This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3403261

Effects of Eye Vergence and Accommodation on Interactions with Content on an AR Magic-lens Display and its Surroundings

Geert Lugtenberg[*]
Nara Institute of Science and Technology

Klen Čopič Pucihar[†]
University of Primorska

Matjaž Kljun[‡]
University of Primorska

Taishi Sawabe[§]
Nara Institute of Science and Technology

Yuichiro Fujimoto[¶]
Nara Institute of Science and Technology

Masayuki Kanbara[‖]
Nara Institute of Science and Technology

Hirokazu Kato[**]
Nara Institute of Science and Technology

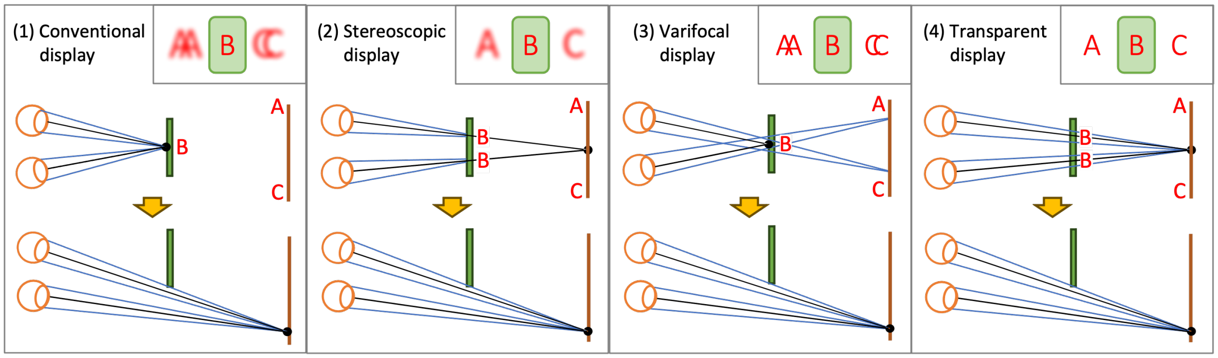

Figure 1: Depending on the type of magic-lens display (green), eye accommodation (blue lines) and/or vergence (black lines) distances must change as the observer shifts from a near display (green) to the far surroundings (brown), or vice versa. The top-right of each scenario shows the first-person view of letters "A," "B," and "C" when the gaze is fixed on letter "B." We considered four display types: (1) Conventional displays. "A" and "C" are out of focus and have double vision. Both the vergence and accommodation distances must change as the gaze shifts (yellow arrow) toward the surroundings. (2) Stereoscopic displays. "A" and "C" are out of focus; only the accommodation must change. (3) Varifocal displays. "A" and "C" are observed with double vision; only the vergence must change. (4) Transparent displays. All three letters in focus and without double vision; no accommodation or vergence changes are necessary.

ABSTRACT

Augmented reality (AR) magic-lens (ML) displays, such as handheld devices, offer a convenient and accessible way to enrich our environment using virtual imagery. Several display technologies, including conventional monocular, less common stereoscopic, and varifocal displays, are currently being used. Vergence and accommodation effects on depth perception, as well as vergence–accommodation conflict, have been studied, where users interact only with the content on the display. However, little research exists on how vergence and accommodation influence user performance and cognitive-task load when users interact with the content on a display and its surroundings in a short timeframe. Examples of this are validating augmented instructions before making an incision andperforming general hand-eye coordinated tasks such as grasping augmented objects. To improve interactions with future AR displays in such scenarios, we must improve our understanding of this influence. To this end, we conducted two fundamental visual-acuity user studies with 28 and 27 participants, while investigating eye vergence and accommodation distances on four ML displays. Our findings show that minimizing the accommodation difference between the display and its surroundings is crucial when the gaze between the display and its surroundings shifts rapidly. Minimizing the difference in vergence is more important when viewing the display and its surroundings as a single context without shifting the gaze. Interestingly, the vergence–accommodation conflict did not significantly affect the cognitive-task load nor play a pivotal role in the accuracy of interactions with AR ML content and its physical surroundings.

Index Terms: Augmented Reality—Human-computer interaction—Video see-through display—Vergence-accommodation

1 INTRODUCTION

Augmented reality (AR) has long been envisioned as a support system for both every day and specialized tasks [4, 15, 47]. The widespread availability of smartphones equipped with dedicated graphical processing units, high-quality cameras, and displays has enabled average users to augment the world using virtual imagery. By utilizing a smartphone camera image, the user can view the physical world together with the digital content as if the handheld device is a transparent screen or lens. Such video see-through displays are also referred to as a magic-lens (ML) displays [8].

The popularity of augmented content in optical see-through displays has recently increased. In contrast to ML displays, optical see-through displays use transparent or semi-transparent materials to superimpose virtual imagery. This method offers a natural way of viewing the real environment as light passes through the material and maintains the user's perspective. This is one of the major advantages of this type of display, because only augmented content

---

[*] e-mail: geert.lugtenberg.gf3@is.naist.jp
[†] e-mail: klen.copic@famnit.upr.si
[‡] e-mail: matjaz.kljun@famnit.upr.si
[§] e-mail: t.sawabe@is.naist.jp
[¶] e-mail: yfujimoto@is.naist.jp
[‖] e-mail: kanbara@is.naist.jp
[**] e-mail: kato@is.naist.jp

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

requires generation and visualization. Thus, no computing power or time has to be spent on the visualization of the natural environment, and could be better spent improving the virtual imagery quality and update speed. However, similar to ML displays, virtual imagery is commonly presented in a fixed image plane. Therefore, if the surroundings and image-plane distances are not aligned, the image falls out of focus. Another drawback of optical see-through displays is that they are expensive and difficult to access. Typically worn on the head, these displays, known as head-mounted displays (HMD), offer the advantage of being hands-free. However, given the limited availability of haptic interactions, the provided input interface can be difficult to operate. Prolonged use of HMDs can lead to fatigue [7, 33] and might, in certain situations, prove cumbersome or impossible owing to the presence of other headgear or environmental constraints.

Displaying augmented content on a handheld device is affordable and provides users with a familiar touchscreen interface. However, this approach has several disadvantages related to vergence and accommodation when interacting with a display or its surrounding environment. *Vergence*, also known as binocular convergence, is a property of stereoscopic systems, in which both eyes rotate to allow light to converge at retinal centers, where vision is the sharpest and most detailed. While *accommodation* is a property of the eye that involves stretching the lens to adjust the focus and maintain clear vision at varying distances, both vergence and accommodation provide strong cues that are essential for accurate depth perception. A challenge arises when using the conventional monoscopic ML setup (Fig. 1.1)), where the eyes converge and are accommodated at the same distance (B) on the display, irrespective of the actual distance to the real surface being visualized. Consequently, when shifting gaze from display (B) to its surroundings (A or C), both eyes must readjust their vergence and accommodation distances. This process is required to bring A or C into focus and remove double vision (top-right of Fig. 1.1). The same readjustment is required when shifting the gaze back to display (B). This process requires time and effort [40], potentially hindering smooth interactions with the augmented content displayed on the ML and its surroundings.

When shifting the gaze between the display and its surroundings, the optimal approach to alleviate eye effort is to have the vergence and accommodation set equally on the surroundings (Fig. 1.4). In such a system, gaze shift does not require eye adjustment, and A, B, and C are always perfectly focused, that is, without double vision. However, creating such a display is technically challenging. An alternative method to alleviate eye effort is to accommodate the display for the surrounding distance (Fig. 1.3), such as in varifocal displays, in which the AR content plane can accommodate any distance. When shifting gaze from the display to its surroundings, only vergence requires change (because the display is monoscopic) to resolve the double vision of A and C (Fig. 1.3 top-right). While this approach appears promising, it introduces a phenomenon termed vergence–accommodation conflict (VAC) [21]; a mismatch between the focusing distance (accommodation) and vergence distance of the eyes, which is common in mixed-reality displays (i.e., augmented and virtual reality). This conflict causes eye strain and is expected to hinder the ability to interact with AR content and its surroundings. Additionally, display systems, such as varifocal displays, require large, complex optics that pose another technical challenge. A final option to alleviate eye effort when shifting the gaze between the ML and its surroundings is to set the vergence distance of the display at or close to the surrounding vergence distance such that only accommodation has to change to achieve perfect focus for A and C (Fig. 1.2). Such a system only requires a stereoscopic display (such as a lenticular screen, multiview [50] or light-field [49] display); however, it also suffers from a vergence–accommodation conflict.

