International Journal of Multimedia Data Engineering and Management
Volume 14 • Issue 1

Comparison of Light Field and Conventional Near-Eye AR Displays in Virtual-Real Integration Efficiency

Wei-An Teng, National Taiwan University, Taiwan
Su-Ling Yeh, National Taiwan University, Taiwan
Homer H. Chen, National Taiwan University, Taiwan*
https://orcid.org/0000-0002-8795-1911

ABSTRACT

Most existing wearable displays for augmented reality (AR) have only one fixed focal plane and hence can easily suffer from vergence-accommodation conflict (VAC). In contrast, light field displays allow users to focus at any depth free of VAC. This paper presents a series of text-based visual search tasks to systematically and quantitatively compare a near-eye light field AR display with a conventional AR display, specifically in regards to how participants wearing such displays would perform on a virtual-real integration task. Task performance is evaluated by task completion rate and accuracy. The results show that the light field AR glasses lead to significantly higher user performance than the conventional AR glasses. In addition, 80% of the participants prefer the light field AR glasses over the conventional AR glasses for visual comfort.

KEYWORDS

3D Display, Augmented Reality, Mixed Reality, Near-Eye Display, Psychophysics, Psychovisual, Vergence-Accommodation Conflict, Visual Search

INTRODUCTION

Existing augmented and virtual reality (AR/VR) devices often suffer from vergence-accommodation conflicts (VAC). Accommodation refers to the adjustment of focal length of a human eye to obtain a clear image of an object, whereas vergence refers to the simultaneous movement of both eyes towards or away from one another in order to align the two retinal images of an object on corresponding retinal points of interest. Consistent accommodation and vergence cues received by the brain can guide the eyes to properly focus on the object. A vergence-accommodation conflict (VAC) occurs when the brain receives mismatching vergence and accommodation cues (Figure 2), which can cause binocular fusion difficulty, visual fatigue, and dizziness to users of augmented and virtual reality (AR/VR). The conflict is more pronounced when the virtual object is closer to the eyes.

Conventional augmented reality (AR) optical see-through displays render stereoscopic images on 2D image planes. These displays are notable for lacking the ability to provide correct focus cues for

DOI: 10.4018/IJMDEM.333609 *Corresponding Author

Figure 1. We present a series of text-based visual search tasks to systematically and quantitatively test human visual performance under the influence of vergence-accommodation conflict (VAC). From left to right: experimental setup, example visual content, and experimental results. Each visual content example includes two side-by-side text blocks. The left text block is displayed by a smartphone, and the right text block by a pair of AR glasses.

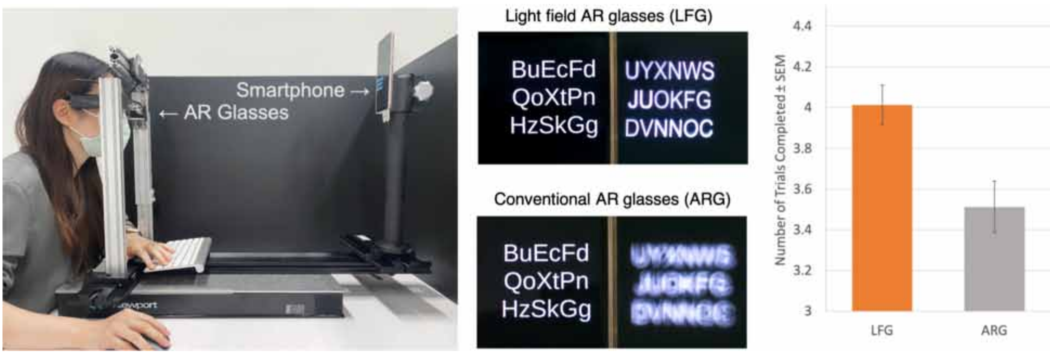
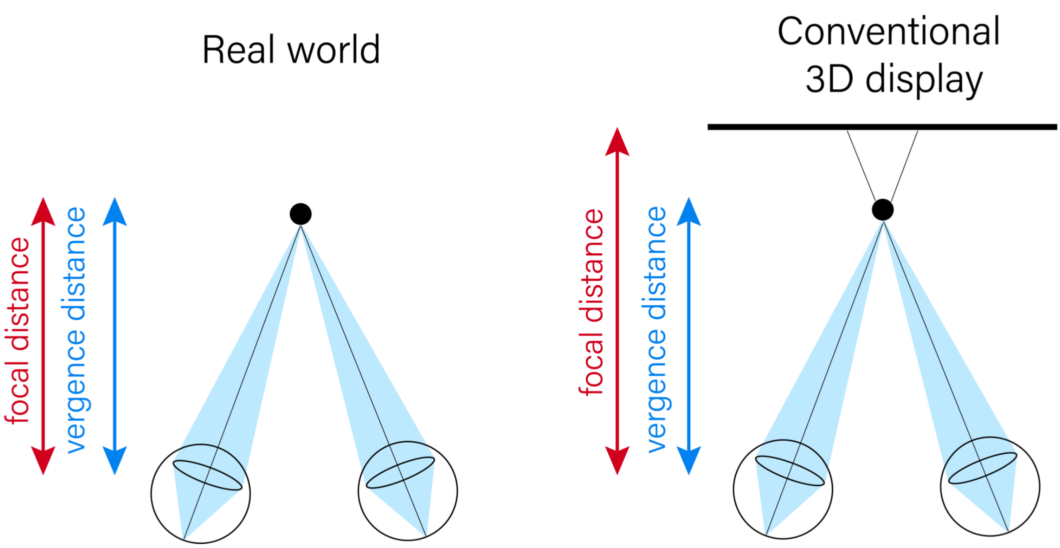


Figure 2. Illustration of the vergence-accommodation conflict. The eyes accommodate and converge at the same distance when watching a real object. But this is not the case when using a conventional 3D AR display.



both accommodation and convergence. Unlike traditional stereoscopic techniques, light field displays, which are VAC-free, work by generating a light field pertaining to a virtual object and projecting it into the user's eyes in the same way as the light field of a real object would travel to the eyes. A light field is a collection of light rays traveling from every point in the space of interest through every possible path (Levoy & Hanrahan, 1996). Since the light field display has continuous focal planes, the user can perceive a clear image of the virtual object at any depth.

The effect of VAC on users has been investigated for conventional AR glasses. While most of the work focuses on the physiological discomfort of users (Hoffman et al., 2008; Shibata et al., 2011; Zou et al., 2015), only a limited amount of work has examined the association between VAC and user visual performance (Gabbard et al., 2019; Daniel & Kapoula, 2019). Furthermore, light field AR display has never been considered in these studies. This motivates us to carry out the research further.

