question: changing the target position between 40-70 cm in depth with a multifocal display enables participants' eyes to correctly verge and accommodate to targets at different (real-world) distances, i.e., to interact with the targets without the VAC. The comparison with a singlefocal display condition, which suffers from the VAC, then allows us to detect the effect of the VAC on pointing in VEs.

3.1 Display Design Considerations

We based the design of our display apparatus on the requirements of our experimental task. In our display, targets need to appear 40, 55, and 70 cm away from the users' eyes and thus the system needs to display images at these distances, too. With such a task design, the front and back mirrors need to be spaced precisely 30 cm apart, which is the same distance as the separation between the left and right targets along the middle mirror, as in Barrera and Stuerzlinger [8]. The mirrors also need to be big enough to display targets 30 cm laterally distant, for three different target widths, plus the shift that is caused by IPD adjustments of the view. The dimensions of the mirrors in front of the participant's eyes were calculated by reflection theory; the participant's vision was limited to the views coming from the mirrors inside the boxes. Beyond the requirements of the experimental task, several other parameters influenced our display design, too.

3.1.1 Image clarity: We followed Patterson guidelines for stereo displays [80] to prevent problems with the image clarity of the display, which might affect depth perception. For this, we focused on the following three aspects: (a) offer a high display resolution to support stereo vision, (b) prevent diplopia by not placing the targets too close to the user, and (c) avoid interocular crosstalk, which occurs when the image for one eye is seen by the other eye [80].

We also used a directional light in the scene to provide a visual cue for the 3D shape of the virtual targets, but made sure that the targets where mostly diffuse. We chose this lighting condition and object material to avoid the potential confound of highlight position and shape serving as a secondary depth cue, i.e., it was not possible to judge target depth based on shading cues (see Figure 7).

3.1.2 Calibration: As the user must move the real wand to the virtually displayed object and the visual and movement distances have to match, we needed to ensure that our display affords the correct perception of distances and target positions. For calibration, we designed and 3D printed a custom 15 x 30 cm calibration platform which enabled us to verify the match between real and virtual content (Figure 3). On the 3D-printed platform, we placed novel, small, custom-designed 2 cm hourglass-like calibration objects at specific locations (both laterally and in depth), which then allowed us to verify the visual alignment between virtual and real objects. For absolute calibration of distances, we also verified with a tape measure that the calibration objects were at the correct distance relative to the viewer. Our new hourglass-like calibration shapes are inspired by a 45°-rotated checkerboard pattern. As they provide a high-contrast visual target regardless of which angle they are seen at, i.e., from either eye of the viewer, they enable us to accurately verify the match between real and virtual for both eyes simultaneously.

3.1.3 Remove image-based depth cues: To avoid confounds, our apparatus prevents participants from using pictorial depth cues to identify the depth of the targets [19]. We achieve this by limiting occlusion and shadows in the environment and use only materials in Unity, but no textures. We also controlled the differences in brightness between the three display panes, to ensure that all virtual content appears at the same brightness, regardless of which mirror it is displayed on.

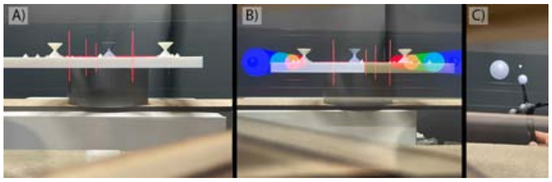


Figure 3: Display calibration setup. A) virtual crosses superimposed on 3D printed calibration device for the right eye. B) Left half of the 3D virtual calibration model superimposed on the real calibration device to verify the match of both scale and pose of the virtual and real objects. Image is taken for the left eye. C) calibrated view in the AR condition, where the top sphere of the wand is superimposed by a virtual sphere. The image also shows the cursor above the wand and a target (spherical shape to left).

3.1.4 Static display: Our display is fixed to a table to avoid the problems associated with using HMDs, such as weight issues that can affect how a display sits on a user's head [13]. We also use a chin and forehead rest to (largely) eliminate participants' head movements, which might affect perception of the stereo display.

3.2 Implementation

We based our multifocal display on Akeley et al.'s [2] design. In their work, the authors used multiple focal planes that helped the user to focus at three, non-linearly distributed depth distances (0.31 m, 0.39 m, and 0.54 m). Through beam splitters, Akeley et al. [2] superimposed images so that each eye sees several images at the same time and can focus on any of the layers. While Akeley et al.'s [2] work showed that their multifocal display can eliminate the VAC for images, their setup did not allow for an investigation of interaction with targets displayed in mid-air. Also, the depth range of their prototype is too small for our purposes. We thus updated Akeley et al.'s [2] design to use the linearly distributed focal distances required for our experimental task. As we are investigating direct interaction with virtual objects in front of the user, the optical path of our system also has to be more complex compared to Akeley et al.'s [2] straight optical system. For this, we use two separate displays positioned to the left and right of the user and reflect their images through a visor to make them appear in front of the user. Finally, to investigate AR interaction, we incorporated see-through capabilities to our display by using beam splitters inside the visor instead of simple mirrors. Figure 4 shows different views of the whole apparatus.

3.2.1 Apparatus: For each eye, our stereo display uses three 10 x 50 cm mirrors placed at a 45° tilt to show the multifocal image to the users. The first two mirrors are beam splitters, while the last is a first-surface mirror that only reflects the view from the display. We also used two 5 x 30 cm beam splitter mirrors in the visor to separate the views for each eye. Based on the two 4K 32" monitors used to display the targets (sitting atop each three-mirror

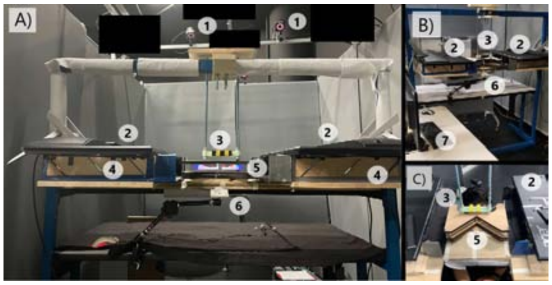


Figure 4: A) Front view of the display, B) side view of the display, and C) top view of the display. The labels show the Optitrack cameras (1), the 4K monitors (2), the head support (3), the display (4), the visor (5), the chin rest (6), and the computer that runs the display (7).

configuration), each mirror effectively shows images with 2680 pixels x 540 pixels resolution. Figure 8 shows the basic design of the system.

Before building the physical system, we created a 3D model of our design (Figure 5A). Using this 3D model, we simulated the FOV and view frustum for each mirror in advance to make sure the apparatus did not obscure the view of the subjects. For instance, we checked if the subjects could see stereo-fusible instances of each target in the same scene, without visual obstructions (Figure 5B). Once we were sure our design was feasible, we laser-cut plywood parts with the correct dimensions. Finally, we painted the inside of the prototype a matte black and used black velvet on the bottom to reduce unwanted reflections and added soft felt around the top edges of the boxes to eliminate stray light from the sides. Figure 5C shows the wood prototype with the mirrors.

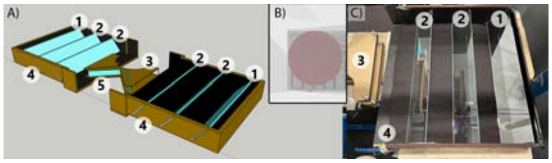


Figure 5: A) 3D model of the display, B) rendering of the participant view of a target (red), and C) final "mirror box" for one eye (the monitor is then placed face-down onto the top of the box). The labels show the first-surface mirror (1), beam splitters (2), the visor (3), the display (4), and the visor's beam splitter (5).

3.2.2 Software. We implemented the system for displaying the targets at the correct location in each mirror in Unity 2019.2.10. We used six virtual cameras located at the user's head position to render the four potential targets. The cameras used off-axis projection to render the image on each mirror with the correct perspective. We rendered the targets either at the top, center, and bottom parts of the screen, which corresponds to images displayed on the front, middle, and back mirror. The optical path for the front targets was 40 cm away from the eye of the user, the two middle targets were at 55 cm, and the back targets at 70 cm. Each target was visible only in the camera corresponding to their mirror position. For the two targets rendered in the middle mirror, we calibrated the FOV of the middle camera so that their images were exactly 30 cm apart on the monitor. Finally, we matched the brightness of each target to cancel the effect of the different optical paths.