Despite the importance of being able to simultaneously see and understand augmented content on a ML display alongside its surroundings, the *individual* influence of eye accommodation, vergence and the existence of VAC on human performance has not yet been studied. In theory, users can use two different strategies when interacting with the display content and its physical surroundings, depending on the location of the physical content. The first is rapidly shifting gaze between the display and its surroundings (when the required physical content is far from the edge of the display), whereas the second strategy involves viewing the display and its surroundings in a single context (when the physical content is close to the edge of the display). In such cases, user performance and task loads can be influenced by the combination of changing vergence and accommodation (Fig. 1.1), changing accommodation (Fig. 1.2), changing vergence (Fig. 1.3), and when the vergence-accommodation conflicts (Fig. 1.2 in .3). If users do not accommodate or verge at the correct distance, the result may be a blurry image or double vision. However, it is unclear which of these has a greater effect on the user performance and task loads. Knowing which technology is better for assisting users interacting with AR content on the display and its surroundings within a short timeframe (while switching gaze or viewing the context as a whole), will increase the success rate of task completion and improve user experience.

To address this knowledge gap, we developed a system capable of recreating all four types of ML display (Fig. 1). We used this system to conduct two visual-acuity user studies ($n = 28$ and $n = 27$) under four vergence and accommodation conditions. During the first user study, we measured user performance and task loads in recognizing eye-test symbols, while rapidly shifting the gaze between the display and its surroundings. In the second study, users perceived the display and its surroundings as a whole and had to recognize eye-test symbols within a short viewing time. Our first finding indicates that when users rapidly shift their gaze between the display and its surroundings, the change in the accommodation distance (Fig. 1.1 and 2) significantly influences performance, whereas eye vergence does not play a major role. Our second finding indicates that when the display and its surroundings are perceived as a single context (as the gaze does not shift, the eyes always verge and accommodate the display), performance improves the most when the vergence distance of the ML display is close to that of its surroundings, such that there is no double vision (Fig. 1.2). The results indicate that stereoscopic displays provide an affordable solution for quick interaction between the display and its surroundings compared with other displays.

2 BACKGROUND & RELATED WORK

To investigate the impact of vergence and accommodation on the interaction between a display and its surroundings, it is essential to understand how these eye phenomena affect everyday vision.

2.1 Accommodation and eye strain

The process of bending light to focus it on the retina involves the contraction or relaxation of the eye lens, adjusting its convexity. When this accommodation process does not function optimally, objects near the eye (hyperopia) or at a certain distance (myopia) appear blurry. Myopia is a common phenomenon that occurs in approximately 23% of people [22], and varies with age, ethnicity and lifestyle. In addition, middle- and older-aged adults often have difficulty seeing things in proximity (i.e., presbyopia). Although these conditions can be corrected with convex or concave lenses, either mounted on spectacles or in the form of contact lenses, neglecting treatment may lead to eye fatigue, headaches, and overall impairment of daily activities. Moreover, the extended use of near-view displays such as computer monitors, tablets, and smartphones causes accommodation-related symptoms and subsequent eye discomfort [18, 24, 25].

In mixed-reality systems, virtual content is often displayed at a fixed focal distance, leading to a discrepancy in the accommodation distance between virtual and real environments, causing either

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3403261

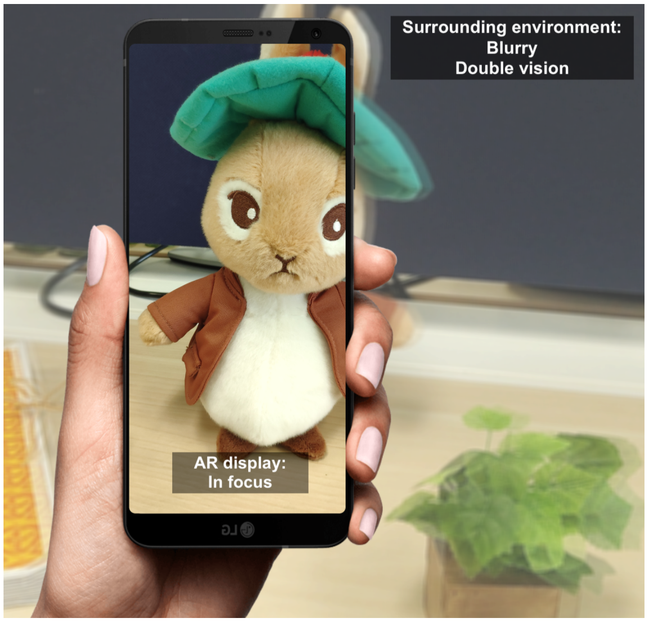

Figure 2: Display content as rendered from the perspective of the user. Focusing on the handheld display causes the surroundings to become blurred and doubled (diplopia) due to eye vergence and accommodation respectively.

to be out of focus. Existing research on AR systems that employ HMD focuses on mitigating this issue with varifocal technology [1], where the focal point is either mechanically changed or multiple focal planes exist. Near-eye light-field displays [23, 34] address the accommodation issue by rendering scenes from various viewpoints, resulting in different depths per viewing angle. Koulieris et al. surveyed state-of-the-art AR and virtual reality (VR) near-eye displays including those with unfixed focal distances [29]. They classified accommodation-supporting near-eye displays as varifocal/multifocal, multiplane, focal surface, and holographic, using a combination of lens optics and screen techniques. Techniques such as the Maxwellian view [51] project a virtual image onto a specific part of the eye retina, most commonly the fovea. This requires careful calibration, considering the positional relationship between the eyes and display and has been previously used in HMDs [37]. Interestingly, AR ML displays, which are commonly implemented in handheld devices or MLs, encounter similar accommodation challenges because the virtual content focal plane is fixed at arm's length. However, they received considerably less attention than their HMD counterparts, which partly motivated the proposed work.

2.2 Vergence and diplopia

Vergence is stimulated by the stereo disparity images created when the two eyes "collaborate" to converge images in a unified binocular vision. It primarily involves rotation of the eyes to ensure that the fixation area of the image falls precisely at the center of the retina in both eyes. However, vergence is also driven by blur and accommodation (see Sect. 2.4). Disparity in the images of objects that do not converge can be (voluntarily) perceived as double vision or diplopia. This effect is illustrated in Fig. 2 within the surroundings, when the gaze is fixed on the AR display. Although synthetic double vision has been used as a depth cue in AR [14, 46], HMDs commonly display disparate images to each eye to facilitate a sense of depth [10, 28, 50]. In mixed-reality studies utilizing projector displays, stereoscopic images are obtained with polarized glasses that filter frames intended for each eye [9, 14, 32]. Additionally, related works using (auto) autostereoscopic AR ML [6, 39, 44, 50] can produce similar depth perceptions.

2.3 Vergence-accommodation conflict

Numerous studies have focused on the fatigue and performance problems arising from the conflict between vergence and accommodation distances [21]. This problem is particularly prevalent in mixed-reality contexts, which are experiencing a surge in popularity. Both VR and AR using HMDs [19, 29, 31, 48, 53, 54] suffer from vergence-accommodation conflicts (VAC), which contribute to lower adoption rates in practical applications. The aforementioned study [29] addressed VAC by matching the binocular disparity of virtual imagery with optical focal cues at various depths. Discomfort studies using HMDs [30] have shown that only focus-adjustable lens designs can accommodate simulated distances to significantly improve comfort. In a ML system similar to that proposed in this study, researchers provided a 3D viewing experience through parallax images with directional rays coming from a Super multi-view (SMV) lenticular lens [41]. By equalizing the eye focus and convergence distance of the virtual image and measuring the accommodative response, they concluded that SMV can reduce the effects of VAC. However, the extent to which VAC would affect performance or task load in a pure stereoscopic ML (Fig. 1.2) or varifocal ML (Fig. 1.3), remains to be explored. Notably, researchers found minimal fatigue and discomfort with viewing distances at TV level (4.5 m) [52]. However, to the best of our knowledge, distances equivalent to arm's length have yet to be evaluated.

2.4 Vergence and accommodation interaction

Eye vergence and accommodation form an interconnected visual system, meaning that changes in one influence or even drive changes in the other; this is the so-called accommodation-convergence reflex. The accommodation response can be driven by blur and stereo disparity between the two eyes. Similarly, accommodative vergence is a blur-driven response that converges or diverges from an eye. Studies have shown that, as accommodative responses deteriorate with age, the interaction in which vergence drives accommodation increases [20]. Therefore, it is expected that age has an impact on the ability to merge a near-view screen with its far-view surroundings, as well as any eye condition that influences the accommodation-convergence reflex. Blurry vision caused by insufficient accommodation affects tasks that require precision such as reading and writing, detailed work, and face (expression) recognition. However, double vision caused by incorrect vergence affects most tasks that rely on accurate binocular vision and depth perception, such as hand-eye coordinated tasks, reaching for objects, and driving. It was observed [36] that small-separation diplopia also negatively affects reading ability. Given that both phenomena affect reading, we decided to use the recognition of text characters to test accuracy in our forthcoming user studies.