In this work, we design a psychovisual experiment to systematically and quantitatively compare the visual performance of participants wearing a pair of light field AR glasses with that of the same participants wearing a pair of conventional AR glasses. To the best of our knowledge, our work is among the first that systematically compare light field AR glasses (VAC-free) with conventional AR glasses (VAC-inherent). We want to find out whether there is a significant difference in user performance between light field AR glasses and conventional AR glasses under the same viewing condition. The results obtained in this experiment lays the foundation for future evaluation of near-eye light field displays.

BACKGROUND

This section reviews the occurrence of VAC, the advanced architecture of conventional 3D displays, light field, and related works based on visual search tasks.

Vergence-Accommodation Conflict

Although AR has received considerable attention from academia and industry over the past decade, most commercially available AR glasses and head-mounted displays have optical focal planes fixed at a certain depth and thereby introduce a mismatch between accommodation and vergence cues. The impact of VAC on conventional AR displays can be divided into three aspects (Hua & Javidi, 2014). First, the focus cues between the virtual object and the real 3D object do not match. Since most conventional AR displays have virtual images formed on a 2D plane at a fixed distance from the user, the eyes of the user are cued to accommodate at the 2D image plane of the display when watching the virtual object. As a result, the user is forced to switch focus frequently between the real 3D object and the 2D image plane in order to integrate real and virtual visual information. The second aspect of VAC is the mismatch between the depth of the 2D image plane and the depth of the virtual object. In contrast to the first aspect, which relates to the usage scenario of virtual-real integration, the second aspect relates to the mismatched distance for accommodation and vergence rendered by the display device. When watching the displayed stereoscopic images, the eyes are cued to converge at the rendered depth and fuse the binocular view, while in the meantime the eyes are forced to accommodate at the source image plane of the device. This condition is different from viewing the real-world scene where accommodation and vergence cues are mostly coupled. The last aspect of VAC relates to the fact that conventional AR devices typically take “all-in-focus” images as input. The synthetic objects at different depths all appear sharp in such images. As a result, the users perceive either all-in-focus or blur images from conventional AR devices. The three aspects of VAC bring the users an uncomfortable viewing experience.

VAC often leads to physiological discomfort or reduction of cognitive and visual performance (Hoffman et al., 2008; Daniel & Kapoula, 2019; Edgar et al., 1994; Yano et al., 2002; Watt et al., 2005). The visual fatigue and physiological discomfort due to VAC have been studied extensively. The incurred common symptoms include eye strain and headaches (Hoffman et al., 2008; Yano et al., 2002; Shibata et al., 2011; Zou et al., 2015). Additionally, studies have shown that VAC impacts visual and cognitive performance by causing confusion to shape and depth perception (Hoffman et al., 2008; Watt et al., 2005). A speed reduction in binocular fusion and cognitive performance has also been observed (Hoffman et al., 2008; Daniel & Kapoula, 2019). These studies show that VAC is a fundamental problem for users of AR and VR displays.

Conventional 3D Display

The conventional 3D display technology typically presents stereoscopic images to both eyes of a user, with the images displayed at a fixed optical focal distance. While such displays can provide 3D sensation, the VAC problem can negatively affect the visual experience of the user. Many advanced 3D architecture designs have been proposed to mitigate the VAC problem for conventional 3D displays. Among them, multi-focal-plane display, extended depth of field display, and vari-focal-plane display are three major ones.

The multi-focal-plane display is a well-known means of addressing the VAC issue (Koetting, 1970; Akeley et al., 2004; Wu et al., 2016). As its name suggests, multi-focal-plane display offers more than one focal plane so that virtual objects at different depths can be properly displayed. The associated algorithm normally divides a 3D scene into a number of depths in accordance with the positions of the multi-focal-planes. Although VAC-related visual discomfort may be alleviated when viewing objects at similar depths on a single focal plane, only a limited number of focal planes can be supported in practice due to cost considerations.

Extended depth of field (EDOF) is another technology that can mitigate the VAC problem. By definition, the depth of field (DOF) is the range of depths that an object can appear in focus without changing accommodation. An EDOF display can be realized using properties of pinhole optics that the DOF can be extended to near infinity when the diameter of a light beam projected into the eye is considerably smaller than the pupil of the eye, analogous to the pinhole camera. Since the DOF is infinite, there is no need for accommodation. The famous Maxwellian view display is one such design (Yuuki et al., 1996). Despite the accommodation-free advantage, EDOF displays suffer from noticeable image quality reduction.

Vari-focal-plane displays vary the optical focal distance dynamically to allow accurate accommodation cues for the eyes (Wann et al., 1995). Low complexity design is the most significant advantage of vari-focal-plane displays. These displays can be implemented by adopting an adjustable lens that moves between the eye and the display panel along the optical axis (Shiwa et al., 1996) or by employing electrically-driven optical elements such as deformable mirrors (Fernández & Artal, 2003), liquid lenses (Kuiper & Hendriks, 2004), and liquid crystal lenses (Ren et al., 2007). An eye-tracker is needed to track the gaze of the eyes so as to adjust the accommodation cues in real-time, which requires extra power consumption and system complexity. Moreover, only the object to be viewed has the correct depth cue, while the other virtual objects not at the same depth are presented with incorrect focus cue.

Although these three advanced architectures for near-eye 3D displays have been proposed or prototyped, most existing commercial 3D displays are still based on conventional stereoscopic principle with left and right images presented to users.

Light Field

The light field has become a popular topic in the past two decades. The term "light field" was coined by Gershun (1936) to describe the radiometric properties of light in a three-dimensional space. There are various ways to parameterize light rays in a light field. For example, the light rays can be represented by a 5D plenoptic function describing the total geometric light distribution (Adelson & Bergen, 1991) or as a 2D slice of a 4D function (Levoy & Hanrahan, 1996).

The development of light field capturing is highly associated with microlens arrays, which capture the 2D spatial and 2D directional information to record the 4D light field. This technique of light field capturing was pioneered and named "integral photography" by Lippmann (1908). The light field can also be generated and displayed. A light field display reproduces how light rays navigate the real world and enter the human eye such that the virtual objects are presented with the correct depth and focus cues. Therefore, unlike traditional stereoscopic techniques, a light field display has continuous focal planes and VAC-free nature.

We have recently received a light field near-eye display prototype from PetaRay that is able to simultaneously display virtual objects at different depths. While the "continuous focal plane" can be verified physically with a camera, we are still interested in the differences between light field AR glasses and conventional AR glasses with regards to user performance and experience.

Visual Search Task and Related Works

Visual search tasks are often executed to evaluate user performance of different types of displays. Smith et al. (2015) designed a text-based visual search task to compare a head-down display with a head-up display regarding user speed, accuracy, and preferences. The study provided a method to quantitatively compare user performance between two different types of displays. Eiberger et al. (2019) discussed the effect of accommodation and vergence change between contents displayed by an optical see-through AR display (virtual) and a body-proximate display (real). The study pointed out the problems of increased reaction time and error rate when integrating virtual and real information through an AR display. However, VAC was not considered since the FOV of the AR display was too narrow to set the vergence distance at the depth of the real content.