As the input device, we used a wand with a soft button on it, like the button on a regular mouse (Figure 6E). IR markers were placed on the wand to enable tracking of the user movements. The wand's dimensions were 21 cm x 13 cm x 5cm, and it weighed 60 grams. As such it was like the wand used in previous work [8, 9]. For 3D tracking of the wand, we used an OptiTrack optical tracking system, consisting of eight 250Hz OptiTrack cameras, calibrated to sub-millimeter precision. We also placed black velvet on the top of the table to minimize reflections that might affect the tracking. The Unity implementation used the OptiTrack data to render the cursor in the correct position. The cursor was a white sphere with a diameter of 0.5cm. We placed this cursor 2 cm above the wand to eliminate any potential negative effects due to diplopia. Figure 3C shows an image of the calibrated virtual wand, superimposed onto the real wand, as well as the cursor with the virtual target used in the study in the AR condition. The average end-to-end system latency was less than 30 ms. To measure the latency, we recorded a real world wand movement in the AR condition with a GoPro Hero3+ at 120Hz and calculated the time elapsed until the virtual wand moved.

We calibrated the field of view of each camera, so when the cursor moves in depth, the rendered cursor "transitions" smoothly between the mirrors. If the cursor was within ± 5 cm of a mirror, well within the "Zone of Clear Vision" [26], the cursor was only visible in a single image. When the cursor was between the depth range displayed by two mirrors, e.g., between 60 and 65 cm from the camera, we show the cursor in both displays (the middle and back ones in the example) and use linear alpha blending between the two layers to display the cursor in between, as in previous work [2, 64]. In other words, if the cursor moves from the front mirror to the middle mirror, we gradually dampen the opacity of the front cursor and increase the opacity of the cursor on the middle mirror until the cursor appears only on the middle mirror. This design might create a VAC problem when the cursor is between mirrors. However, previous work has shown that for rapid aimed targeting movements, which are what Fitts' law predicts, the eyes focus *on the target location*, i.e., the gaze *precedes* the cursor [95], which means that any potential visual artifacts should not substantially affect the pointing performance.

4 USER STUDY

The main goal of our user study was to compare user performance for virtual hand pointing at targets with different depths in VR and AR stereo display conditions, with and without the VAC.

4.1 Hypotheses

H1 - the presence of the VAC negatively affects user 3D selection performance for targets at different visual depth in peri-personal space. We expect that pointing at targets at different visual depths in a multifocal stereo display without the VAC will exhibit reduced hand movement times and increased throughput compared to using a singlefocal stereo display that suffer from the VAC. We based this hypothesis on the problems that the VAC causes for the visual system, including fatigue [6, 33, 34, 39, 77], reduced performance [28, 37, 39, 98], and high cognitive load [20]. More specifically, Gabbard et al. [29] found that continuously shifting visual focus between different distances resulted in significant reductions in (viewing) task performance, reduced comfort, and increased eye fatigue. The importance of H1 is that it quantifies the negative performance effect for virtual hand selection in VR and AR HMDs that use only singlefocal stereo displays to render content.

H2 - 3D selection performance in multifocal displays is better than in singlefocal displays. As multifocal displays present the virtual content at different focal planes, making it like viewing the real world [16, 26], we hypothesize that user performance using multifocal displays exceeds singlefocal displays. The importance of H2 is that commercial multifocal HMDs are slowly becoming available and quantifying their benefits is important.

H3 - the presence of real objects affects 3D selection performance in a multifocal display. While VR headsets are designed to show virtual content, see-through AR HMDs enable users to perceive real-life environmental cues, including lighting and texture. For instance, during a 3D selection task, users could potentially perceive the depth change of a movement better due to the motion parallax of their real hand [19]. Previous research identified no significant difference between VR and AR HMDs for 3D selection [9]. Yet, this difference might be a consequence of the fact that real world content is often at a different focal depth than the virtual content [4, 25, 29]. Thus, we hypothesize that user performance using AR multifocal HMDs is better than using VR multifocal HMDs. The importance of H3 is that when virtual and real content are displayed at the same focal plane, performance is not impacted when the user switches their attention between the real world and virtual information.

4.2 Methodology

We study 3D target selection via a Fitts' law task, which first involves using a wand to move the cursor into the target and then using a button to select the target. For this task to be successful, users need to accurately perceive the position of the target in 3D, i.e., not only laterally, but also in depth, and then move their hand to intersect the virtual target with the cursor above the wand. By following the Fitts' law methodology and using the throughput measure [61], we can compare the performance of different users regardless of their pointing strategies, i.e., we can compare users that are a bit slow but more precise with faster but slightly less careful users.

4.2.1 Participants: We recruited 24 unpaid participants from the local university community (14 male, 13 female, 1 "prefer not to say"). Three participants were left-handed, and the rest right-handed. Seventeen participants were less than 25 years old, six between 26-31 years old, and one over 35 years old, with an average age of 23.6, SD = 4.6. All participants measured normal when tested for stereo viewing capability through a random dot stereo test [43] and used their dominant hand for the task. Our setup permitted participants to wear their glasses, if the needed them to correct their vision.

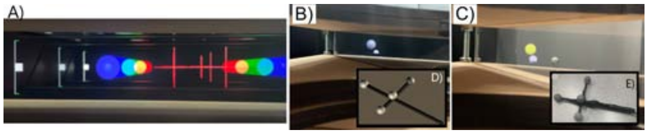


Figure 6: A) The calibrated headset view for the left-eye in the VR condition. The three cylinders on the left- and right-hand sides align and can be perceived as one whole big cylinder each. The three crosses on the left side are also aligned so participant cannot see the two back crosses. B) VR view during the experiment. C) AR view during the experiment. D) Virtual wand. E) Physical wand. Since the camera used to take these picture did not fit completely inside the headpiece, they should only be considered representative views.

4.2.2 System: In this experiment, we used a PC running Windows 10, with Intel i7-2700K CPU, 16GB Ram, and an NVIDIA GeForce GTX 1080 graphics card. See Figure 6e for the wand used in the AR condition and Figure 6d for the 3D model of the wand used in the VR condition. We rendered the 3D model at the same position as the real one to avoid potential confounds between conditions.

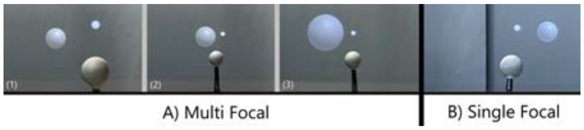


Figure 7: Exemplary views for the different display conditions: A) multifocal display condition, (1) the user focuses on the target in the farthest plane, while the cursor is in the closest plane and thus out of focus. (2) target and cursor in the same focal plane (in this case the farthest plane). (3) focus on the target in the closest plane, with the cursor being in the farthest plane (and thus out of focus). B) single focal display condition. Both cursor and target are always in focus regardless of their depth.

4.2.3 Display: We used the custom multifocal stereo display described in the previous section, see Figure 4. The software was updated to support both tested conditions: single-display and multifocal (Figure 8). In the multifocal condition, each target at a different depth was visible in a different mirror, corresponding to the distance from the user. In the single-display condition, all targets appeared *on the back most mirror*, i.e., the one that is 70 cm away from the user's head. We chose the 70 cm display to match the distance used in previous work [8], so that we can directly compare results. This distance also approximates how current HMDs display virtual content. Figure 7 shows exemplary views for the different display conditions.

Our configuration guarantees that the VAC occurs in the single-display condition, while it does not occur in the multifocal condition.

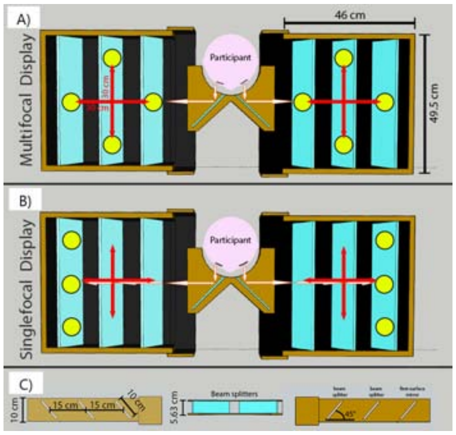


Figure 8: Display conditions: A) multifocal display condition, and B) singlefocal display condition. Red arrows illustrate the (virtual) hand movement path, and white arrows illustrate the visual path. C) side, respectively, front view of the system diagram with dimensions.

To differentiate between the VR and AR conditions, we put black cardboard pieces behind the pair of beam splitters in the visor, so that participants were unable to see their hand in the VR condition. Instead, they saw only a virtual representation of the wand. See Figure 6B. In the AR condition, participants were looking towards a wall covered in grey fabric. We did not show the virtual wand in the AR condition but showed a virtual cursor on top of the real wand. A photograph of the calibrated AR view of the wand and target is shown in Figure 6C.

4.3 Procedure

After participants had read the consent form and signed it, we administered a random dot stereo test [43] to see if they could merge targets. If participants passed the test, i.e., they did not see double images, we measured their interpupillary distance (IPD) using a 3rd party mobile app called "Eye Measure" [38]. To ensure a correct measurement, we verified the IPD at least two times through the app. Then, we seated the participants in front of the apparatus visor on an adjustable chair that does not swivel and adjusted the chair height. We then asked participants to place their chin on the chinrest, which was adjusted to their head size and potentially rotated to avoid blocking their arm movements, and also asked them to place their forehead against the head support. This ensured that their eyes were in (approximately) the same location during the experiment. To eliminate the potential confound of vertical disparity, we also verified that their eye level matched the horizontal visual plane of the apparatus, i.e., the horizontal plane where all targets appeared. Participants used their dominant hand to perform the tasks.