Theories in the field of optics and ophthalmology describe eye vergence and accommodation reaction times to accommodative stimuli such as blur, apparent size, and distance. The authors of [26] reported an average reaction time of 0.62 and 0.56 s for "far-to-near" and "near-to-far" accommodation change, respectively. The convergence response times were considerably faster, averaging approximately 0.20 s. This observation is interesting, as it suggests that the accommodative response time is a bottleneck when the fixation point shifts between different distances. It follows logically that reducing the accommodation distance would make reaction times faster up to the point where the vergence time becomes an issue. To the best of our knowledge, these reaction times have not yet been studied in the context of AR ML displays. Therefore, in our first user study (Study A), we verify whether these response times are similar in the context of a rapidly shifting gaze between the ML display (near plane) and the surroundings (far plane). In the second

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

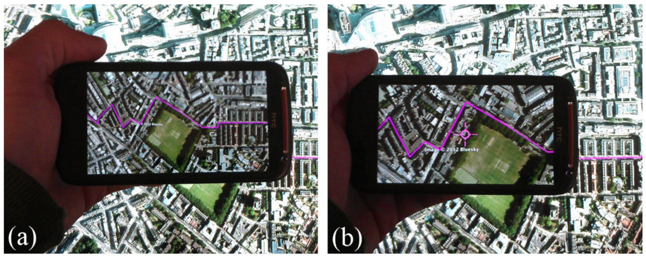

Figure 3: (a) Dual-view problem in video see-through MLs comes from the fact that the camera is positioned on the back of the device and thus has a different perspective than the observer. (b) When rendered from the perspective of the user, the dual-view is not present. Image taken from [11] with permission from the authors.

user study (Study B), a fixed gaze was maintained to investigate the resulting effects from incongruous vergence and accommodation – blur and diplopia, respectively.

2.5 Magic-lens systems

Previous studies on ML systems have predominately focused on handheld devices owing to their accessibility and technological advantages. Contemporary smartphones and tablets are equipped with a range of built-in sensors that are useful in AR applications. There is a difference between an ML that simply visualizes an on-device camera image as a background [8] and an ML that visualizes geometrically correct views within the lens area, as seen from the user's perspective [5, 11, 42]. In [12], the researchers investigated both types, as shown in Fig. 3, and found that users consider the real environment and ML as separate views when using device perspective rendering. This perception also holds true when a user-perspective ML is not sufficiently performant. In such cases, users will rapidly shift their gaze between the ML area and the surrounding real environment to interact with both rather than looking at the scene as a whole. This observation serves as the motivation for the first user study (Study A), in which the strategy of rapid gaze shifting was employed. Other optical see-through ML displays include heads-up displays (HUD) [43] used in aviation that allow pilots to see the runway even in bad weather conditions, and automotive industries [27, 35], to provide information such as speed or navigation on their windshield whilst driving. Particularly with the increasing prevalence of electric cars equipped with HUDs, there is a growing frequency of gaze switches between virtual information displayed at the HUD distance and real-world information. Consequently, understanding the impact of each visual system on performance becomes increasingly important.

2.6 Context and depth switching

Owing to the popularity of mixed reality, and its increasing usage for task support in the industry, previous studies have investigated the effect that switching depth layers (i.e. between the real physical environment and virtual imagery) has on human performance in visual tasks. Eiberger et al. measured task completion time and error rate on a combination of optical see-through HMD and a body-proximate display at 30 cm [16]. They found that during a visual search task, when content is on different depth layers (i.e. it is necessary to refocus and verge), performance significantly decreased compared to a visual search on a single depth layer. Using a monocular near-eye display, Gabbard et al. similarly showed decreased performance when focal distance needed to switch and that repetitions increased visual fatigue [17]. When replicated and extended these studies by using a binocular AR Haploscope, Arefin et al. additionally found that only increasing focal switching distance degraded performance in the binocular condition [2]. In these studies, researchers maintain consistent accommodation and vergence distances, likely due to the well-documented adverse effects of VAC. Consequently, the individual impacts of these visual systems (Fig. 1.2 and 1.3) on performance remain unclear, which is the main goal of this research.

3 USER STUDIES

3.1 Hypotheses

Our objective was to measure visual-acuity under different conditions of vergence and accommodation distances, influenced by different types of ML displays located at arm's length and far surroundings. To interact with the (augmented) content on the ML display and its surroundings, we investigated two strategies: (1) rapid gaze shifting between the ML display and its far surroundings and (2) fixed gaze on the ML display while looking at the display and its surroundings in a single context. According to the literature [26], the accommodation response is significantly slower than the vergence response when gaze fixation moves from near-to-far or from far-to-near. Therefore, we hypothesize as follows:

> **(H1)** When rapidly shifting gaze, reducing the eye accommodation distance of a ML display in relation to its surroundings, results in faster interaction with AR content, more so than reducing the vergence distance.

Using the second strategy, eyes do not re-accommodate nor reconverge on different focal distances. However, artifacts resulting from misaligned accommodation distance (blur) and vergence distance (double vision) persist. Therefore, we hypothesize as follows:

> **(H2)** When gaze is fixed on a ML display, reducing its eye vergence and accommodation distance in relation to its surroundings will result in a more accurate merging of the AR content.

Given the prevalent findings in the literature (Sect. 2.3) regarding fatigue and eye strain in VR scenarios caused by vergence–accommodation conflicts, we also expect the following:

> **(H3)** A mismatch between the eye accommodation distance and vergence distance of a ML display requires the viewer to concentrate more, resulting in a higher perceived task load and less accurate interaction between the AR content and its surroundings.

3.2 System design

We considered a typical scenario in which a handheld device was used as an AR system. The average near-working distance between the user's eyes and the smartphone screen is between 32 cm and 36 cm [3]. Accordingly, we fixed the user's head at 35 cm away from the ML *near display*. This ML *near display* was mounted and was not adjustable by the user. For the surroundings, we selected a plane 1 m away from the user, which is referred to as the *far display*. A top-down schematic is shown in Fig. 5.

3.2.1 Accommodation distance

Through defocused blurring, the eye can be accommodated on a surface that emits or reflects light. In the proposed setup, shown in Fig. 5, we were able to move the *near display* closer and further along the depth axis of the viewer (dotted line). Accommodation was controlled by altering this distance.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3403261

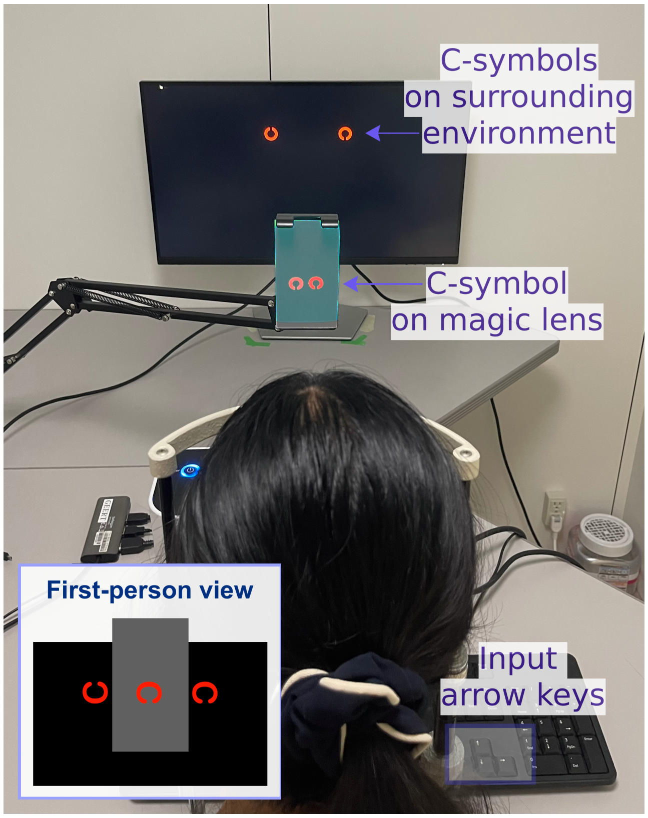

Figure 4: Experimental setup with the perspective from behind the participant, and their first-person view. The system is running a task for Study B, where three symbols are displayed: two outer symbols are displayed in the background on the surrounding display, whereas the middle symbol is stereo-projected onto the ML display. By wearing polarizing 3D glasses, the participant perceived only one symbol in the middle, with its depth determined by the amount of disparity. Arrow keys on a keyboard in front of the participant allowed for input of the direction of the openings of the C-symbols during the experiments. The diameter of the C-symbol on the ML display was 1.9 cm and the C-symbols on the surrounding display were scaled so that all three sizes appeared equal to the viewer.