The design of our experiment was inspired by the work of Gabbard et al. (2019), who examined the effects of context switching and focal distance switching on the virtual-real integration performance of users wearing optical see-through AR displays. Context switching refers to the act of switching attention between real and virtual information, whereas focal distance switching refers to accommodation changes. Attention encompasses both external and internal attention, distinguishing between paying attention to outside visual objects and paying attention to inner thoughts. The participants had to frequently switch attention between real and virtual visual content since the two types of content were placed at different distances. However, the VAC issue for AR was not considered for two possible reasons. First, the AR display used in the study is monocular and does not involve convergence in the visual perception of virtual content. Second, whether VAC exists for monocular AR remains to be seen despite the fact that accommodation and convergence are coupled such that a change in one drives a change in the other. Although light field displays are considered the ultimate technology for resolving the VAC problem, we have yet to be aware of any work investigating differences in user performance between light field AR displays and conventional AR displays.

EXPERIMENT

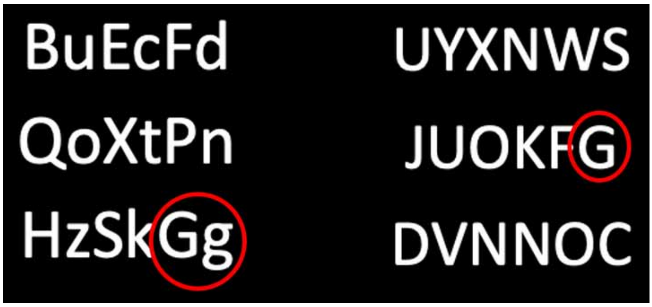
Our experiment was performed to examine the effect of vergence and accommodation cues have on user visual performance. It was designed to compare the visual performance of a participant in two cases: one where the participant wore a pair of light field AR glasses and another where the participant wore a pair of conventional AR glasses. The occurrence and severity of VAC depend on the display principle of these devices. Since the two types of AR glasses have the same eyepiece but differ in how visual information is projected to the retina, the existence of VAC for conventional AR glasses would be the main factor contributing to the differences, if any, in user performance.

We put the real and virtual information together in a visual search to model a usage scenario of virtual-real integration for AR glasses. Since the text-based visual search task adopted in the experiment demands high visual acuity, participants are more likely to accommodate at the actual depth of rendered texts (Iavecchia et al., 1988). This ensures that the participants are able to accommodate at the depth as expected, which is important as it would be difficult to measure the accommodation objectively.

Experimental Design

The participants were required to accomplish a series of tasks that involved the integration of information from both real and virtual worlds. The experimental setup is shown in Figure 1. In a normal, bright indoor room, two identical setups were placed side-by-side on the table, one for light field AR glasses and the other for conventional AR glasses. Each pair of AR glasses was fixed on aluminum extrusion rails, and a smartphone was attached to a rod clamp on an optical support rod mounted on a carrier. The carrier could be manually moved along a short optical rail in the horizontal direction, where the short optical rail was mounted on another optical carrier to allow for manual adjustment in the back-and-forth direction on a long optical rail. Participants were asked to sit in a

Figure 3. Example text blocks presented to participants in the experiment. The left text block contains a pair of identical letters, referred to as the target letter. The right text block contains between zero and three copies of the target letter. The number of times the target letter appears in the right text block is referred to as the target number. In this example, the target letter is G and the target number is 1.



chair to see both the real and virtual worlds through the AR glasses and control the display program through a Bluetooth mouse and a Bluetooth keyboard.

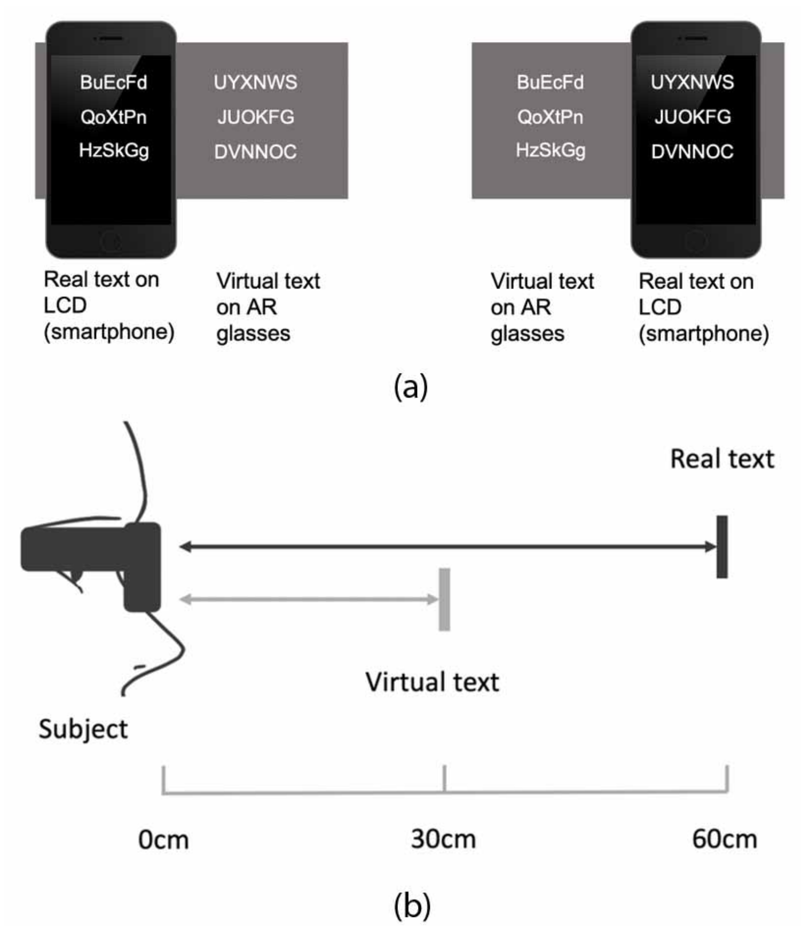
In the experiment, participants were presented with two text blocks (Figure 3) displayed side by side: one by a pair of AR glasses and the other by the smartphone. The left block contained a sequence of upper-case and lower-case letters that appear in alternating order. More precisely, an upper-case letter was followed by a lower-case letter, which in turn was followed by an upper-case letter. The right block contained only a sequence of upper-case letters. Participants were instructed to pick out which letter (termed the "target letter") in the left block appeared twice consecutively, with one in upper case and the other in lower case. Taking the left text block in Figure 3 as an example, the target letter would be G since "Gg" appears consecutively. After the target letter was found, the participants were asked to switch their attention to the right block to count the number (termed the "target number") of times the target letter appeared in the right block. The right block was designed to have zero target letters at the minimum and three target letters at the maximum. Taking the same example in Figure 3, the target number would be 1 since G appears once in the right text block. The process of finding a target letter and its target number constitutes a "trial."