We then adjusted the visual display to take the IPD of the individual participant into account. While we preset the virtual camera parameters based on calibration results from a pilot study, we additionally calibrated the view for each participant to achieve the highest possible visual accuracy. For this, we used a configuration scene that consists of three virtual red crosses for each participant's eye at 40, 55, and 70 cm from the origin, respectively (Figure 6A). The left crosses were rendered half the IPD left of the center and vice versa for the right view. These crosses were aligned with the optical center of each eye at the correct IPD, i.e., when the system was set up correctly, they visually aligned perfectly. If the external IPD measurement was slightly off, we asked participant to tell us what fine-adjustments were needed to align the three crosses for each eye, one eye at a time. This process allowed us to compensate for any potential error in the app measurements.

Once the display was calibrated (Figure 6A), we verified that participants experienced correct stereo vision. For this, we displayed a single object, and the experimenter asked the participants how many objects they saw. During this stage, we also confirmed that participants perceived the cursor to "belong" to the wand in the AR condition. Then, using a test scene, we instructed the participants on the main pointing task and encouraged them to practice until they felt comfortable with it. In this test scene, we used cube-shaped targets to avoid any direct learning effect for the real experiment (where we used spheres). We described the main task in the display design section above (section 3) but clarify further here that it involves reciprocally selecting 3D targets with back-and-forth motions, and participants were instructed to do this "as quickly and as accurately as possible", an instruction that is common in Fitts' law experiment [62]. Participants pushed the button on the wand to select a target. As two targets positioned along the line of sight can occlude each other, we displayed alternating targets at the participant's eye level in all conditions. For a valid selection, the cursor had to be inside the target sphere. Following previous work [8, 9], we also highlighted the targets when the cursor was inside the target to provide visual feedback. If a participant missed a target, the system played an error sound using the PC's speakers. During the experiment and to reduce fatigue, participants were required to take a mandatory break of 60 seconds between different movement types, where we showed a screen titled "REST", and 5 minutes between AR/VR conditions. Additionally, participants were permitted to take a break of 2 minutes between trial rounds, whenever they felt tired. During the study no participants took such an optional break.

After the experiment was complete, the participants filled a survey that asked about the ease of use, perceived speed, and fatigue level for each condition of the experiment. We also asked them about their preferred condition.

5 EXPERIMENTAL DESIGN

We used a $2_{DisplayType} \times 2_{ARVR} \times 2_{MovementDirection}$ within-subjects design. The independent variables were display type (single- vs multifocal display), AR vs VR, and the two movement directions (lateral vs depth). To vary the task, we used three different target sizes. Based on the 30 cm target distance, our task used three distinct IDs between 3.38 and 4.52 bits. The target size and target position (front-back or left-right) changed randomly for each set of 11 pointing trials. Display type, AR/VR, and Movement Direction conditions were counterbalanced across all subjects using a Latin square design. We recorded 11 trials per target ID. Each participant completed 3 repetitions, for a total of 792 trials ($2_{DisplayType} \times$

$2_{ARVR} \times 2_{MovementDirection} \times 3_{ID} \times 11_{Trial} \times 3_{Repetition}$). Across all 24 participants, we recorded a total of 19,008 trials. We removed 12 double-clicks (0.1% of the data), where participants had accidentally clicked the button more than once on a target.

The dependent variables were movement time (ms), error rate (percentage of targets missed), and throughput (bps). We also analyzed the movement paths using target re-entry events, speed, ballistic, and correction times [75]. Both ballistic and correction times were calculated using Nieuwenhuizen's method [75].

5.1 Results

Data was normally distributed for throughput (Skewness (S) = 0.52, Kurtosis (K) = 0.64), speed (S = 0,43, K = 0.28) and correction time (S = 0,6, K = 0.41). The measures of movement time (S = 0.47, K = 0.71), ballistic time (S = 0.47, K = 0.95) and correction distance (S = -0.17, K = -0.59) were normal after log-transformation. The two other dependent variables, error rate and target re-entry, were not normally distributed and we used the aligned rank transform [101] before analysis with ANOVA. We did not remove outliers in movement time and error distance, as this would have removed mostly data for movements in the view direction, i.e., depth movements. The results were analyzed using repeated measures ANOVA with $\alpha = 0.05$. We applied Huynh-Feldt correction when the ϵ was less than 0.75. One-way statistical results are summarized in Table 1. For brevity, we do not list interactions that are not significant.

Table 1: Statistical results of the user study.

	Display Type			AR\VR			Mov. Direction			ID		
Measure	F(1,23)	p	η^2	F(1,23)	p	η^2	F(1,23)	p	η^2	F	p	η^2
Movement Time	**6.688**	**<0.05**	**0.225**	2.692	n.s.	0.105	**14.371**	**<0.001**	**0.385**	**F(1.151,26.480) = 342.913**	**<0.001**	**0.937**
Error rate	**8.510**	**<0.01**	**0.27**	0.541	n.s.	0.23	.453	n.s.	0.021	**F(2,46) = 35.295**	**<0.001**	**0.605**
Throughput	**9.941**	**<0.01**	**0.302**	**4.314**	**<0.05**	**0.158**	**21.932**	**<0.001**	**0.448**	**F(1.382,31.787) = 165.959**	**<0.001**	**0.878**
Target re-entry	0.022	n.s.	0.001	3.362	n.s.	0.128	**6.304**	**<0.05**	**0.215**	**F(2,46) = 438.874**	**<0.001**	**0.95**
Speed	1.045	n.s.	0.043	0.99	n.s.	0.004	**29.528**	**<0.001**	**0.562**	F(1,209,27.806) = 0.559	n.s.	0.024
Ballistic Time	**7.342**	**<0.05**	**0.242**	0.194	n.s.	0.008	**53.286**	**<0.001**	**0.669**	**F(1.133,26.055) = 43.353**	**<0.001**	**0.653**
Correction time	0.539	n.s.	0.023	2.68	n.s.	0.104	2.111	n.s.	0.084	**F(1.050,24.159) = 128.656**	**<0.001**	**0.848**
Correction distance	4.257	n.s.	0.156	2.099	n.s.	0.084	2.377	n.s.	0.094	**F(1.327,30.53) = 401.845**	**<0.001**	**0.946**

5.1.1 Movement Time. Per the statistical results shown in Table 1, time results were significant for display type, movement direction, and ID. According to these results, subjects were faster with the multifocal display than the singlefocal one, and faster with lateral movements compared to movements in the view direction. The detailed results are illustrated in Figure 9(a) and Figure 9(b) for display type and movement direction, respectively. We also observed a significant interaction between movement direction and display type in Table 1, which is shown in Figure 10(a). According to the interaction results, it took longer to execute the task along the z-axis in the singlefocal display condition.

5.1.2 Error rate. The error rate was significantly different for display type and ID, in Table 1. According to these results, subjects had a lower error rate with the multifocal display than with singlefocal. The detailed analysis of this results for display type is shown in Figure 9(c).

5.1.3 Throughput. According to the statistical results shown in Table 1, throughput results were significant for display type AR/VR movement direction and ID. According to the results, participants' throughput was higher with multifocal displays (Figure 9(d)), for AR (Figure 9(e)), and for movements along the view direction (Figure 9(f)). There was also an interaction between display type and axis in Table 1, where we observed higher throughput with multifocal display in depth movements, as shown in Figure 10(b).