3.2.2 Vergence distance

In order to control the distance on which the user's eyes converge, we employed stereoscopic rendering. In this type of rendering, each eye is presented with a different disparity image. Binocular disparity links these two images for different eyes as one, forming a strong depth cue. One approach to achieve stereoscopic rendering is by using an autostereoscopic display that uses a parallax barrier, e.g. lenticular lens array, to visualize stereo images. However, when we want to adjust the accommodation distance, an autostereoscopic display poses problems because its viewing angle, in combination with the distance, is fixed. Furthermore, these displays are susceptible to crosstalk between eye images. An easier solution is to use a stereoscopic projector and adjust its focal length according to the *near display* (Fig. 4 and Fig. 6), in combination with active shutter glasses. By wearing active shutter glasses the user perceives a stereo image of the ML surface, allowing us to control the vergence variable.

3.2.3 Stereoscopic rendering

There are limitations to the depth effect created by binocular disparity images in stereoscopic rendering. As a rule of thumb, the interaxial distance should not exceed 1/30th of the convergence distance [38], assuming an interocular distance of 6.35 cm. In the proposed setup, this enables a range of 30 cm to 45.7 cm of comfortable stereo viewing for the *near-ML display* and a range of 75 cm to 105 cm for the *far display* (Fig. 5). The desired vergence distance of 105 cm away from the viewer projected onto the *near display* was not possible. Similarly, verging 35 cm away from the viewer while projecting onto the *far display* was not possible.

We performed a pilot study with nine participants using the described system, in order to determine acceptable values for eye-acuity symbol size and horizontal distance, as well as color and disparity capabilities. Using a trial and error approach in which the examiner continually adjusted the binocular disparity, it was found that the range in which the user could perceive a symbol clearly in a 3D space ranged from approximately 25 cm to 70 cm from the viewer in the *near*-accommodation *display* setup (orange range in Fig. 5). In the *far*-accommodation setup, the clear viewing range was approximately 75 cm to 105 cm from the viewer. These values varied slightly per participant, depending on eye health, age, ability to focus, and interocular distance and increased with familiarity and practice. Therefore, calibration of each user's stereoscopic ability and practice before using the system is recommended and is part of the forthcoming user studies.

3.2.4 Eye-acuity symbols

We measured eye acuity by the participant correctly discerning three symbols in a row. A common practice in eye examinations is to use Snellen chart symbols [13], which consist of capital letters progressively becoming smaller. In the aforementioned pilot study, where Sloan letters (a subset of the Snellen chart) were used, participants needed a considerable amount of time to input their answers on a keyboard and to take their head off the mount to have a clear view of the keyboard letters. When testing the verbal confirmation of the answer letters, accidental input of an incorrect answer was more likely due to confusion between participant's intended answer (pronunciation and shared attention) and the examiner's interpretation of the answer. Therefore, we used Landholt C-symbols [45] to test eye acuity. These consist of C-symbols that can have openings in one of the four directions: up, down, left, right.

We wanted a high contrast between the symbols and background on the surrounding display to make it easier to discern the openings of the symbols. Therefore, we dimmed the light and used a black background with red colour symbols. Red was found to be a good trade-off between contrast and the resulting ghosting effect from stereoscopic projection; when using white symbols (highest contrast), some users were able to perceive two images on the stereoscopic display instead of one 3-dimensional image.

When tasked with discerning symbols, their size is an important factor. Together with the viewing time, these two variables would make our user study complex and would require long experimentation times to gather sufficient repetitions per condition. Therefore, we used C-symbols of a fixed size (1.9 cm in diameter) that were determined in the pilot study to be just discernible (visual angle of approximately 5 degrees) on the far-surrounding display while focusing on the ML (see Fig. 4). Increasing this minimum size is expected to positively affect symbol recognition. This mimics a real scenario in which the user looks at the display and its surroundings in a single context.

3.3 Experimental conditions

In the experiments, we used a 2 × 2 within-subjects design with four combinations of vergence and accommodation variables, each of which could be *near* or *far*. In the accommodation-near condition, a stereoscopic image was projected onto the ML display (example seen in Fig. 4), and the vergence distance was either on the display (*near*) or 70 cm away from the participant (*far*). In the accommodation *far*

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

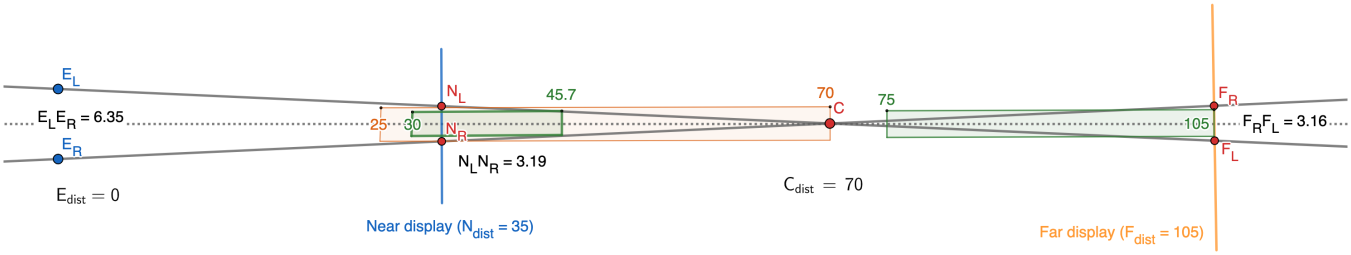

Figure 5: Binocular fusion limitations of stereoscopic images that contributed to how we built our proposed setup. Values in centimeters. Average human interocular distance is 6.35 cm between left eye E_L and right eye E_R, a *near display* is 35 cm away and a *far display* 105 cm away from the viewer. Green sections are the comfortable viewing ranges outside of which the viewer can experience difficulty merging the stereoscopic images. They are defined by the local minima and maxima of convergence point C where $N_LN_R < \frac{C_{dist}}{30}$ and $F_RF_L < \frac{C_{dist}}{30}$. The orange section is the stereoscopic distance range that could be obtained by participants in the user study using the proposed system.

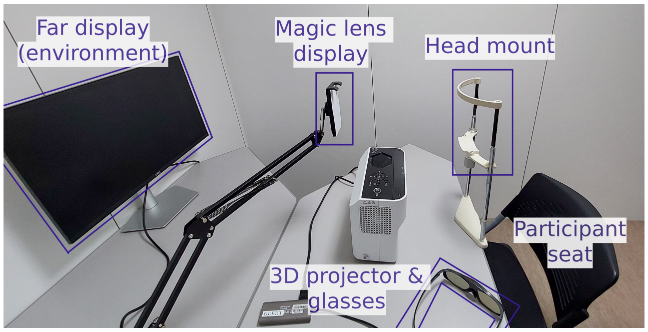

Figure 6: Experimental environment from the side. The distance from the *far display* (surrounding environment) to the ML was 70 cm, and the distance from the ML (*near display*) to the head mount was 35 cm. The short-throw 3D projector on the table projects a stereoscopic image onto a white sheet (6.5 cm × 12 cm) mounted on a movable clamp, representing the near display.

condition, the stereoscopic image was projected onto the surrounding display, and the vergence distance was set to either the surroundings distance of 105 cm (*far*) or 75 cm away from the participant (*near*). These two conditions were counterbalanced among the participants.

3.4 Study A: Shifting gaze

3.4.1 Apparatus

For the surrounding display we vertically mounted a 1 m high by 2 m wide white plastic panel on a table. Projection onto this panel was performed using an Optoma WU515ST projector. In front of this panel was positioned another 0.15 m × 0.15 m white plastic panel that functioned as the ML display. We used a Barco F50 WQXGA projector to display the stereoscopic side-by-side images. The 3D projector's image was projected onto the surrounding display in the accommodation-*far* condition. To observe the stereoscopic image, we used active shutter glasses with digital light processing (DLP) technology.

3.4.2 Task

Participants were directed to use arrow keys on a keyboard placed in front of them to indicate the direction of the openings of three C-symbols, proceeding from left to right. For example, in the round shown in Fig. 4, the participant presses the up key, down key, and down key again. During this first experiment, the left and right C-symbols were placed horizontally 1 m (50.8 degrees) apart behind the ML display. This was determined to be sufficiently wide for the participants to be forced to shift their gaze between symbols.

Following a three-second countdown shown in the top-center of the surrounding display, the three C-symbols would appear simultaneously in random orientations. Participants were asked to focus on the position of the left symbol during the countdown and, as soon as the symbols appeared, to shift their gaze from left to middle and then to right while pressing their choice of arrow key (three times). For each symbol, we captured response correctness and response time.

3.4.3 Participants and procedure

We recruited 28 university students (14 males and 14 females) aged between 18 and 32 (M=21, SD=3.25). Five participants had prior experience with mixed-reality systems, whereas 22 had no or limited experience. We verified good visual acuity (with correction if necessary) in all participants.