The use of alternating upper- and lower-case letters forces participants to read through the text block carefully. This design precludes the target letter being detected at first glance simply through shape recognition. The letters "I," "J," and "L" were also eliminated from the left text block since the upper- and lower-case versions of these letters resemble one another and would affect the accuracy of the experiment.

Figure 4 shows how the displayed text was arranged in the experiment. In Figure 4 (a), the text displayed on the smartphone is called the real text because the participant could see it directly as a real object through the optical combiner of the AR glasses, whereas the text projected to the eyes of the participant is called the virtual text because it is a virtual object. Each text block was placed at a depth of either 30 cm or 60 cm, as shown in Figure 4 (b). In addition, the real text could be displayed on either the left or right side of the visual field, and the virtual text could be displayed on either a conventional or a light field AR display. Therefore, there were a total of $2\times2\times2=8$ possible display configurations for each pair of AR glasses. Each participant went through the display configurations in a random order to ensure the validity of the experiment. The angular size of the displayed text was fixed at 4.76 degrees throughout the experiment, meaning that the relative size of the text was consistent no matter where it was located.

Figures 5 and 6 show some example display configurations used in the experiment, where the camera was set to focus on the real text in all display configurations. As shown in Figure 5, the virtual text displayed by the light field AR glasses is in focus when it is at the same depth as the real text, and image ghosting appears when it is at a different depth. These figures demonstrate that the light

Figure 4. Illustration of the placement of real and virtual text blocks in the visual field of a participant. (a) We randomly exchanged the location of the real and virtual texts, but the text block on the left always contained alternating upper- and lower-case letters. (b) The real and virtual text blocks appeared at either 30 cm or 60 cm from the participant.



field AR glasses are free of VAC. In contrast, as shown in Figure 6, the virtual text displayed by the conventional AR glasses appears blurry in all display configurations since the conventional AR glasses have a fixed focal distance of 2.5 meters. The depth of the virtual text can only be changed by adjusting the disparity of the stereo image pair.

Before commencing the experiment, we conducted a pre-test to ensure the validity of the experimental design, collect comments from the participants, and record any issues. In some cases, participants could not align the virtual text presented to each eye (binocular fusion difficulty), but could read and accomplish the trials using only one eye. Such cases were considered invalid since the task was concerned with "binocular" vision. Only results generated with successful binocular alignment were retained. During the experiment we also asked participants to verify if they could completely align the virtual texts presented to both eyes. Participants whose eyes diverged in the middle of the experiment were asked to stop the trials until they could align the virtual text again. The timer would not stop in this case. This is important for the experiment because binocular fusion difficulty mostly occurs in concurrence with severe VAC. In addition, we found that some participants exhibited a higher task completion rate when the real text appeared on the left side of the visual field. A possible explanation for this is that while finding the unknown target letter

Figure 5. Display configurations with the virtual text displayed by light field AR glasses. (a) An illustration of the setup. (b) A display configuration with the virtual text (on the left) and the real text (on the right) at 30 cm. (c)-(f) Other display configurations with the real text on the left and the virtual text on the right. Note that when the virtual text is captured by a camera, uneven brightness appears. The uneven brightness is much less pronounced when the virtual text is viewed by human eyes.

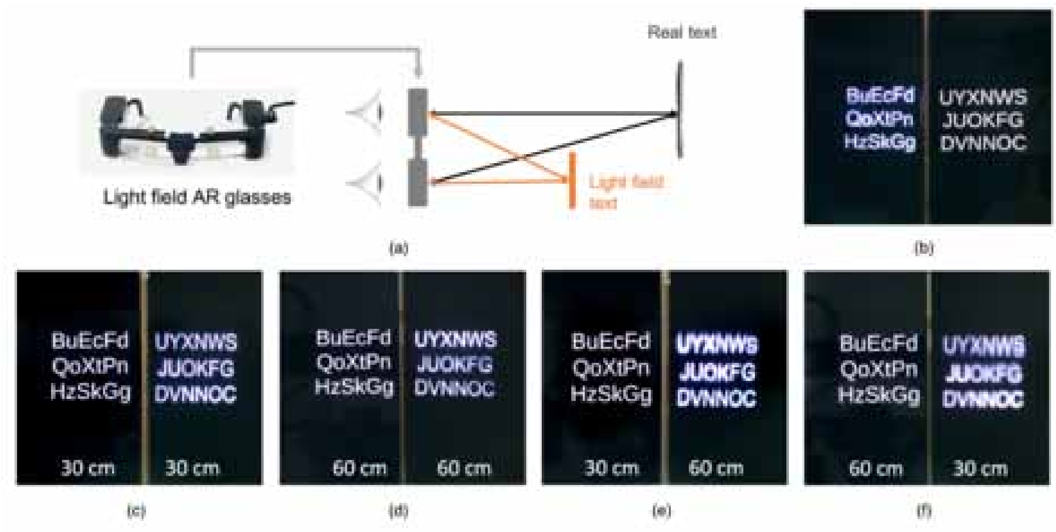
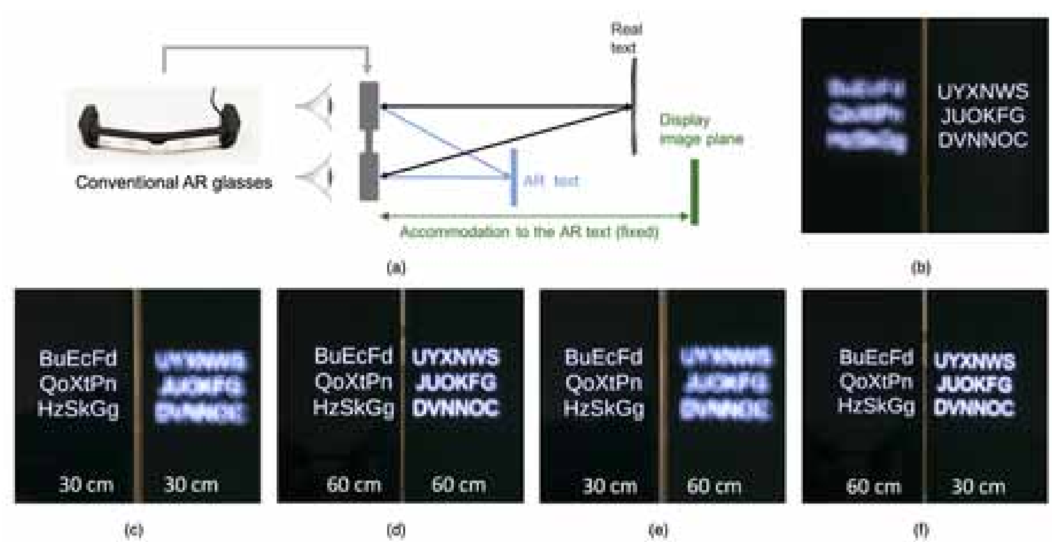


Figure 6. Display configurations with the virtual text displayed by conventional AR glasses. (a) An illustration of the setup. The eyes converge to the depth of the virtual text and accommodate to the image plane of the conventional AR glasses. (b) A display configuration where the virtual text (on the left) and the real text (on the right) at 30 cm. (c)-(f) Other display configurations with the real text on the left and the virtual text on the right.



requires intense concentration, the shape of the target letter is already known when finding the target number; participants can still identify target letters in the right text through shape recognition even if the virtual text appeared blurred. Hence, we counterbalanced the text placement to avoid response bias from affecting the result.