5.1.4 Movement Trajectory results.

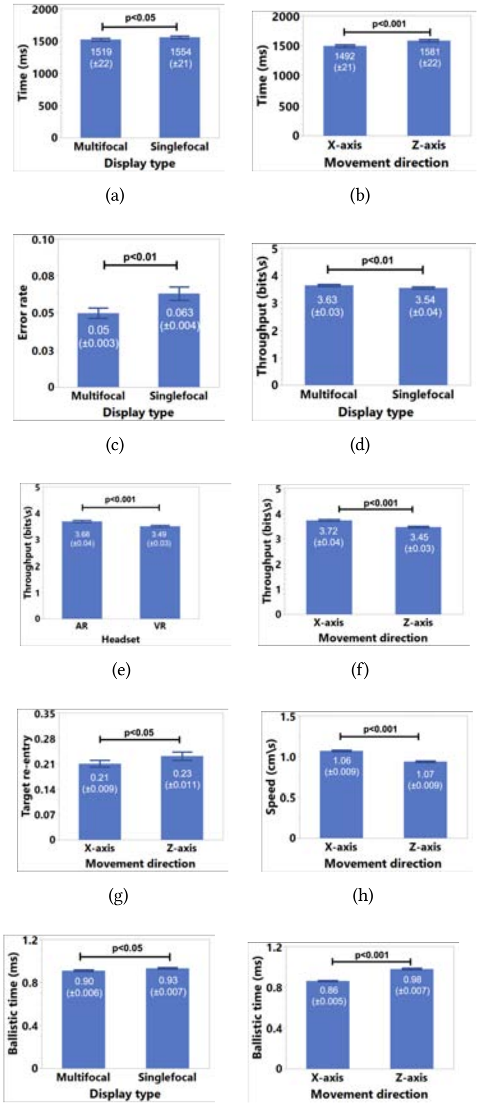
- *Re-entry*: Target re-entry was significant for movement direction, see Figure 9(g), and ID. According to these results, we observed fewer target re-entries for movements in the x-axis. Also, there was an interaction effect between display type and axis, where participants exhibited more target re-entries in z-axis movements compared to the x-axis for the singlefocal display condition.
- *Speed*: Speed results were significant for movement direction, shown in Figure 9(h), where participants were faster in the x-axis.
- *Ballistic time*: Ballistic time was significant for display type and movement direction, Figure 9(i), Figure 9(j), respectively. We also found a significant difference for ID. According to these results, we observed a higher ballistic time in the singlefocal display condition and for z-axis movements.
- *Correction time*: Correction time results were significant for ID.
- *Correction distance*: Correction distance results were significant for ID. There was also an interaction effect between display type and headset, where we observed a higher correction distance with the VR multifocal display compare to the AR one. See Figure 11(a). There was also an interaction effect between display type, headset and axis, where AR multifocal displays had a smaller correction distance than the VR one for depth movements.

5.1.5 Questionnaire results. At the end of the experiment, we asked subjects to fill a questionnaire to share their opinions. We asked subjects if they recognized the difference between the multifocal and singlefocal display conditions and all stated that they perceived them to be different. While nine subjects preferred the multifocal display, nine others preferred neither, and the remaining six preferred singlefocal.

We also asked subjects which display condition they preferred, on a 7-point Likert scale (1-strongly prefer VR and 7-strongly prefer AR). Only one subject preferred the AR condition, while most preferred VR (Mean = 3.69, SD = 0.87). Yet, in the comments one participants stated, *"I preferred AR because I was able to see my hand and it felt more like an extension rather than a different embodied interaction (like in VR)."* and another one said *"It was much easier to [perceive] distance during AR."*

When we asked subjects about their arm fatigue using a 7-point Likert scale (1-feel rested, 4-feel normal and 7-feel tired), the average response was normal (Mean = 4.34, SD = 1.53). However, in the comments two participants mentioned that they had to stretch their arms to reach the farthest targets, i.e., those at 70 cm.

Finally, we asked participants in the questionnaire if they experienced any motion sickness symptoms, such as dizziness, eyestrain, nausea, vomiting, sweating, and burping. None perceived nor exhibited such symptoms.



(a) (b)

(c) (d)

(e) (f)

(g) (h)

(i) (j)

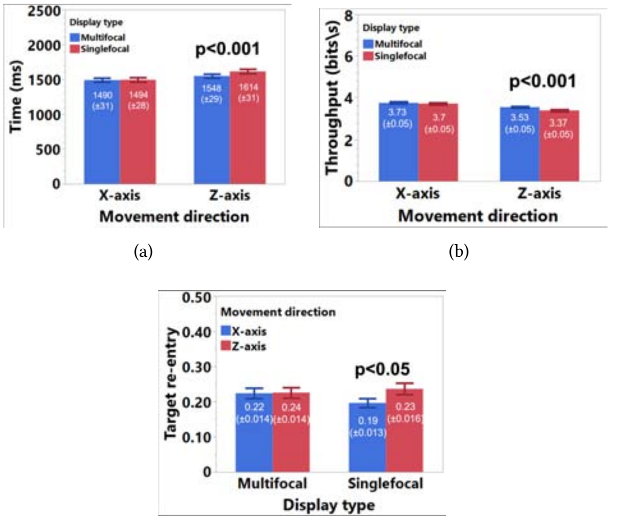
Figure 9: (a) Detailed time results for display type, (b) movement direction, error rate results for (c) display type, throughput results for (d) display type, (e) headset and (f) movement direction, target re-entry results for (g) movement direction, speed results for (h) movement direction, ballistic time results for (i) display time and (j) movement direction. Error bars represent the standard error of the mean.

6 DISCUSSION

We designed a multifocal stereo display that allows the user to focus on targets at different distances. The main purpose of this device was to investigate the theory that the vergence-accommodation conflict (VAC) affects user interaction performance in stereo display systems. Our results confirm that the VAC has a significant detrimental impact on user performance for pointing motions to targets at different visual depths. See Figure 10.

6.1 The Impact of the Vergence and Accommodation Conflict on User Interaction

Our first and main contribution identifies the impact of the VAC on the user interaction for pointing motions to targets at different visual depth. When comparing single- and multifocal display performance, our results show that participants were slower and their throughput significantly decreased for the singlefocal conditions for movements along the depth axis (30 cm depth change, from 40 cm to 70 cm). Yet, we did not observe a corresponding difference for lateral movements (both targets 55 cm away from the user, no change in depth). See Figure 10(a) for time and Figure 10(b) for throughput.



(a) (b)

(c)

Figure 10: Display and movement direction interaction results for (a) time and (b) throughput. (c) Target re-entry vs. display type. Error bars represent the standard error of the mean.

Looking more in-depth at our data, we found that the difference in movement time between selection with a multifocal versus a singlefocal display is on average 35 ms. This is substantially less than the average latency for vergence eye movements (160-200 ms) and lens accommodation (300-400 ms) [6]. We thus expect that this small difference is due to the way our participants executed the pointing task. Based on the repetitive nature of the pointing movement, we expect that our participants moved their arm even before having a correct depth perception of the target. Then, and only after the first ballistic movement, they corrected their movement to reach the target. Gabbard et al.'s [29] results support our hypothesis. They found that when people read text after a change in focal depth, they make more errors at the beginning of the sentence than at the end, meaning that their participants started reading the sentence before they completely focused on the text. In our work, the singlefocal condition was slower, which we believe to be caused by the VAC, which made it harder to perceive the correct target depth. The target re-entry data supports this argument, as there is a significant difference between depth and lateral movements for singlefocal displays, but not for multifocal displays. Given that participants exhibited more target re-entries for singlefocal displays depth movements this means that participants had more trouble finding the correct target depth (Figure 10(c)).

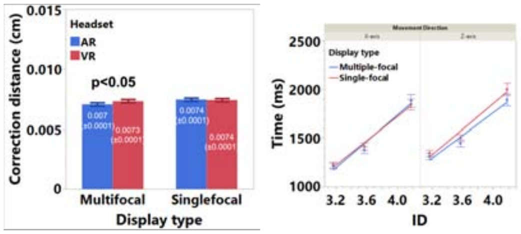
Even though the results here identify a seemingly small difference, i.e., 35 ms between singlefocal and multifocal display conditions, this amount might still play a crucial role in VR/AR systems, as interaction is sensitive to end-to-end latency [82, 93]. Research in surgical teleoperation has shown that end-to-end latencies below 100 ms are needed to support efficient interaction [72], and we expect similar results to hold for VR and AR systems. As the VAC in singlefocal stereo displays effectively delays interaction, a value of 35 ms is not negligible, relative to the 100 ms identified in that work. Further, and as accommodation to varying, less predictable target depths can be expected to take longer, we expect the penalty due to the VAC to be larger in practice, specifically for non-repetitive tasks that are more representative of real-world interaction in VR/AR applications. In such situations and because accommodation can take up to 400 ms in the worst case [6], one could then expect the negative effect due to the VAC to be larger and to potentially even reach the 100 ms range. We plan to run a replication study with targets at random depths to test this hypothesis in the future.

Overall, our results identify that the presence of the VAC affects movements with a change in depth negatively. There, users need to constantly deal with changing (and conflicting) vergence and accommodation cues to correctly hit the target. Yet, the absence of a VAC seems to globally decrease user performance, as the performance for movements with no change in depth was similar between display conditions. In this task condition, participants moved their gaze along the same depth during the task, with a constant VAC, regardless of the display condition. Based on this, we can confirm H1, as our results show that the VAC negatively affects user 3D selection performance in peri-personal space for movements with a change of depth.

6.2 Study of Interaction with Multifocal Displays

6.2.1 Single- vs. multifocal displays: The second contribution of this work identifies the impact of multifocal displays on user interaction. In the comparison between single- and multifocal displays, we found significant differences for time, error rate, throughput, and ballistic time. According to these results, participants were slower, made more errors, and their throughput performance significantly decreased with singlefocal displays. Based on this, we believe we can confirm H2.