This study consisted of two phases: preparation and data-gathering. In the preparation phase, the participants were asked to sit on a chair in front of the experimental setup, and the intention of the experiment was explained. They were then asked to wear shutter glasses, and the light in the room was dimmed. The examiner checked their ability to see the stereo depth by displaying a C-symbol on the ML display and asking the participant to estimate their distance from them. This was performed with various degrees of stereo disparity and was repeated on the surrounding display.

In the data-gathering phase, the examiner adjusted the setup according to the counterbalanced conditions of accommodation and vergence, followed by a practice round. This round consisted of 10 repetitions of the task described in Sect. 3.4.2, during which participants were allowed to ask questions. Following sufficient practice, when ready, participants were asked to focus and complete the task 40 times as quickly as possible. When finished, the participants were asked to take off their shutter glasses, and rest their eyes, during which the examiner re-adjusted the experiment setup according to the counterbalanced condition of accommodation and vergence. Then the second phase (practice and task completion) was repeated.

The entire procedure lasted approximately 30 minutes including three breaks of 3 minutes. The study received prior approval from the institutional review board.

3.5 Study B: Fixed gaze

3.5.1 Apparatus

We used a Dell desktop monitor in front of which was a device holder mount with a white surface that served as a ML display surface. We used a Ricoh PJWX4153N projector with stereoscopic side-by-side projection in combination with BenQ YDD3PG active shutter glasses that worked with the DLP-link technology of the projector. The two displays were placed in a row on two tables and a head mount was attached to the front table. The mount maintained

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3403261

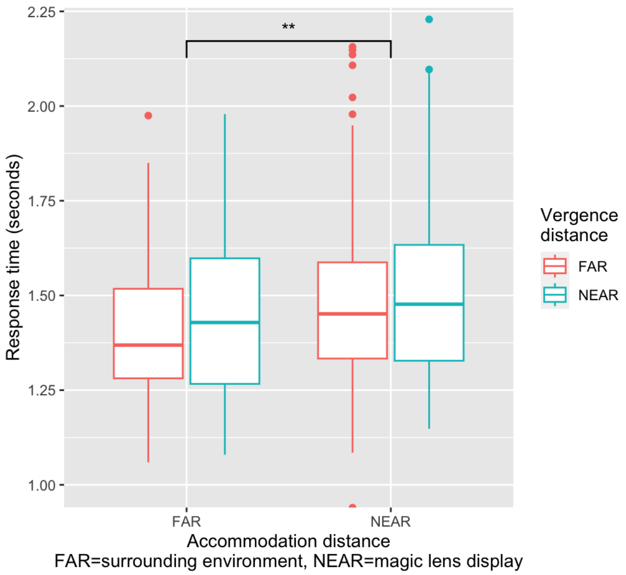

Figure 7: Study A performance results. The graph shows the average response time when shifting gaze between symbols on the surrounding (far plane) and a ML display (near plane) under four conditions of ML vergence distance and accommodation distance. Altering the accommodation distance of the ML display has a significant ($p < .01$) effect on response time, while vergence does not.

the participants' head stability in a calibrated position throughout the experiment. The Dell desktop monitor functioned as the surrounding display and was replaced by the stereo projector in the *far*-accommodation condition. The setup for the *near*-accommodation condition is shown in Fig. 6.

3.5.2 Task

Participants were instructed to input the opening directions of three C-symbols using arrow keys on a keyboard in front of them. In this study, the symbols on the surrounding display were placed horizontally close to the ML (from the perspective of the user; Fig. 4), such that they were as close as possible to the participant's foveal vision.

The three C-symbols were displayed simultaneously for a duration that progressively decreased as the rounds advanced. In the pilot study (n=9), we established that all participants could achieve near 100% accuracy given > 0.5 s of symbol visualization time under any condition in our setup, with prior practice. When the visualization time was < 0.1 s, accuracy sharply declined and varied significantly from participant to participant, approaching the accuracy of the majority classifier. Therefore, the visualization duration for the C-symbols was set to start at 0.5 s, with 0.05 s decrements ending at 0.1 s, with 10 repetitions per visualization time, resulting in a total of 80 repetitions per condition. Participants were instructed to always focus on the *middle symbol*, and not to change their eye focus to the left or right symbols. During the study design we discussed using an eye tracker to verify the participant's eye focus. However, the active shutter glasses prevented us from using screen-based eye tracking technology. Wearing both shutter glasses and eye tracker glasses proved impractical, especially if additional correctional lenses were used. When the C-symbols are not visible, a question mark (?) was displayed in the middle symbol position so that the participants could focus on the accommodation and vergence distances of the current condition. In this study, we captured the symbol response correctness (three binary points per round).

3.5.3 Participants and procedure

We recruited 27 participants (18 males, 9 females) aged between 22 and 39 years, with a mean age of 28.1 years (SD=5.01). All participants were graduate university students, 3 of whom had in-depth knowledge of mixed reality, 6 had some experience, and 18 had no experience. Most participants (26) had previously wore shutter glasses to view a 3D movie and had no problems with depth-effect perception. We verified good visual acuity (with correction, if necessary) in 26 participants; one participant was excluded because of astigmatism.

This study consisted of three phases. In the first phase, participants were asked to fill out a demographic information questionnaire, and their eye acuity was discussed focusing on: any history of eye conditions, what is their prescription of glasses or contact lenses, and are they currently wearing corrections. Next, they were asked to sit on a chair in front of the experimental setup, and the intention of the study was explained. The participants were assured of three breaks between the next four rounds. After further clarification when requested, the room was dimmed.

In the second phase, the participants were asked to put on the stereo shutter glasses and put their chin on the head mount, with their forehead touching the top of the mount (see Fig. 4). The examiner then verified their ability to see stereoscopic depth by displaying a Landholt C-symbol on the ML display with zero disparity, and asked the participant to indicate how far the symbol appeared from them. The examiner then increased the disparity to simulate the far condition, and repeated the questions accordingly. The participants were familiarized with their tasks (section 3.5.2) followed by a practice round. During practice, the symbols were visible for 0.5 s and repeated 20 times, taking an average of 4 min.

The third (main) phase was repeated four times, once for each vergence and accommodation condition. First, the examiner sets up the ML display and the stereoscopic rendering according to the current conditions. After confirming that the participant was ready, a three-second countdown was displayed on the surrounding display, after which three C-symbols appeared in random configurations. After the viewing period elapsed, the symbols were substituted with question marks, prompting participants to input the directional openings of the three C-symbols using the arrow keys. This was repeated ten times, after which the viewing time period was decreased by one step. When it reached 0.1 s, the text “Finished” was displayed, and the condition round ended. Following each condition, participants were asked to take off their glasses, step outside the dimmed room to rest their eyes for 5 min, and complete the NASA TLX questionnaire. They were also asked to grade the condition between 1 and 4, based on how well they thought they performed, where 1 = worst and 4 = best.

4 RESULTS

4.1 Overall performance

Study A

In Fig. 7, the average response time per condition of vergence and accommodation is visualized for only rounds in which all three symbols are correctly answered. We applied a two-way–repeated-measures analysis of variance (ANOVA), which revealed no statistically significant interaction between eye vergence and accommodation ($F_{1,27} = 0.122$, $p = 0.730$). A simple main effects analysis showed that accommodation had a significant effect on response time ($F_{1,27} = 8.711$, $p < .01$), whereas eye vergence did not ($F_{1,27} = 1.768$, $p = 0.195$). The average round response time in the conditions in which the eyes had to re-accommodate when shifting gaze to and from the ML display (accommodation distance = NEAR) was 1.510 s (SD = 0.258), and when they did not re-accommodate (accommodation distance = FAR), the average response time was 1.431 s (SD = 0.203).

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

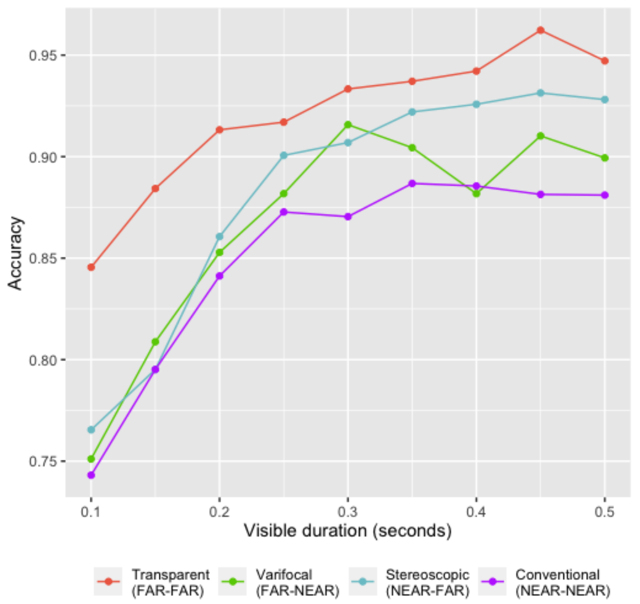

Figure 8: Study B accuracy results per display type for different visibility durations. The legend shows the vergence and accommodation distances for each display type; for example "Stereoscopic (NEAR–FAR)" is a display condition where accommodation is set to the near plane and vergence is set to the far plane. The vertical axis shows the accuracy of discerning the opening of C-symbols with a visible duration in seconds on the horizontal axis. As the visible duration decreased, the accuracy decreased as well.