Participants

There was a total of 20 participants (7 females and 13 males) in the experiment. The participants ranged in age from 19 to 24 years old (mean=23) and were recruited from our university. We focused on this age range to ensure the reliability of the results because accommodation amplitude degrades with age (Duane, 1922). Participants were required to have normal or corrected vision between $+0.75$ and -0.75 diopters. Before the experiment started, all participants were tested for their ability to control the vergence of their eyes and perform binocular fusion with both types of AR glasses. Specifically, they were asked to look at a virtual text located at 30 cm, the closest text distance in the experiment. Participants who were unable to fuse the stereo text of the conventional AR glasses were disqualified from participating in the experiment. All the participants were also asked to do the Ocular Surface Disease Index test (Walt et al., 1997) to ensure that they had no eye-related diseases that may have affected their vision.

Apparatus

A pair of light field (LF) AR glasses developed by PetaRay and a pair of BT-30C AR glasses made by Epson were used in the experiment. The former is a binocular near-eye AR display with 960×540 spatial resolution. The latter is a conventional binocular near-eye AR display with 640×720 spatial resolution and 2.5 m optical focal distance.

A Xiaomi Mi8 smartphone was used to display the real text. It had a 6.21-inch screen with a resolution of 1080×2248 pixels.

All custom programs ran on the three display devices (two pairs of AR glasses and one Xiaomi Mi8) were written in C# and driven by the Unity engine. A Bluetooth keyboard and a Bluetooth mouse were used to control the display devices.

Experimental Procedures

Each participant was required to perform the visual search task through the eight different display configurations for each pair of AR glasses. Each participant was given 25 seconds to perform a run consisting of five trials for a display configuration. Once the run was completed or the time was up, all display devices would go blank and enter a 60-second break. The participants were instructed to close their eyes and rest during the break to avoid carrying over any eye strain, and were instructed to prepare for the next run five seconds before the end of each break. In addition, participants were asked to rate their eye strain level using the 5-point Likert scale shown in Table 1 before and after each run (Hoffman et al., 2008).

From the pre-test, we learned that the participant's task completion rate and accuracy of the experiment depend on their experience level. Those who were not familiar with the visual search completed less than 50% of the initial run on average. However, the completion rate improved to an average of 60% after one or more runs of practice. Therefore, we let the participants practice for one random display configuration. This helped participants to learn how to find the target letters and

Table 1. Eye strain level on a five-point Likert scale

1	2	3	4	5
Very Fresh	Fine	Mild Strain	Moderate Strain	Severe Strain

target numbers and how to switch from one trial to the next by simultaneously pressing the return key on the keyboard and clicking the mouse.

Independent Variables

We manipulated three independent variables: the type of glasses (AR or LF), the distance of the virtual text (near or far), and the distance of the real text (near or far).

The existence of VAC is the main difference between the two types of glasses used in this study. The BT-30C AR glasses have a fixed optical focal plane, while the light field AR glasses allow the user to focus at any depth. In the subsequent discussions, the BT-30C AR glasses is referred to as ARG and the light field AR glasses as LFG.

The virtual text was presented by either the LFG or the ARG in the experiment, and the real text was displayed by the Xiaomi 8. The text distance was set to 30 cm (near) or 60 cm (far) for both real and virtual texts. In the following discussions, we use V30, V60, R30, and R60 to denote the near virtual text distance, far virtual text distance, near real text distance, and far real text distance, respectively.

Dependent Variables

We recorded the number of trials completed (ranging from 0 to 5), the frequency of miscounted errors (ranging from 0 to 15), and the eye strain level (ranging from 1 to 5) for each participant in each run of the experiment. An interview with each participant was conducted after the experiment to collect further feedback on eye strain, binocular fusion difficulty, and the overall user experience of the two pairs of AR glasses.

Each participant's task completion score was calculated by the number of trials completed for each display configuration, ranging from 0 to 5. We measured miscounted errors by the number of over- or under-counted target letters in each run. If an incorrect target letter was initially picked from the left text, the miscounted error was set to 3, the maximum possible value of miscounted error in a single trial. The eye strain level was subjectively rated by the participants using the Likert 5-point scale shown in Table 1 before and after each trial, with 1 for no pain and 5 for severe strain.

RESULTS AND DISCUSSION

We hypothesize that the participants performing the visual search tasks would achieve higher virtual-real integration efficiency with the LFG. The virtual-real integration efficiency refers to the task completion scores and accuracy achieved by the participants in the visual search tasks. This section analyzes the dependent variables of the experiments by conducting repeated measures ANOVA, planned comparisons, one-tailed paired *t*-test, and correlation test. The repeated measures ANOVA is a well-known technique that compares the means of one or more variables from repeated observations. Typically, it is used to rate the participants performing at more than two time points. A paired *t*-test is sufficient for only two time points, but for more time points a repeated measures ANOVA is required. Repeated measures reduce the error variance.

Task Completion Score

A three-way repeated measures ANOVA ($2\times2\times2$) is conducted on the number of trials completed with a 95% confidence interval across the four display configurations (V30-R30, V30-R60, V60-R30, and V60-R60 listed in Table 2) of the two pairs of AR glasses. The ANOVA result shows main effects of glasses ($F\left(1{,}19\right) = 22.07, p < .001, \cdot^{2} = .140$) and virtual text distance ($F\left(1{,}19\right) = 29.00, p < .001, \cdot^{2} = .133$), and an interaction between the virtual and the real text distances ($F\left(1{,}19\right) = 4.57,\ p = .046, \cdot^{2} = .018$). In ANOVA, an interaction is defined as when the

Table 2. Display configurations of the two types of AR glasses for data analysis

Virtual Text Distance	Real Text Distance	Type of Glasses and Display Configuration	
V30	R30	AR, V30-R30	LF, V30-R30
	R60	AR, V30-R60	LF, V30-R60
V60	R30	AR, V60-R30	LF, V60-R30
	R60	AR, V60-R60	LF, V60-R60

difference in the means of the response between the levels of one factor is not the same across all levels of another factor.)