6.2.2 Multifocal AR and VR displays: When looking at the difference between AR and VR multifocal displays, we found a difference in terms of throughput. In detailed analysis, we also identified that VR multifocal displays exhibit a larger correction distance than AR multifocal displays (Figure 11(a)). This difference is not present for singlefocal AR/VR displays. Given that the multifocal AR display condition exhibited higher throughput and less correction distances, i.e., participants were able to judge the depth of the target better, we believe that the additional cues afforded by the AR condition together with the ability of displaying virtual elements at the same depth as physical ones allow users to better identify the target depth. This result seems to contradict previous results for pointing in VR and AR HMDs [10], but we point out that this previous work used only singlefocal displays. We see our results as support for H3.



(a) (b)

Figure 11: (a) Display type and VR/AR interaction results for correction distance. (b) Fitts' law model for single- and multifocal displays. Error bars represent the standard error of the mean.

6.2.3 Fitts' Law Analysis for Multifocal Displays: We also confirmed that movement times follow Fitts' Law, as there is a linear relationship between IDs and Time for multifocal (R^2 = 0.96) and singlefocal displays (R^2 = 0.98). The throughput results also corroborate this as, for example, the throughput values are 3.8 for ID 3.1, 3.6 for ID 3.5 and 3.07 for ID 4.18 for the singlefocal condition, all of which are similar to previous work [8, 9].

6.3 Cross-study Comparison

For the design of our multifocal display apparatus, we focused on building a device that would enable us to directly compare our results with previous work [8, 9]. Another motivation was to be able to directly verify that our new display works correctly. Matching previous work, we found a significant difference between IDs, where larger IDs have worse performance than smaller IDs for throughput, movement time, error rate, target re-entry, ballistic time, correction time and correction distance. However, we also identified that our participants exhibited lower average pointing times and throughput for all condition (See Table 2) as observed in previous work. Further, participants were slower and their throughput performance decreased for depth movements.

Table 2: Pointing performance comparison with previous work.

Singlefocal display	Lateral target movement time	Depth target movement time	Lateral target throughput	Depth target throughput
Our experiment	1494 ms (sd = 491)	1614 ms (sd = 530)	3.7 bits/s (sd = 0.92)	3.4 bits/s (sd = 0.81)
Barrera & Stuerzlinger [8]	1104 ms (sd = 372)	1450 ms (sd = 715)	4.6 bits/s (sd = 1.8)	3.6 bits/s (sd = 1.48)
Batmaz el al. [9]	920 ms (sd = 328)	1090 ms (sd = 333)	4.6 bits/s (sd = 1.4)/s	3.6 bits/s (sd = 1.03)

Here, we discuss potential variables that could have caused this overall decrease in performance. We also describe how these possible limitations did not impact our participants' ability to comfortably perceive stereo views at the correct distances with our custom multifocal display, even if they impacted the overall performance of the experiment.

- Bio-mechanical limits of movement: When the subjects pointed to the targets, user performance could be significantly affected by biomechanical constrains. Such limitation can be observed in the upper extremity, such as the plane of shoulder exertion, which affects the used muscles [3, 68, 91]. Another issue is that hand movements that go over the mid-body line are more complex (and thus slower) than those that do not [74, 86]. In our study, the subjects had to position themselves in a 45 cm wide space. Even though we tried to reduce postural discomfort in the experiment and subjects were able to reach all the targets, the relatively cramped space could have affected user performance. Further, some participants took longer to reach targets at 70 cm. While this may have increased the absolute pointing time for 70 cm targets, we still observed a significant difference between the single- and multifocal display conditions for depth motions, which means that any potential issues affected both conditions equally.
- Task fatigue: Since 3D mid-air pointing is not a daily task for participants, some subjects could have experienced increased fatigue during the experiment. To combat this, we forced subjects to take regular breaks. Still, they did not report significant arm fatigue at the end of the study. Thus, we believe that arm fatigue did not affect the performance of the participants.
- Visual fatigue: Another potential explanation of our results is visual fatigue (potentially) caused by the need to constantly transition between targets that were subject to varying degrees of the VAC [6]. Still, none of our participants complained about eyestrain, so we believe this was not a notable concern. Also, during the change between AR and VR conditions, participants were encouraged to get up and rest for a couple of minutes.

  Further, we carefully counterbalanced single- and multifocal and AR/VR conditions during the experiment. Thus, while it is potentially possible that switching between multifocal and singlefocal displays for each task could increase visual fatigue, we believe that we eliminated this potential confound through the counterbalancing scheme.
- Display Problems: As we always displayed targets at the three main focal distances, targets were displayed perfectly, i.e., without blending or other potential deficiencies. Still, we used alpha blending when the *cursor* was between focal planes in the multifocal condition. Thus, we cannot rule out that minor visual artifacts for the cursor might have been visible. However, we do not see this as a notable concern, as previous work has shown that for targeting movements the eyes focus on the target location, i.e., *precede* the cursor [95]. This means that any visual artifacts should not have affected the pointing performance in a substantial way.
- Motion Sickness: Another potential issue in stereo display systems is motion sickness, which is a characteristic of undesirable display systems [6]. However, when looking at the questionnaire results, we found that participants did not experience motion sickness symptoms when using our system, which means that this issue is unlikely to have played a role.

6.4 Implications for the Field of Virtual & Augmented Reality

Our outcomes point to several implications for VR and AR display systems and/or applications.

- Building on previous work that identified this issue [8, 9], our results provide strong evidence that application designers should avoid displaying targets at varying depths to avoid a drop in user interaction performance with current singlefocal stereo display systems, including display walls and VR/AR HMDs.
- For multifocal stereo displays, and in addition to a far focal plane, system designers should consider including (at least) two focal planes within reach of the user. Considering the zone of clear vision [26], a set of focal planes at 40, 65, and 200 cm might be sensible, but this is–by far–not the only option. Thus, additional research will be needed to verify what constitutes an appropriate set of focal planes that works well for both viewing of as well as interaction with virtual content.
- We expect our outcomes to apply to all display systems that have only a single focal plane within arm's reach. Yet, our results also show the importance of evaluating user interaction with and without a change in depth for different types of stereo displays, such as varifocal, light-field, and holographic displays. Our identification of the benefits of display systems that do not suffer from the VAC also motivate further work on reach-in volumetric displays.
- Previous studies, such as [8, 15, 18] proposed different models to predict human pointing in (visual) depth. The clear difference in lateral and depth movements observed in the work presented here highlights the importance of additional research on such models.

7 CONCLUSION

This paper presented a comparison of multifocal and singlefocal as well as AR vs VR displays for 3D virtual hand pointing. We identified that a (sufficiently large) change in target depth negatively affects virtual hand interaction in peri-personal space with singlefocal displays. In contrast, multifocal displays do not suffer from this issue. Given this outcome, we believe that the VAC present in singlefocal stereo display systems fundamentally impacts user performance negatively. Overall, our results match Barrera and Stuerzlinger's [8] results for 3D TVs, and Batmaz et al.'s [9] results for current VR and AR headsets. We also identify a statistically better performance for AR displays.

Our work presents the first 3D selection study with a multifocal display and analyzes user performance. Moreover, our work has implications for the design of all kinds of stereo displays, especially

HMDs. For the common case of VR/AR systems that afford interaction within arm's reach, e.g., by tracking the hand with a Leap Motion or the sensors of the HoloLens 2 or the Varjo XR-3, our results points to a need to support multiple focal layers to enable more efficient 3D selection.

In the future, we are planning to investigate additional target configurations to characterize the impact of the VAC even more accurately.

ACKNOWLEDGMENTS

Many thanks for Nicholas Liang for designing the first virtual prototype of the multifocal display.