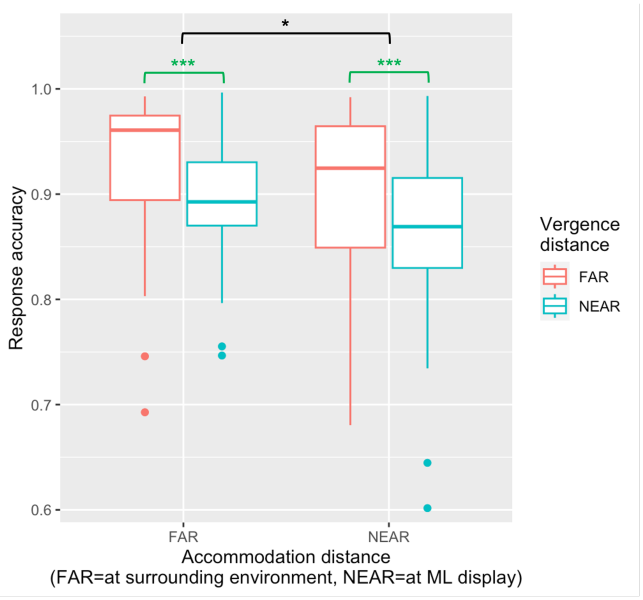

Figure 9: Study B performance results. It shows the average response accuracy when fixing the gaze on a symbol on the (near plane) ML display and symbols on the surrounding (far plane) under different conditions of vergence and accommodation distances of the ML. Altering the vergence distance of the ML display has a significant effect ($p < 0.001$, green) on the response accuracy, and so does altering the accommodation distance ($p < 0.05$, black); however, it has a smaller effect.

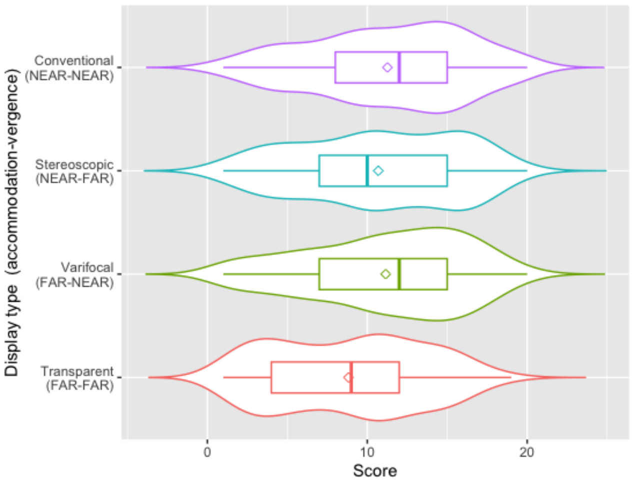

Figure 10: Results from the NASA TLX questionnaire on task load per condition averaged over all subscales for Study B. The outer edge of the violin plot shows the relative number of responses for each score.

Table 1: Subjective grading (1 to 4) of performance per condition and display type for Study B

Display type	Condition		Mean grade	SD
	Accommodation	Vergence		
Transparent	FAR	FAR	3.58	0.86
Varifocal	FAR	NEAR	2.15	0.92
Stereoscopic	NEAR	FAR	2.62	0.85
Conventional	NEAR	NEAR	1.65	0.89

Study B

Fig. 8 shows the proportion of correct responses per visibility duration in seconds, for all four combinations of conditions. In Fig. 9 this result is averaged over all visibility periods. Owing to the non-normal distribution of accuracy data, we applied an Aligned Rank Transform followed by a two-way–repeated-measures ANOVA. The results showed that changing the eye vergence led to a statistically significant difference in accuracy ($F_{1,25} = 35.801$, $p < .001$). Moreover, changing the eye accommodation distance also led to a statistically significant difference in accuracy ($F_{1,25} = 6.10$, $p < .05$), but there was no significant interaction between vergence and accommodation.

4.2 Task load and questionnaire

The NASA TLX questionnaire scores were averaged, as shown in Fig. 10. The display types had similar mean scores, with the transparent display having the lowest perceived task load (M = 8.82, SD = 4.73). The conventional (M = 11.26, SD = 4.87) and varifocal (M = 11.15, SD = 4.91) displays had similar higher values, whereas the stereoscopic display had a value in the middle (M = 10.69, SD = 5.01). We applied a two-way ANOVA, which showed that eye vergence had a significant impact on task load ($F_{1,99} = 7.396$, $p < .01$) whereas accommodation only showed weak evidence of influencing task load ($F_{1,99} = 2.973$, $p < .1$). The analysis revealed no significant interaction between vergence and accommodation.

Table 1 presents the participants' subjective grading of the four conditions, between 4 (best) and 1 (worst). The subjective grade was the highest (M = 3.58, SD = 0.86) for the transparent display and the lowest for the conventional display (M = 1.65, SD = 0.89).

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2024.3403261

5 DISCUSSION

To interact with the (augmented) content on the ML display and its surroundings, we investigated two strategies: (1) rapid gaze shifting between the ML display and its surroundings and (2) fixed gaze on the ML display while viewing the display and its surroundings in a single context. To better understand how different conditions of vergence and accommodation distances affect user performance, we conducted two visual-acuity experiments, which are discussed hereafter.

5.1 Interacting with AR display and its surroundings by rapid gaze shifting

In our first hypothesis (H1), we predicted that when rapidly shifting the gaze, reducing the eye accommodation distance of the AR ML display in relation to its surroundings would result in a faster interaction between the AR content and its surroundings, more so than reducing the vergence distance. The results of Study A support this hypothesis: When the accommodation distance between the display and its surroundings was reduced, a significant reduction in symbol reading time was observed (on average 0.077 s). This decrease in reading time demonstrates a faster interaction between the AR content and its surroundings. However, there was no significant impact when the difference in the vergence distances was reduced. This is in line with the literature [26], where the accommodation response was found to be substantially slower than the vergence response when gaze fixation moved from near-to-far or far-to-near. This explains why accommodation was the prevailing human factor in this visual-acuity experiment. We furthermore observe that, in line with prior studies, focal distance switching caused reduced performance [2, 17] and switching both visual distances had the worst performance [16].

However, it is still interesting to note that when the eyes did not have to re-accommodate when shifting gaze, decreasing the vergence distance did not produce a significantly faster response time (Fig. 7: FAR-accommodation distance plots). It is possible that our experimental design did not have a sufficiently high resolution to capture small effects. Nevertheless, these results show that, in AR support tasks that require rapid shifting of gaze fixation between a display and its surroundings, matching the accommodation distance of the augmented content with that of its surroundings accelerates the task. For instance, a varifocal display (third image in Fig. 1) can help read augmented instructions on a surgery support display, where a surgeon often and rapidly shifts between AR instructions and its surroundings. While this type of display might induce discomfort owing to the vergence–accommodation conflict, we did not find evidence of its significant impact on cognitive-task load or task performance (i.e., no statistical analysis showed an interaction effect between vergence and accommodation). This suggests that the vergence-accommodation conflict is not a prevalent human factor in such scenarios.

5.2 Interacting with AR and its surroundings by viewing both as a single context

In our second hypothesis (H2), we predicted that when the gaze is fixed on a near-ML display, reducing eye vergence and accommodation distance in relation to its surroundings will result in a more accurate merging of AR content and its surroundings. This was based on the assumption that the resulting blurring and double vision from disparate vergence and accommodation, respectively, hinder accurate detection of the surroundings. The results of Study B provide evidence to support this hypothesis. The eye-acuity symbol identification accuracy was the highest when both eye vergence and accommodation distances were similar between the *near display* (ML display) and the far surroundings. For example, accuracy is best for a transparent display (Fig. 9, FAR-FAR condition), and worst when both distance differences are largest, as is the case when using a conventional ML display (NEAR-NEAR condition). This effect holds true over short and relatively long durations of an interaction switching task and increases as the task time shortens (the red line is far above other lines in Fig. 8 for short durations).