As Figure 7 shows, the LFG ($M = 4$, $SD= .391$) leads to better performance than the ARG ($M = 3.513$, $SD = .562$), especially when the virtual text appears at 30 cm. For both pairs of AR glasses, the user scores higher at V60 ($M = 3.994$, $SD = .377$) than at V30 ($M = 3.519$, $SD = .545$). Figure 8 shows the illustration of the interaction between the virtual and the real text distances. The two lines corresponding to different real text distances cross each other in the plot. The existence of line crossing implies that the real text distance does not dominate the participant's performance. Given a fixed virtual text distance, the participants perform better when the real and the virtual texts are at the same distance than when they are at different distances. This result is not surprising because vergence and accommodation cannot change instantly; it takes more time for participants to switch between texts at different distances than texts at the same distance.

To further understand how the virtual and real text distances affect the task completion score for both pairs of AR glasses, we conduct Bonferroni's post hoc test (p-values were adjusted with Bonferroni correction for comparing a family of 6 groups) on the interaction between the virtual and real text distances and show the results in Table 3, where the configuration combinations are divided into six groups. It can be seen that the p-values of Groups 3, 4, and 5 are significant, which indicates that the participants act faster with virtual text at 60 cm (V60) than at 30 cm (V30).

More insights into the post hoc analysis results can be gained by examining the display configurations of each group. The result of Group 3 suggests that, with both real and virtual texts at

Figure 7. Task completion scores for four display configurations for both pairs of AR glasses

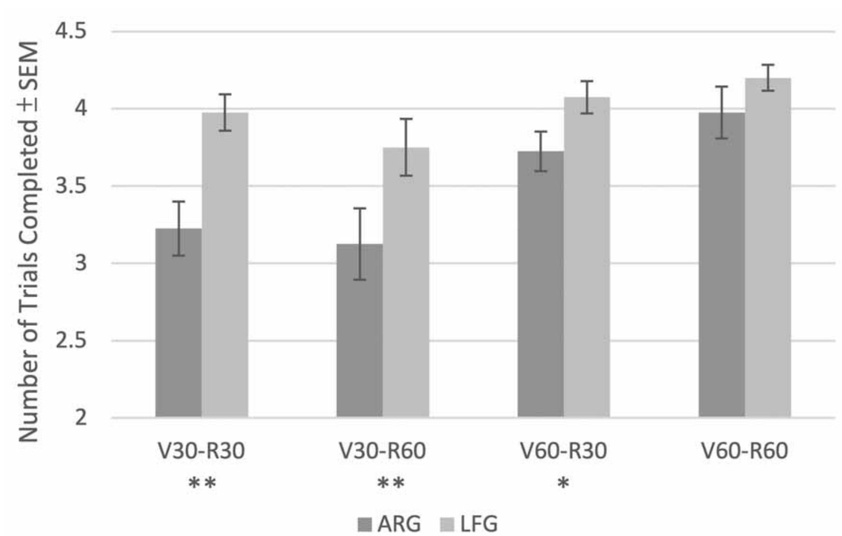


Figure 8. Illustration of the interaction between the virtual and the real text distances

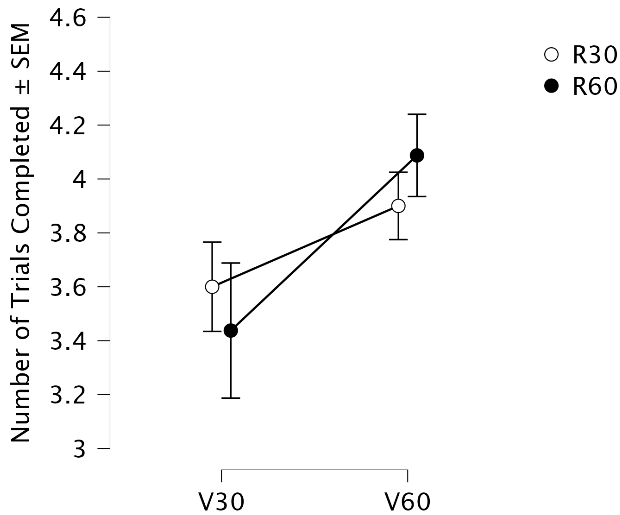


Table 3. Post hoc test of the interaction between virtual and real text distances

Group #	Config. 1	Config. 2	df	*t*-Value	*p*-Value
1	V30-R60	V30-R30	19	−1.357	1.000
2	V60-R30	V30-R30	19	2.494	.103
3	V60-R60	V30-R30	19	3.926	**.002****
4	V60-R30	V30-R60	19	3.725	**.004****
5	V60-R60	V30-R60	19	5.403	**< .001*****
6	V60-R60	V60-R30	19	1.566	.754

$^{**}p < .01$. $^{***}p < .001$.

the same distance, participants are able to complete more trials when both texts are at 60 cm (V60-R60, $M = 4.088$, $SD = .454$) than when both texts are at 30 cm (V30-R30, $M = 3.6$, $SD = .476$). The result of Group 5 suggests that, with the real text fixed at 60 cm (R60), participants get higher scores when the virtual text is placed at the far distance (V60-R60) than at the near distance (V30-R60, $M = 3.438$, $SD = .773$). The results of Groups 3 and 5 are intuitively understandable because the binocular fusion problem is more pronounced for near-distance text than for far-distance text and because the change of vergence needs some time to effect. However, the result of Group 4 is less intuitive. It shows that, with virtual and real texts at different distances, a higher task completion score is achieved with the virtual text at far distance (V60-R30, $M = 3.9$, $SD = .366$), although both display configurations require a change of vergence. We think the result is dominated by the poor performance of the ARG at V30, as we can see from Figure 7 that the number of trials completed by participants for the ARG at V30 is significantly less than that at V60.

To compare the LFG with the ARG in terms of task completion scores for the four display configurations shown in Figure 7, we conduct planned comparisons with one-tailed paired *t*-tests. The test results are shown in Table 4. Among the four display configurations, the LFG has significantly higher task completion scores than the ARG in three of them (V30-R30, $t(19) = 3.572$, $p = .001$, V30-R60, $t(19) = 2.663$,

Table 4. Results of the planned comparisons between the LFG and the ARG

Display Configuration	df	*t*-Value	*p*-Value
V30-R30	19	3.572	**.001****
V30-R60	19	2.663	**.008****
V60-R30	19	2.101	**.025***
V60-R60	19	1.308	.103

$^*p < .05$. $^{**}p < .01$.

$p = .008$, V60-R30, $t\left(19\right) = 2.101$, $p = .025$). This result indicates that, with the LFG, participants complete the trials faster when the virtual text appears at V30 or when switching vergence is required (V30-R60 and V60-R30). The task completion scores for the two pairs of AR glasses are statistically comparable under the V60-R60 display configuration. With the virtual text located at V60 in the V60-R30 and V60-R60 configurations, the task completion score for both pairs of AR glasses is less different than the case where the virtual text appears at V30. This is especially true for V60-R60, for which there is no discernable score difference between the two pairs of AR glasses. The result is reasonable because participants wearing the ARG were impacted less by VAC with virtual text at 60 cm than that at 30 cm.