REFERENCES

[1] Kaan Akşit, Ward Lopes, Jonghyun Kim, Peter Shirley, and David Luebke. 2017. Near-eye Varifocal Augmented Reality Display Using See-through Screens. *ACM Trans. Graph.* 36, 6, Article 189 (Nov. 2017), 13 pages. https://doi.org/10.1145/3130800.3130892

[2] Kurt Akeley, Simon J. Watt, Ahna Reza Girshick, and Martin S. Banks. 2004. A Stereo Display Prototype with Multiple Focal Distances. In *ACM SIGGRAPH 2004 Papers* (Los Angeles, California) *(SIGGRAPH '04)*. Association for Computing Machinery, New York, NY, USA, 804–813. https://doi.org/10.1145/1186562.1015804

[3] Nicholas T. Antony and Peter J. Keir. 2010. Effects of posture, movement and hand load on shoulder muscle activity. *Journal of Electromyography and Kinesiology* 20, 2 (apr 2010), 191–198. https://doi.org/10.1016/j.jelekin.2009.04.010

[4] Mohammed Safayet Arefin, Nate Phillips, Alexander Plopski, Joseph L. Gabbard, and J. Edward Swan. 2020. Impact of AR Display Context Switching and Focal Distance Switching on Human Performance: Replication on an AR Haploscope. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 571–572. https://doi.org/10.1109/VRW50115.2020.00137

[5] Claudia Armbrüster, Marc Wolter, Torsten Wolfgang Kuhlen, W. Spijkers, and Bruno Fimm. 2008. Depth Perception in Virtual Reality: Distance Estimations in Peri- and Extrapersonal Space. *CyberPsychology & Behavior* 11, 1 (feb 2008), 9–15. https://doi.org/10.1089/cpb.2007.9935

[6] Takehiko Bando, Atsuhiko Iijima, and Sumio Yano. 2012. Visual fatigue caused by stereoscopic images and the search for the requirement to prevent them: A review. *Displays* 33, 2 (2012), 76–83. https://doi.org/10.1016/j.displa.2011.09.001

[7] Martin S. Banks, Joohwan Kim, and Takashi Shibata. 2013. Insight into vergence/accommodation mismatch. *Head- and Helmet-Mounted Displays XVIII: Design and Applications* 8735 (2013), 873509. https://doi.org/10.1117/12.2019866

[8] Mayra Donaji Barrera Machuca and Wolfgang Stuerzlinger. 2019. The Effect of Stereo Display Deficiencies on Virtual Hand Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM Press, New York, NY, 14. https://doi.org/10.1145/3290605.3300437

[9] A. U. Batmaz, M. D. B. Machuca, D. M. Pham, and W. Stuerzlinger. 2019. Do Head-Mounted Display Stereo Deficiencies Affect 3D Pointing Tasks in AR and VR?. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 585–592. https://doi.org/10.1109/VR.2019.8797975

[10] Anil Ufuk Batmaz, Aunnoy K. Muttasim, Morteza Malekmakan, Elham Sadr, and Wolfgang Stuerzlinger. 2020. Touch the Wall: Comparison of Virtual and Augmented Reality with Conventional 2D Screen Eye-Hand Coordination Training Systems. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR'20)*. 11.

[11] James M. Brown and Naomi Weisstein. 1988. A spatial frequency effect on perceived depth. *Perception & Psychophysics* 44, 2 (mar 1988), 157–166. https://doi.org/10.3758/BF03208708

[12] Gerd Bruder, Frank Steinicke, and Wolfgang Stuerzlinger. 2013. 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)*. IEEE, 115–118. https://doi.org/10.1109/3DUI.2013.6550207

[13] Lauren E. Buck, Mary K. Young, and Bobby Bodenheimer. 2018. A comparison of distance estimation in HMD-based virtual environments with different HMD-based conditions. *ACM Transactions on Applied Perception* 15, 3 (2018). https://doi.org/10.1145/3196885

[14] Canon. 2021. MREAL S1.

[15] Yeonjoo Cha and Rohae Myung. 2013. Extended Fitts' law for 3D pointing tasks using 3D target arrangements. *International Journal of Industrial Ergonomics* 43, 4 (2013), 350 – 355. https://doi.org/10.1016/j.ergon.2013.05.005

[16] Jen-Hao Rick Chang, Anat Levin, B. V. K. Vijaya Kumar, and Aswin C. Sankaranarayanan. 2020. Towards Occlusion-Aware Multifocal Displays. *ACM Trans. Graph.* 39, 4, Article 68 (July 2020), 15 pages. https://doi.org/10.1145/3386569.3392424

[17] Yuan Chen, Junwei Sun, Qiang Xu, Edward Lank, Pourang Irani, and Wei Li. 2021. Empirical Evaluation of Moving Target Selection in Virtual Reality using Egocentric Metaphors. In *IFIP Conference on Human-Computer Interaction*. Springer.

[18] Logan D Clark, Aakash B Bhagat, and Sara L Riggs. 2020. Extending Fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies* 139 (2020), 102413.

[19] James E. Cutting and Peter M. Vishton. 1995. 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*. Academic Press, San Diego, California, USA, 69–117.

[20] François Daniel and Zoï Kapoula. 2019. Induced vergence-accommodation conflict reduces cognitive performance in the Stroop test. *Scientific Reports* 9, 1 (2019), 1–13. https://doi.org/10.1038/s41598-018-37778-y

[21] David Dunn, Praneeth Chakravarthula, Qian Dong, and Henry Fuchs. 2018. Mitigating vergence-accommodation conflict for near-eye displays via deformable beamsplitters. In *Digital Optics for Immersive Displays*, Bernard C. Kress, Wolfgang Osten, and Hagen Stolle (Eds.), Vol. 10676. International Society for Optics and Photonics, SPIE, 196 – 208. https://doi.org/10.1117/12.2314664

[22] David Dunn, Cary Tippets, Kent Torell, Petr Kellnhofer, Kaan Akşit, Piotr Didyk, Karol Myszkowski, David Luebke, and Henry Fuchs. 2017. Wide Field Of View Varifocal Near-Eye Display Using See-Through Deformable Membrane Mirrors. *IEEE Transactions on Visualization and Computer Graphics* 23, 4 (2017), 1322–1331. https://doi.org/10.1109/TVCG.2017.2657058

[23] Frank H. Durgin, Dennis R. Proffitt, Thomas J. Olson, and Karen S. Reinke. 1995. Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance* 21, 3 (1995), 679–699. https://doi.org/10.1037/0096-1523.21.3.679

[24] G. N. Dutton, A. Saaed, B. Fahad, R. Fraser, G. McDaid, J. McDade, A. Mackintosh, T. Rane, and K. Spowart. 2004. 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* 18, 1 (jan 2004), 27–34. https://doi.org/10.1038/sj.eye.6700541

[25] Anna Eiberger, Per Ola Kristensson, Susanne Mayr, Matthias Kranz, and Jens Grubert. 2019. Effects of Depth Layer Switching between an Optical See-Through Head-Mounted Display and a Body-Proximate Display. In *Symposium on Spatial User Interaction* (New Orleans, LA, USA) *(SUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 15, 9 pages. https://doi.org/10.1145/3357251.3357588

[26] Ian M Erkelens and Kevin J MacKenzie. 2020. 19-2: Vergence-Accommodation Conflicts in Augmented Reality: Impacts on Perceived Image Quality. *SID Symposium Digest of Technical Papers* 51, 1 (2020), 265–268. https://doi.org/10.1002/sdtp.13855 arXiv:https://sid.onlinelibrary.wiley.com/doi/pdf/10.1002/sdtp.13855

[27] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.

[28] Tetsuya Fukushima, Masahito Torii, Kazuhiko Ukai, James S. Wolffsohn, and Bernard Gilmartin. 2009. The relationship between CA/C ratio and individual differences in dynamic accommodative responses while viewing stereoscopic images. *Journal of Vision* 9, 13 (2009), 1–13. https://doi.org/10.1167/9.13.1

[29] Joseph L. Gabbard, Divya Gupta Mehra, and J. Edward Swan. 2019. Effects of AR Display Context Switching and Focal Distance Switching on Human Performance. *IEEE Transactions on Visualization and Computer Graphics* 25, 6 (2019), 2228–2241. https://doi.org/10.1109/TVCG.2018.2832633

[30] Evan D. Graham and Christine L. MacKenzie. 1996. Physical versus virtual pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '96)*. ACM Press, New York, New York, USA, 292–299. https://doi.org/10.1145/238386.238532

[31] Ali Hasnain, Pierre-Yves Laffont, Shukri Bin Abdul Jalil, Kutluhan Buyukburc, Pierre-Yves Guillemet, Samuel Wirajaya, Liqiang Khoo, Teng Deng, and Jean-Charles Bazin. 2019. Piezo-actuated varifocal head-mounted displays for virtual and augmented reality. In *Advances in Display Technologies IX*, Jiun-Haw Lee, Qiong-Hua Wang, and Tae-Hoon Yoon (Eds.), Vol. 10942. International Society for Optics and Photonics, SPIE, 40 – 50. https://doi.org/10.1117/12.2509143

[32] Robert F. Hess, Long To, Jiawei Zhou, Guangyu Wang, and Jeremy R. Cooperstock. 2015. Stereo Vision: The Haves and Have-Nots. *i-Perception* 6, 3 (jun 2015), 1–5. https://doi.org/10.1177/2041669515593028

[33] David M. Hoffman, Ahna R. Girshick, Kurt Akeley, and Martin S. Banks. 2008. Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision* 8, 3 (mar 2008), 33.1–30. https://doi.org/10.1167/8.3.33

[34] Hyungki Hong and Seok Hyon Kang. 2015. Measurement of the lens accommodation in viewing stereoscopic displays. *Journal of the Society for Information Display* 23, 1 (jan 2015), 19–26. https://doi.org/10.1002/jsid.303

[35] HTC. 2021. VIVE PRO 2. https://www.vive.com/ca/

[36] Xinpeng Huang, Ping An, Fengyin Cao, Deyang Liu, and Qiang Wu. 2019. Light-field compression using a pair of steps and depth estimation. *Opt. Express* 27, 3 (Feb 2019), 3557–3573. https://doi.org/10.1364/OE.27.003557

[37] Anke Huckauf, Mario H. Urbina, Irina Böckelmann, Lutz Schega, Rüdiger Mecke, Jens Grubert, Fabian Doil, and Johannes Tümler. 2010. Perceptual issues in optical-see-through displays. *Proceedings - APGV 2010: Symposium on Applied Perception in Graphics and Visualization* 1, 212 (2010), 41–48. https://doi.org/10.1145/1836248.1836255

[38] Bonlook Inc. 2021. EyeMeasure. https://apps.apple.com/us/app/eyemeasure/id1417435049

[39] J. Iskander, Mohammed Hossny, and S. Nahavandi. 2019. Using biomechanics to investigate the effect of VR on eye vergence system. *Applied Ergonomics* 81, August 2018 (2019), 102883. https://doi.org/10.1016/j.apergo.2019.102883

[40] ISO. 2015. ISO 9241-400:2012 Ergonomics of human-system interaction - Part 411: Evaluation methods for the design of physical input devices.