When the ML display acted like a stereoscopic display, rendering content at a vergence distance close to the surrounding distance (Fig. 9: accommodation=NEAR and vergence=FAR), also resulted in an improvement in accuracy over the conventional ML display. This suggests that simply using a display with stereoscopic capabilities and rendering the content close to the detected surroundings would allow for a more accurate user experience when the task involved requires the viewer to see both the surroundings and AR ML content simultaneously. Rendering content at a distance where eye accommodation is close to the surroundings also improves accuracy (FAR–NEAR), but to a lesser degree than vergence. This advantage of stereoscopic displays is beneficial because varifocal displays, which can change accommodation distances or have a specific set of focal planes, are less common than stereoscopic displays and often require large optics. Furthermore, stereoscopic displays are relatively affordable in contrast to varifocal displays. The results also allowed us to conclude that double vision was the prevalent human factor compared to blurred surroundings in our scenario, which was the case when users did not accommodate or verge at the correct distance.

5.3 Vergence-accommodation conflict

We also hypothesized (H3) that an AR ML display with mismatched vergence and accommodation distances would require higher physical and mental demands and would result in a less accurate interaction with AR content and its surroundings. Our results did not indicate a significant difference in the task load when vergence and accommodation were mismatched. On the varifocal and stereoscopic displays, the aggregated task load score (Fig. 10) was only slightly higher than the score for a transparent display where the conditions matched (FAR–FAR), and there was no significant difference from the conventional display (NEAR–NEAR). We observed an overall lower task load score in the transparent condition, as expected. The eye vergence and accommodation distances were identical for the *near* display and matched the surroundings. However, the scores were extremely dispersed under all conditions. This high standard deviation indicates that our method of measuring eye comfort is less predictable and difficult to generalize. However, a larger dispersion is expected in subjective data, and additional analysis is necessary. However, our results indicate that mismatching vergence and accommodation distance do not affect the cognitive-task load when users are required to merge their surroundings with an ML display. Furthermore, as already mentioned, no interaction effect of vergence and accommodation was detected in Studies A or B, again suggesting that vergence–accommodation conflict was not the key to accurate interaction with AR content and its surroundings.

5.4 User preference

Finally, when asked in Study B, under which condition users preferred to obtain the most correct answers, participants ranked the transparent display condition as the highest (Table 1). This was again expected, as the eye strain was the lowest, and the measured accuracy was also the highest under this condition. The second-most preferred display was stereoscopic, as in the NEAR–FAR condition. From the questionnaire responses, it seems that the participants had fewer problems with content being out of focus (as was the case in the nonmatching accommodation distance condition) than with double vision. This binocular diplopia occurs when the eyes converge in front of or behind a focal plane, as is the case under nonmatching vergence conditions. This further supports our recommendation for utilizing a stereoscopic ML display with a matching vergence distance, because when a strategy of fixing the gaze on the ML display is employed, the AR content on the display can be merged faster and

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

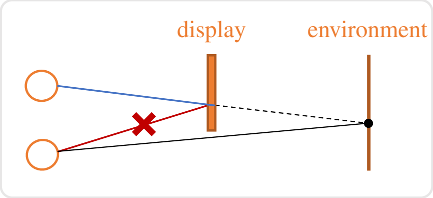

Figure 11: Diplopia (double vision) problem in ML displays. In edge-cases, the surrounding environment is visible for one eye (black) but blocked by the display for the other eye (blue). Owing to stereoscopic disparity, focusing on the farther surroundings causes double vision (blue and red) of the closer display. Blackening the image of the display for one eye (red) or rendering it transparent may alleviate this problem.

more accurately with its surroundings.

6 LIMITATIONS AND FUTURE WORK

One limitation of our studies was the dependence on the size of the symbols and the accuracy of discerning them. The symbol sizes were determined in a pilot study (section 3.2.3) and remained consistent throughout the experiments. However, it remains uncertain whether the accuracy is still affected by minimizing the eye accommodation and vergence distances when using larger-sized symbols. Furthermore, it is possible that visual tasks with many details, such as reading small texts, benefit more from minimizing the accommodation discrepancy, whereas depth-heavy tasks benefit more from minimizing the vergence discrepancy. This relationship should be investigated further in future studies.

Another limitation of our study design is the cognitive load of discriminating a symbol and matching it with the correct input key. It is easier for a user to input their answers when the eye acuity symbols are all equal or when two symbols (orientations) match. In future work we hope to separate this difficulty factor.

As highlighted in the Discussion section, stereoscopic disparity also causes diplopia, that is, double vision of objects or environments on which the eyes are not verging. In cases where the focus point in the surroundings is visible to one eye but occluded by the display for the other (Fig. 11), diplopia may hinder the ability to merge the two views. In future work, we plan to investigate the effects of creating a truly monoscopic ML (e.g., blackening the image on the ML for one eye or only for the border-case eye, as shown in Fig. 11, red line). It would be interesting to verify whether rendering content from a user's perspective has any impact on the performance shown in this study.

Vergence and accommodation are strong depth-cue providers. In this study, we did not focus on depth perception. Our results suggest that decreasing the difference between accommodation or vergence distances as displayed on an AR display and real surroundings will improve a viewer's task performance, but does not take into account the effect that this has on depth perception. Future work will need to verify the trade-offs between merging performance and depth-cue quality, based on the type of action performed.

Finally, although participants were given task instructions for shifting gaze (Study A) or focusing on a single point (Study B), we used limited methods available to verify that these instructions were strictly followed. Although we did not find an uneven distribution of accuracy over the three symbol locations, employing eye tracking in future studies would be beneficial.

7 CONCLUSION

This study investigated eye vergence and accommodation distances in a typical scenario where users interacted with the content on a ML display at arm's length, as well as with its surroundings, in a short timeframe. We discussed the issues posed by these visual processes in contemporary mixed-reality displays and highlighted the lack of related materials concerning ML displays.

Two fundamental visual-acuity user studies were conducted in which both visual processes were compared by changing the display distances between near (arm's length) and far (1 m, similar to an office desk environment). In the first study, users interacted with the content on a near-ML display and its surroundings by rapidly shifting their gaze. We found that eye accommodation was bottlenecked and that reducing the distance of accommodation decreased the time needed to identify eye-acuity symbols. In the second user study, users focused on the ML display to view it and its surroundings as a merged, single context. We found that minimizing the eye's vergence distance discrepancy helps users most in accurately identifying eye-acuity symbols. Additionally, minimizing accommodation distance had a positive effect; however, the extent was less than that in the first study's results. These results coincided with the participants' subjective task performance and preferences. Thus, in a situation where the user has to frequently or rapidly compare content on a *near display*, such as a handheld device with AR support, with the physical surroundings, it is beneficial to reduce or equalize the stereoscopic vergence distance of that content, as well as its accommodation distance, in relation to the surroundings. Furthermore, if there are negative effects resulting from conflicting vergence–accommodation in a ML setup, they have no significant impact on the cognitive-task load, nor are they detected as key to accurate interaction with the AR ML and its surroundings within a short timeframe.

REFERENCES

[1] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks. A stereo display prototype with multiple focal distances. *ACM transactions on graphics (TOG)*, 23(3):804–813, 2004.

[2] M. S. Arefin, N. Phillips, A. Plopski, J. L. Gabbard, and J. E. Swan. The effect of context switching, focal switching distance, binocular and monocular viewing, and transient focal blur on human performance in optical see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2014–2025, 2022.

[3] Y. Bababekova, M. Rosenfield, J. E. Hue, and R. R. Huang. Font size and viewing distance of handheld smart phones. *Optometry and Vision Science*, 88(7):795–797, 2011.

[4] J. L. Bacca Acosta, S. M. Baldiris Navarro, R. Fabregat Gesa, S. Graf, et al. Augmented reality trends in education: a systematic review of research and applications. *Journal of Educational Technology and Society, 2014, vol. 17, núm. 4, p. 133-149*, 2014.

[5] D. Baričević, C. Lee, M. Turk, T. Höllerer, and D. A. Bowman. A hand-held ar magic lens with user-perspective rendering. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 197–206, 2012. doi: 10.1109/ISMAR.2012.6402557

[6] M. Berning, D. Kleinert, T. Riedel, and M. Beigl. A study of depth perception in hand-held augmented reality using autostereoscopic displays. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 93–98. IEEE, 2014.

[7] V. Biener, S. Kalamkar, N. Nouri, E. Ofek, M. Pahud, J. J. Dudley, J. Hu, P. O. Kristensson, M. Weerasinghe, K. Čopič Pucihar, M. Kljun, S. Streuber, and J. Grubert. Quantifying the effects of working in vr for one week. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3810–3820, 2022. doi: 10.1109/TVCG.2022.3203103

[8] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose. Toolglass and magic lenses: the see-through interface. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pp. 73–80, 1993.