Error Rate

The error rate is defined as the sum of miscounted errors over the total number of trials completed, the accuracy is defined as $1 - error\,rate$, and the task completion rate is defined as the total number of trials completed over the total number of trials. In our experiment, the total number of trials is 800 for each pair of AR glasses.

We perform a one-tailed paired *t*-test on each participant's error rate of the two pairs of AR glasses. The result shows no significant difference between the error rate of both pairs of AR glasses ($t\left(19\right) = .526$, $p = .303$). Each participant miscounted 2.35 letters on average in the experiment, which is surprisingly low given that each participant completed an average of 60.2 trials. As Table 5 shows, although the overall error rate is very low, the LFG has a lower miscounted error sum and a higher task completion rate than the ARG. This result indicates that the LFG leads to more accurate user performance than the ARG. We also notice that all participants tended to count out all the target letters before moving to the next trial. On the one hand, such a tendency may be the main reason for the low error rates. On the other hand, this tendency may slow the participants down during the task when they are impacted by the VAC, which can be seen when comparing the task completion performance of the two pairs of AR glasses.

Eye Strain Level

Participant ratings reveal little information about the relationship between the experimental factors (type of glasses, virtual text distance, and real text distance) and the rating of eye strain level. Overall,

Table 5. Total number of trials, sum of miscounted errors, task completion rate, and error rate for the ARG and the LFG

	ARG	LFG
Total number of trials completed	562	642
Sum of miscounted errors	24	23
Task completion rate	70.25%	**80.25%**
Error rate	4.27%	**3.58%**

the average eye strain level increases with the number of runs with a 0.79 positive correlation coefficient (Figure 9), while the ratings vary a lot for each participant. For example, one participant rated the eye strain level as "(1) Very Fresh" throughout the whole experiment. Also, a few participants ended up with the state of "(5) Severe Strain" with the rating of "(2) Fine" at the beginning of the experiment. The results suggest that the 60-second break may be too short and caused the fatigue to carry over to the experiment in the next display configuration.

User Experience and Feedback

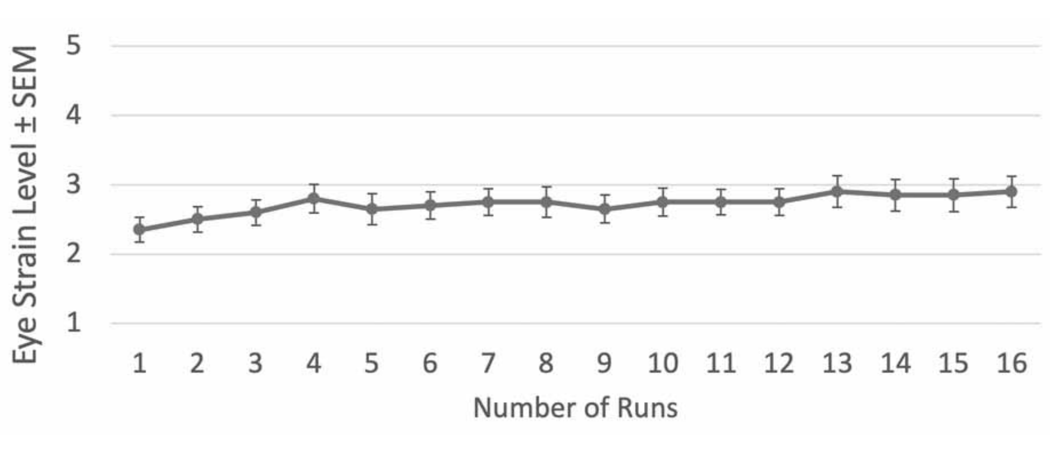
The participants reported some issues during the experiment. While using the ARG under the V30-R30 configuration, some participants reported blurry vision when switching their gazes between the virtual and the real text. We also noted that a participant described the process of focusing on the near virtual object as "two-step focusing," first adjusting the vergence to V30 and then accommodating to the optical focal distance (2.5 m) of the ARG. About half of the participants experienced the two-step focusing process, and it took them more time to fuse binocular view with the ARG than with the LFG. Five participants said they saw dim text displayed by the LFG, but this did not impact them to successfully perform the trials.

The participant feedback indicates that 80% (16 out of 20) of the participants preferred the LFG over the ARG because they felt it was easier to focus on the light field text than the AR text. Two participants, however, did not feel any difference between the two types of AR glasses. There were also two participants in favor of using the ARG because they felt the text displayed by the ARG was sharper and brighter than that displayed by the LFG, whereas most participants did not mention the sharpness and brightness of the virtual text displayed by the two types of AR glasses. One of the two participants who preferred the ARG experienced the impact of VAC (eye strain and binocular fusion difficulty) and completed fewer trials with the ARG (20) than with the LFG (25). It is likely that this participant's preference was not related to VAC because the sharpness and brightness of the text are not associated with the existence of VAC.

Discussion

In this study, we compare the user performance of light field AR glasses with conventional AR glasses through a series of visual search tasks. This is an efficient way to amplify the impact of VAC and is reliable for evaluating user performance. However, visual search tasks are not a normal usage scenario for AR glasses. Therefore, future experiments could include a more diverse series of

Figure 9. The average eye strain level increases as the number of runs grows. There is a positive correlation between the eye strain level and the number of runs.



tasks. In addition, due to the required accommodation of this experiment, we limited the age range of participants. In future experiments, we hope to recruit a larger sample and expand the age range of the participants to (12, 70) to capture gender differences and age effects. We believe experimental results from a wider demographic range will be more convincing.

When designing this experiment, all letters from the Latin alphabet were used except "I," "J," and "L." These letters were excluded because combinations of these visually-similar letters could confuse the participants and hence influence text recognition results. However, the participant feedback suggests two other letter combinations may influence text recognition. First, participants took more time and concentration to identify the target letter when the letters "W," "V," and "Y," appeared in the same text block. Second, when the upper-case and the lower-case versions of a target letter resemble each other (e.g., C, O, P, etc.), participants could find the target letter by recognizing its shape without reading and recognizing the target letter. Allowing participants to substitute an easier task (shape recognition) for the original task (text recognition) may unwantedly influence the experimental results. While it is unclear if the same level of concentration and accommodation are required for "shape recognition" and "text recognition" (i.e., reading), the difference between these tasks may affect the task completion score. The confounding effects of these specific combinations can be addressed by either explicitly controlling these combinations or changing task design to completely avoid the issue.

In addition, although we collected self-reported eye strain ratings, we did not find any relevance between eye strain ratings and the three experimental factors (type of glasses, virtual text distance, and real text distance). However, this does not suggest that these factors are insignificant. It is possible that the task completion time is too short to experience significant differences in eye strain. Therefore, to make these differences more significant, experiment run-time should be extended and the number of runs should be increased. Additionally, each display configuration should be either tested on different days to avoid carry-over fatigue or arranged by a Latin Square to counterbalance the impact of display configuration order on experimental results.