[41] Afsoon Jamali, Comrun Yousefzadeh, Colin McGinty, Douglas Bryant, and Philip Bos. 2018. LC lens systems to solve accommodation/convergence conflict in three-dimensional and virtual reality displays. *Optical Engineering* 57, 10 (2018), 1. https://doi.org/10.1117/1.oe.57.10.105101

[42] Izabelle Janzen, Vasanth K. Rajendran, and Kellogg S. Booth. 2016. Modeling the Impact of Depth on Pointing Performance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 188–199. https://doi.org/10.1145/2858036.2858244

[43] Bela Julesz. 1964. Binocular depth perception without familiarity cues: Random-dot stereo images with controlled spatial and temporal properties clarify problems in stereopsis. *Science* 145, 3630 (1964), 356–362.

[44] KarlG. 2018. Magic Leap Review Part 1 – The Terrible View Through Diffraction Gratings. https://www.kguttag.com/2018/09/26/magic-leap-review-part-1-the-terrible-view-through-diffraction-gratings/

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

[46] Hanyuool Kim, Issei Takahashi, Hiroki Yamamoto, Satoshi Maekawa, and Takeshi Naemura. 2014. Mario: Mid-air augmented reality interaction with objects. *Entertainment Computing* 5, 4 (2014), 233–241.

[47] Jonghyun Kim, Youngmo Jeong, Michael Stengel, Kaan Akşit, Rachel Albert, Ben Boudaoud, Trey Greer, Joohwan Kim, Ward Lopes, Zander Majercik, Peter Shirley, Josef Spjut, Morgan McGuire, and David Luebke. 2019. Foveated AR: Dynamically-Foveated Augmented Reality Display. *ACM Trans. Graph.* 38, 4, Article 99 (July 2019), 15 pages. https://doi.org/10.1145/3306346.3322987

[48] Joohwan Kim, David Kane, and Martin S. Banks. 2014. The rate of change of vergence-accommodation conflict affects visual discomfort. *Vision Research* 105 (2014), 159–165. https://doi.org/10.1016/j.visres.2014.10.021 arXiv:NIHMS150003

[49] Robert Konrad, Nitish Padmanaban, Keenan Molner, Emily A. Cooper, and Gordon Wetzstein. 2017. Accommodation-Invariant Computational near-Eye Displays. *ACM Trans. Graph.* 36, 4, Article 88 (July 2017), 12 pages. https://doi.org/10.1145/3072959.3073594

[50] Regis Kopper, Doug A. Bowman, Mara G. Silva, and Ryan P. McMahan. 2010. A human motor behavior model for distal pointing tasks. *International Journal of Human-Computer Studies* 68, 10 (oct 2010), 603–615. https://doi.org/10.1016/j.ijhcs.2010.05.001

[51] G. A. Koulieris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt. 2019. Near-Eye Display and Tracking Technologies for Virtual and Augmented Reality. *Computer Graphics Forum* 38, 2 (2019), 493–519. https://doi.org/10.1111/cgf.13654 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13654

[52] George-Alex Koulieris, Bee Bui, Martin S. Banks, and George Drettakis. 2017. Accommodation and Comfort in Head-Mounted Displays. *ACM Trans. Graph.* 36, 4, Article 87 (July 2017), 11 pages. https://doi.org/10.1145/3072959.3073622

[53] Gregory Kramida. 2016. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics* 22, 7 (2016), 1912–1931. https://doi.org/10.1109/TVCG.2015.2473855 arXiv:arXiv:1011.1669v3

[54] Douglas Lanman and David Luebke. 2013. Near-eye Light Field Displays. *ACM Trans. Graph.* 32, 6, Article 220 (Nov. 2013), 10 pages. https://doi.org/10.1145/2508363.2508366

[55] Magic Leap. 2019. Magic Leap.

[56] LightSpace. 2021. mpS3D. https://www.lightspace3d.com/index.php/Products/mps3d-vr-headset/

[57] Chiuhsiang J. Lin, Betsha T. Abreham, and Bereket H. Woldegiorgis. 2019. Effects of displays on a direct reaching task: A comparative study of head mounted display and stereoscopic widescreen display. *International Journal of Industrial Ergonomics* 72, December 2018 (2019), 372–379. https://doi.org/10.1016/j.ergon.2019.06.013

[58] Chiuhsiang Joe Lin and Bereket Haile Woldegiorgis. 2015. Interaction and visual performance in stereoscopic displays: A review. *Journal of the Society for Information Display* 23, 7 (jul 2015), 319–332. https://doi.org/10.1002/jsid.378

[59] Chiuhsiang Joe Lin and Bereket H. Woldegiorgis. 2017. Egocentric distance perception and performance of direct pointing in stereoscopic displays. *Applied Ergonomics* 64 (oct 2017), 66–74. https://doi.org/10.1016/j.apergo.2017.05.007

[60] Chiuhsiang Joe Lin, Bereket H. Woldegiorgis, and Dino Caesaron. 2014. Distance estimation of near-field visual objects in stereoscopic displays. *Journal of the Society for Information Display* 22, 7 (jul 2014), 370–379. https://doi.org/10.1002/jsid.269

[61] I Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction* 7, 1 (1992), 91–139.

[62] I. Scott MacKenzie and William Buxton. 1992. Extending Fitts' Law to Two-Dimensional Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) *(CHI '92)*. Association for Computing Machinery, New York, NY, USA, 219–226. https://doi.org/10.1145/142750.142794

[63] Kevin J. MacKenzie, Ruth A. Dickson, and Simon J. Watt. 2012. Vergence and accommodation to multiple-image-plane stereoscopic displays: "real world" responses with practical image-plane separations? *Journal of Electronic Imaging* 21, 1 (2012), 1 – 9. https://doi.org/10.1117/1.JEI.21.1.011002

[64] Kevin J. MacKenzie, David M. Hoffman, and Simon J. Watt. 2010. Accommodation to multiple-focal-plane displays: Implications for improving stereoscopic displays and for accommodation control. *Journal of Vision* 10, 8 (07 2010), 22–22. https://doi.org/10.1167/10.8.22 arXiv:https://jov.arvojournals.org/arvo/content_public/journal/jov/933477/jov-10-8-22.pdf

[65] Kevin J. MacKenzie and Simon J. Watt. 2010. Eliminating accommodation-convergence conflicts in stereoscopic displays: Can multiple-focal-plane displays elicit continuous and consistent vergence and accommodation responses?. In *Stereoscopic Displays and Applications XXI*, Andrew J. Woods, Nicolas S. Holliman, and Neil A. Dodgson (Eds.), Vol. 7524. International Society for Optics and Photonics, SPIE, 414 – 423. https://doi.org/10.1117/12.840283

[66] Guido Maiello, Manuela Chessa, Fabio Solari, and Peter J. Bex. 2015. The (In)Effectiveness of Simulated Blur for Depth Perception in Naturalistic Images. *PLOS ONE* 10, 10 (10 2015), 1–15. https://doi.org/10.1371/journal.pone.0140230

[67] Andrew Maimone, Andreas Georgiou, and Joel S. Kollin. 2017. Holographic near-eye displays for virtual and augmented reality. *ACM Transactions on Graphics* 36, 4 (2017), 1–16. https://doi.org/10.1145/3072959.3073624

[68] Alison McDonald, Bryan R. Picco, Alicia L. Belbeck, Amy Y. Chow, and Clark R. Dickerson. 2012. Spatial dependency of shoulder muscle demands in horizontal pushing and pulling. *Applied Ergonomics* 43, 6 (nov 2012), 971–978. https://doi.org/10.1016/j.apergo.2012.01.005

[69] John P. McIntire, Paul R. Havig, and Eric E. Geiselman. 2014. Stereoscopic 3D displays and human performance: A comprehensive review. *Displays* 35, 1 (jan 2014), 18–26. https://doi.org/10.1016/j.displa.2013.10.004

[70] MICROSOFT. 2020. HoloLens 2. https://www.microsoft.com/en-us/hololens

[71] Haruki Mizushina, Junya Nakamura, Yasuhiro Takaki, and Hiroshi Ando. 2016. Super multi-view 3D displays reduce conflict between accommodative and vergence responses. *Journal of the Society for Information Display* 24, 12 (2016), 747–756. https://doi.org/10.1002/jsid.520

[72] Hajime Morohashi, Kenichi Hakamada, Takahiro Kanno, Kenji Kawashima, Harue Akasaka, Yuma Ebihara, Eiji Oki, Satoshi Hirano, and Masaki Mori. 2021. Social implementation of a remote surgery system in Japan: a field experiment using a newly developed surgical robot via a commercial network. *Surgery today* (2021), 1–10.