[9] O. Bimber, G. Wetzstein, A. Emmerling, and C. Nitschke. Enabling view-dependent stereoscopic projection in real environments. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*, pp. 14–23. IEEE, 2005.

[10] O. Cakmakci and J. Rolland. Head-worn displays: a review. *Journal of display technology*, 2(3):199–216, 2006.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

[11] K. Čopič Pucihar, P. Coulton, and J. Alexander. Evaluating dual-view perceptual issues in handheld augmented reality: device vs. user perspective rendering. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pp. 381–388, 2013.

[12] K. Čopič Pucihar, P. Coulton, and J. Alexander. The use of surrounding visual context in handheld ar: device vs. user perspective rendering. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 197–206, 2014.

[13] P. T. de Jong. A history of visual acuity testing and optotypes. *Eye*, pp. 1–12, 2022.

[14] D. Drascic and P. Milgram. Perceptual issues in augmented reality. In *Stereoscopic displays and virtual reality systems III*, vol. 2653, pp. 123–134. Spie, 1996.

[15] M. Dunleavy, C. Dede, and R. Mitchell. Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of science Education and Technology*, 18:7–22, 2009.

[16] A. Eiberger, P. O. Kristensson, S. Mayr, M. Kranz, and J. Grubert. Effects of depth layer switching between an optical see-through head-mounted display and a body-proximate display. In *Symposium on Spatial User Interaction*, pp. 1–9, 2019.

[17] J. L. Gabbard, D. G. Mehra, and J. E. Swan. Effects of ar display context switching and focal distance switching on human performance. *IEEE transactions on visualization and computer graphics*, 25(6):2228–2241, 2018.

[18] S. Gowrisankaran and J. E. Sheedy. Computer vision syndrome: A review. *Work*, 52(2):303–314, 2015.

[19] J. Guo, D. Weng, H. B.-L. Duh, Y. Liu, and Y. Wang. Effects of using hmds on visual fatigue in virtual environments. In *2017 IEEE Virtual Reality (VR)*, pp. 249–250. IEEE, 2017.

[20] G. Heron, W. N. Charman, C. M. Schor, et al. Age changes in the interactions between the accommodation and vergence systems. *Optometry and vision science*, 78(10):754–762, 2001.

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

[22] B. A. Holden, T. R. Fricke, D. A. Wilson, M. Jong, K. S. Naidoo, P. Sankaridurg, T. Y. Wong, T. J. Naduvilath, and S. Resnikoff. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology*, 123(5):1036–1042, 2016.

[23] F.-C. Huang, D. P. Luebke, and G. Wetzstein. The light field stereoscope. In *SIGGRAPH emerging technologies*, pp. 24–1, 2015.

[24] S. Jaiswal, L. Asper, J. Long, A. Lee, K. Harrison, and B. Golebiowski. Ocular and visual discomfort associated with smartphones, tablets and computers: what we do and do not know. *Clinical and Experimental Optometry*, 102(5):463–477, 2019.

[25] K. Kaur, B. Gurnani, S. Nayak, N. Deori, S. Kaur, J. Jethani, D. Singh, S. Agarkar, J. R. Hussaindeen, J. Sukhija, et al. Digital eye strain-a comprehensive review. *Ophthalmology and therapy*, 11(5):1655–1680, 2022.

[26] A. Khurana, A. Khurana, and B. Khurana. *Theory and Practice of Optics and Refraction*. Modern system of ophthalmology (MSO) series. Elsevier India, 2017.

[27] H. Kim, X. Wu, J. L. Gabbard, and N. F. Polys. Exploring head-up augmented reality interfaces for crash warning systems. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pp. 224–227, 2013.

[28] N. Kim, A.-H. Phan, M.-U. Erdenebat, A. Alam, K.-C. Kwon, M.-L. Piao, and J.-H. Lee. 3d display technology. *Disp. Imag*, 1(1):73–95, 2014.

[29] G. A. Koulieris, K. Akşit, M. Stengel, R. K. Mantiuk, K. Mania, and C. Richardt. Near-eye display and tracking technologies for virtual and augmented reality. *Computer Graphics Forum*, 38(2):493–519, 2019.

[30] G.-A. Koulieris, B. Bui, M. S. Banks, and G. Drettakis. Accommodation and comfort in head-mounted displays. *ACM Transactions on Graphics (TOG)*, 36(4):1–11, 2017.

[31] G. Kramida. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE transactions on visualization and computer graphics*, 22(7):1912–1931, 2015.

[32] D. M. Krum, E. A. Suma, and M. Bolas. Augmented reality using personal projection and retroreflection. *Personal and Ubiquitous Computing*, 16:17–26, 2012.

[33] M. Lambooij, W. IJsselsteijn, M. Fortuin, I. Heynderickx, et al. Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of imaging science and technology*, 53(3):30201–1, 2009.

[34] D. Lanman and D. Luebke. Near-eye light field displays. *ACM transactions on graphics (TOG)*, 32(6):1–10, 2013.

[35] J.-h. Lee, I. Yanusik, Y. Choi, B. Kang, C. Hwang, J. Park, D. Nam, and S. Hong. Automotive augmented reality 3d head-up display based on light-field rendering with eye-tracking. *Optics Express*, 28(20):29788–29804, 2020.

[36] B. Lijka, S. Toor, and G. Arblaster. The impact of diplopia on reading. *The British and Irish Orthoptic Journal*, 15(1):8, 2019.

[37] J. Lin, D. Cheng, C. Yao, and Y. Wang. Retinal projection head-mounted display. *Frontiers of Optoelectronics*, 10:1–8, 2017.

[38] L. Lipton. *Foundations of the stereoscopic cinema: a study in depth*. Van Nostrand Reinhold Company, 1982.

[39] X.-L. Ma, R.-Y. Yuan, L.-B. Zhang, M.-Y. He, H.-L. Zhang, Y. Xing, and Q.-H. Wang. Augmented reality autostereoscopic 3d display based on sparse reflection array. *Optics Communications*, 510:127913, 2022.

[40] J. Maxwell, J. Tong, and C. M. Schor. Short-term adaptation of accommodation, accommodative vergence and disparity vergence facility. *Vision Research*, 62:93–101, 2012. doi: 10.1016/j.visres.2012.03.013

[41] H. Mizushina, J. Nakamura, Y. Takaki, and H. Ando. Super multi-view 3d displays reduce conflict between accommodative and vergence responses. *Journal of the Society for Information Display*, 24(12):747–756, 2016.

[42] P. Mohr, M. Tatzgern, J. Grubert, D. Schmalstieg, and D. Kalkofen. Adaptive user perspective rendering for handheld augmented reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 176–181, 2017. doi: 10.1109/3DUI.2017.7893336

[43] R. L. Newman. *Head-up displays: Designing the way ahead*. Routledge, 2017.

[44] A. Olwal, C. Lindfors, J. Gustafsson, T. Kjellberg, and L. Mattsson. Astor: An autostereoscopic optical see-through augmented reality system. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*, pp. 24–27. IEEE, 2005.

[45] B. Rassow and Y. Wang. Correlation of letter optotypes with landholt ring for different degrees of visual acuity. *Klinische Monatsblatter fur Augenheilkunde*, 215(2):119–126, 1999.

[46] A. Rastogi. Design of an interface for teleoperation in unstructured environments using augmented reality displays. University of Toronto, 1996. Master's Thesis.

[47] D. Schmalstieg and T. Hollerer. *Augmented reality: principles and practice*. Addison-Wesley Professional, 2016.

[48] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. Visual discomfort with stereo displays: effects of viewing distance and direction of vergence-accommodation conflict. In *Stereoscopic Displays and Applications XXII*, vol. 7863, pp. 222–230. SPIE, 2011.

[49] R. Swannack and O. D. A. Prima. An assessment of human depth understanding in handheld light-field displays.

[50] H. Urey, K. V. Chellappan, E. Erden, and P. Surman. State of the art in stereoscopic and autostereoscopic displays. *Proceedings of the IEEE*, 99(4):540–555, 2011.

[51] G. Westheimer. The maxwellian view. *Vision research*, 6(11-12):669–682, 1966.

[52] S. Yano, S. Ide, T. Mitsuhashi, and H. Thwaites. A study of visual fatigue and visual comfort for 3d hdtv/hdtv images. *Displays*, 23(4):191–201, 2002.

[53] R. Zabels, K. Osmanis, M. Narels, U. Gertners, A. Ozols, K. Rūtenbergs, and I. Osmanis. Ar displays: Next-generation technologies to solve the vergence–accommodation conflict. *Applied Sciences*, 9(15):3147, 2019.

[54] Y. Zhou, J. Zhang, and F. Fang. Vergence-accommodation conflict in optical see-through display: Review and prospect. *Results in Optics*, 5:100160, 2021.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/