CONCLUSION

This paper describes an empirical study comparing the differences between a light field AR display and a conventional AR display in virtual-real integration efficiency. The results show that, the light field AR display leads to higher task completion scores than the conventional AR display. Although the ratings of eye strain level do not differ much in each display configuration under the current experimental design, the subjective ratings reveal that the majority of participants preferred the light field AR glasses over the conventional AR glasses. Overall, this work demonstrates the strengths of light field display technology for AR.

ACKNOWLEDGMENT

This work was supported in part by National Science and Technology Council of Taiwan under Contracts 110-2221-E-002-108-MY3 and 112-2218-E-002-035-MBK and in part by National Taiwan University under Contracts 111L880602 and 112L900902.

REFERENCES

Adelson, E. H., & Bergen, J. R. (1991). The Plenoptic Function and the Elements of Early Vision. In M. S. Landy & H. Anthony Movshon (Eds.), *Computational Models of Visual Processing*. MIT Press.

Akeley, K., Watt, S. J., Girshick, A. R., & Banks, M. S. (2004). A stereo display prototype with multiple focal distances. *ACM Transactions on Graphics*, *23*(3), 804–813. doi:10.1145/1015706.1015804

Daniel, F., & Kapoula, Z. (2019). Induced vergence-accommodation conflict reduces cognitive performance in the Stroop test. *Scientific Reports*, *9*(1), 1247. doi:10.1038/s41598-018-37778-y PMID:30718625

Duane, A. (1922). Studies in monocular and binocular accommodation with their clinical applications. *American Journal of Ophthalmology*, *5*(11), 865–877. doi:10.1016/S0002-9394(22)90793-7

Edgar, G. K., Pope, J. C. D., & Craig, I. R. (1994). Visual accomodation problems with head-up and helmet-mounted displays? *Displays*, *15*(2), 68–75. doi:10.1016/0141-9382(94)90059-0

Eiberger, A., Kristensson, P. O., Mayr, S., Kranz, M., & Grubert, J. (2019). Effects of Depth Layer Switching between an Optical See-Through Head-Mounted Display and a Body-Proximate Display. *Symposium on Spatial User Interaction*, 1–9. doi:10.1145/3357251.3357588

Fernández, E. J., & Artal, P. (2003). Membrane deformable mirror for adaptive optics: Performance limits in visual optics. *Optics Express*, *11*(9), 1056–1069. doi:10.1364/OE.11.001056 PMID:19465970

Gabbard, J. L., Mehra, D. G., & Swan, J. E. (2019). Effects of AR Display Context Switching and Focal Distance Switching on Human Performance. *IEEE Transactions on Visualization and Computer Graphics*, *25*(6), 2228–2241. doi:10.1109/TVCG.2018.2832633 PMID:29994003

Gershun, A. (1939). The Light Field (P. Moon & G. Timoshenko, Trans.). *Journal of Mathematics and Physics, 18*, 51-151. (Original publication 1936)

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision (Charlottesville, Va.)*, *8*(3), 33–33. doi:10.1167/8.3.33 PMID:18484839

Hua, H., & Javidi, B. (2014). A 3D integral imaging optical see-through head-mounted display. *Optics Express*, *22*(11), 13484–13491. doi:10.1364/OE.22.013484 PMID:24921542

Iavecchia, J. H., Iavecchia, H. P., & Roscoe, I., Stanley N. (1988). Eye Accommodation to Head-Up Virtual Images. *Human Factors, 30*(6), 689–702. 10.1177/001872088803000605

Koetting, R. A. (1970). Stereopsis in presbyopes fitted with single vision contact lenses. *Optometry and Vision Science*, *47*(7), 557–561. doi:10.1097/00006324-197007000-00006 PMID:5270396

Kuiper, S., & Hendriks, B. H. W. (2004). Variable-focus liquid lens for miniature cameras. *Applied Physics Letters*, *85*(7), 1128–1130. doi:10.1063/1.1779954

Levoy, M., & Hanrahan, P. (1996). Light Field Rendering. *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, 31–42. doi:10.1145/237170.237199

Lippmann, G. (1908). Épreuves réversibles donnant la sensation du relief. *J. Phys. Theor. Appl.*, *7*(1), 821–825. doi:10.1051/jphystap:019080070082100

Ren, H., Fox, D. W., Wu, B., & Wu, S.-T. (2007). Liquid crystal lens with large focal length tunability and low operating voltage. *Optics Express*, *15*(18), 11328–11335. doi:10.1364/OE.15.011328 PMID:19547490

Shibata, Kim, Hoffman, & Banks. (2011). *Visual discomfort with stereo displays: Effects of viewing distance and direction of vergence-accommodation conflict*. 10.1117/12.872347

Shiwa, S., Omura, K., & Kishino, F. (1996). Proposal for a 3-D display with accommodative compensation: 3DDAC. *Journal of the Society for Information Display*, *4*(4), 255–261. doi:10.1889/1.1987395

Smith, M., Streeter, J., Burnett, G., & Gabbard, J. L. (2015). Visual search tasks: The effects of head-up displays on driving and task performance. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 80–87. doi:10.1145/2799250.2799291

Walt, J., Rowe, M., & Stern, K. (1997). Evaluating the functional impact of dry eye: The Ocular Surface Disease Index. *Drug Information Journal*, *31*(1436), b5.

Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Vision Research*, *35*(19), 2731–2736. doi:10.1016/0042-6989(95)00018-U PMID:7483313

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision (Charlottesville, Va.)*, *5*(10), 7. doi:10.1167/5.10.7 PMID:16441189

Wu, W., Llull, P., Tosic, I., Bedard, N., Berkner, K., & Balram, N. (2016). Content-adaptive focus configuration for near-eye multi-focal displays. *2016 IEEE International Conference on Multimedia and Expo (ICME),* 1–6. doi:10.1109/ICME.2016.7552965

Yano, S., Ide, S., Mitsuhashi, T., & Thwaites, H. (2002). A study of visual fatigue and visual comfort for 3D HDTV/HDTV images. *Displays*, *23*(4), 191–201. doi:10.1016/S0141-9382(02)00038-0

Yuuki, A., Itoga, K., & Satake, T. (2012). A new Maxwellian view display for trouble-free accommodation. *Journal of the Society for Information Display*, *20*(10), 581–588. doi:10.1002/jsid.122

Zou, B., Liu, Y., Guo, M., & Wang, Y. (2015). EEG-Based Assessment of Stereoscopic 3D Visual Fatigue Caused by Vergence-Accommodation Conflict. *Journal of Display Technology*, *11*(12), 1076–1083. doi:10.1109/JDT.2015.2451087