[73] Rahul Narain, Rachel A. Albert, Abdullah Bulbul, Gregory J. Ward, Martin S. Banks, and James F. O'Brien. 2015. Optimal Presentation of Imagery with Focus Cues on Multi-plane Displays. *ACM Trans. Graph.* 34, 4, Article 59 (July 2015), 12 pages. https://doi.org/10.1145/2766909

[74] Kephart Newell C. 1962. *The Slow Learner in the Classroom.* C. E. Merrill Books, Columbus. 292 pages.

[75] Karin Nieuwenhuizen, Jean-Bernard Martens, Lei Liu, and Robert van Liere. 2009. Insights from Dividing 3D Goal-Directed Movements into Meaningful Phases. *IEEE Computer Graphics and Applications* 29, 6 (Nov 2009), 44–53. https://doi.org/10.1109/MCG.2009.121

[76] Oculus. 2020. Quest 2. https://www.oculus.com/

[77] Heeseok Oh, Sanghoon Lee, and Alan Conrad Bovik. 2016. Stereoscopic 3D visual discomfort prediction: A dynamic accommodation and vergence interaction model. *IEEE Transactions on Image Processing* 25, 2 (2016), 615–629. https://doi.org/10.1109/TIP.2015.2506340

[78] Nitish Padmanaban, Robert Konrad, Emily A. Cooper, and Gordon Wetzstein. 2017. 4-3: Invited Paper: Gaze-contingent Adaptive Focus Near-eye Displays. *SID Symposium Digest of Technical Papers* 48, 1 (2017), 23–25. https://doi.org/10.1002/sdtp.11566 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/sdtp.11566

[79] Nitish Padmanaban, Robert Konrad, Tal Stramer, Emily A. Cooper, and Gordon Wetzstein. 2017. Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proceedings of the National Academy of Sciences of the United States of America* 114, 9 (2017), 2183–2188. https://doi.org/10.1073/pnas.1617251114

[80] Robert Patterson. 2009. Review Paper: Human factors of stereo displays: An update. *Journal of the Society for Information Display* 17, 12 (2009), 987. https://doi.org/10.1889/jsid17.12.987

[81] Robert Patterson and Wayne L. Martin. 1992. Human stereopsis. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 34, 6 (dec 1992), 669–92. https://doi.org/10.1177/001872089203400603

[82] Andriy Pavlovych and Wolfgang Stuerzlinger. 2009. The tradeoff between spatial jitter and latency in pointing tasks. In *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems*. ACM, 187–196.

[83] Sowmya Ravikumar, Kurt Akeley, and Martin S. Banks. 2011. Creating effective focus cues in multi-plane 3D displays. *Opt. Express* 19, 21 (Oct 2011), 20940–20952. https://doi.org/10.1364/OE.19.020940

[84] Rebekka S. Renner, Boris M. Velichkovsky, and Jens R. Helmert. 2013. The perception of egocentric distances in virtual environments - A review. *Comput. Surveys* 46, 2 (nov 2013), 1–40. https://doi.org/10.1145/2543581.2543590 arXiv:arXiv:1502.07526v1

[85] Jannick P. Rolland, Myron W. Krueger, and Alexei Goon. 2000. Multifocal planes head-mounted displays. *Appl. Opt.* 39, 19 (Jul 2000), 3209–3215. https://doi.org/10.1364/AO.39.003209

[86] W. N. Schofield. 1976. Do children find movements which cross the body midline difficult? *Quarterly Journal of Experimental Psychology* 28, 4 (1976), 571–582. https://doi.org/10.1080/14640747608400584

[87] Gurjot Singh, Stephen R. Ellis, and J. Edward Swan. 2018. The Effect of Focal Distance, Age, and Brightness on Near-Field Augmented Reality Depth Matching. *IEEE Transactions on Visualization and Computer Graphics* PP, c (2018), 1. https://doi.org/10.1109/TVCG.2018.2869729 arXiv:1712.00088

[88] Gurjot Singh, J. Edward Swan, J. Adam Jones, and Stephen R. Ellis. 2010. 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)*. ACM Press, New York, New York, USA, 149–156. https://doi.org/10.1145/1836248.1836277

[89] R. William Soukoreff and I. Scott MacKenzie. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies* 61, 6 (dec 2004), 751–789. https://doi.org/10.1016/j.ijhcs.2004.09.001

[90] Frank Steinicke, Klaus H Hinrichs, Johannes Schöning, and Antonio Krüger. 2008. Multi-touching 3D data: Towards direct interaction in stereoscopic display environments coupled with mobile devices. In *Proc. AVI Workshop on Designing Multi-Touch Interaction Techniques for Coupled Public and Private Displays*. Citeseer, 46–49.

[91] Helmut Strasser and Karl-Werner Müller. 1999. Favorable movements of the hand-arm system in the horizontal plane assessed by electromyographic investigations and subjective rating. *International Journal of Industrial Ergonomics* 23, 4 (mar 1999), 339–347. https://doi.org/10.1016/S0169-8141(98)00050-X

[92] J. Edward Swan, Gurjot Singh, and Stephen R. Ellis. 2015. Matching and Reaching Depth Judgments with Real and Augmented Reality Targets. *IEEE Transactions on Visualization and Computer Graphics* 21, 11 (nov 2015), 1289–1298. https://doi.org/10.1109/TVCG.2015.2459895

[93] Robert J. Teather, Andriy Pavlovych, Wolfgang Stuerzlinger, and I. Scott MacKenzie. 2009. Effects of tracking technology, latency, and spatial jitter on object movement. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI '09)*. IEEE, 43–50. https://doi.org/10.1109/3DUI.2009.4811204

[94] Robert J. Teather and Wolfgang Stuerzlinger. 2011. Pointing at 3D targets in a stereo head-tracked virtual environment. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI '11)*. IEEE, 87–94. https://doi.org/10.1109/3DUI.2011.5759222

[95] Julian J. Tramper and C. C. A. M. Gielen. 2011. Visuomotor Coordination Is Different for Different Directions in Three-Dimensional Space. *Journal of Neuroscience* 31, 21 (2011), 7857–7866. https://doi.org/10.1523/JNEUROSCI.0486-11.2011 arXiv:https://www.jneurosci.org/content/31/21/7857.full.pdf

[96] Takaaki Ueno and Yasuhiro Takaki. 2018. Super multi-view near-eye display to solve vergence–accommodation conflict. *Opt. Express* 26, 23 (Nov 2018), 30703–30715. https://doi.org/10.1364/OE.26.030703

[97] Varjo. 2021. VR-3.

[98] Cyril Vienne, Laurent Sorin, Laurent Blondé, Quan Huynh-Thu, and Pascal Mamassian. 2014. Effect of the accommodation-vergence conflict on vergence eye movements. *Vision Research* 100 (2014), 124–133. https://doi.org/10.1016/j.visres.2014.04.017

[99] W.H. Welch, P.M. Greco, R. Abovitz, Y. Munk, and S.A. Miller. 2017. Virtual and Augmented Reality Systems and Methods.

[100] Austin Wilson and Hong Hua. 2019. Design and demonstration of a vari-focal optical see-through head-mounted display using freeform Alvarez lenses. *Opt. Express* 27, 11 (May 2019), 15627–15637. https://doi.org/10.1364/OE.27.015627

[101] 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

[102] Roberts Zabels, Krišs Osmanis, Mārtiņš Narels, Uģis Gertners, Ainārs Ozols, Kārlis Rūtenbergs, and Ilmārs Osmanis. 2019. AR Displays: next-generation technologies to solve the vergence–accommodation conflict. *Applied Sciences* 9, 15 (2019), 3147.

[103] Charles M. Zaroff, Magosha Knutelska, and Thomas E. Frumkes. 2003. Variation in Stereoacuity: Normative Description, Fixation Disparity, and the Roles of Aging and Gender. *Investigative Ophthalmology & Visual Science* 44, 2 (feb 2003), 891. https://doi.org/10.1167/iovs.02-0361