Effects of AR Display Context Switching and Focal Distance Switching on Human Performance

Joseph L. Gabbard, Divya Gupta Mehra, & J. Edward Swan II, *Senior Member, IEEE*

Abstract—In augmented reality (AR) environments, information is often distributed between real world and virtual contexts, and often appears at different distances from the user. Therefore, to integrate the information, users must repeatedly switch context and refocus the eyes. To focus at different distances, the user's eyes must accommodate, which when done repeatedly can cause eyestrain and degrade task performance. An experiment was conducted that examined switching context and focal distance between a real and an AR environment, using a text-based visual search task and a monocular optical see-through AR display. Both context switching and focal distance switching resulted in significantly reduced performance. In addition, repeatedly performing the task caused visual fatigue to steadily increase. Performance was particularly poor for virtual text presented at optical infinity, and for target letters that participants tried to read before their eyes had completely accommodated to a new focal distance. The results show that context switching and focal distance switching are important AR user interface design issues.

Index Terms— H.5.1: Multimedia Information Systems — Artificial, Augmented, and Virtual Realities; H.5.2: User Interfaces — Ergonomics, Evaluation / Methodology, Screen Design, Style Guides

1 Introduction

Augmented reality (AR), which superimposes virtual symbology and text onto real world views, intends to enhance users' understanding of the world. By providing users with proximal and relevant virtual information, AR has been shown to assist user task performance across a number of domains, including military (Livingston et al. [1]), transportation and aerospace (Regenbrecht et al. [2]), healthcare (Barsom et al. [3]), logistics (Schwerdtfeger et al. [4]), and maintenance and assembly (Webel et al. [5]), to name a few. However, the virtual information, coupled with a complex physical world, creates the potential for excessive perceptual and cognitive load. In particular, occupational AR users must simultaneously perceive, comprehend, and assimilate virtual and real world visual information while concurrently performing one or more real world tasks. And while these challenges can be mitigated by good visual user interface design, other challenges arise from the nature of the hardware itself, as well as the interplay of visual demands imposed by both the AR hardware and real-world environments.

Among these challenges is that the overwhelming majority of today's AR optical see-through head-worn and head-up display technologies present AR graphics at a fixed *focal distance*—the distance at which the eye's lens must accommodate to see in sharp focus. Given that real world objects of interest occur at a range of distances, fixed focal distance displays require users to continuously switch both accommodation and attention between the real-world scene and virtual graphics. Frequent shifting of gaze to different focal depths can result in excessive strain on the accommodation mechanism of the eye (Neveu et al. [6], Lambooij et al. [7]). Placing additional strain on users' accommodation mechanism can further negatively affect an AR system's overall usability by increasing eye fatigue (Lambooij et al. [7], Miller et al. [8]), biasing distance estimations (Roscoe [9]), increasing the likelihood of missed targets (Edgar et al. [10]), limiting the ability to fuse binocular images (Wu et al. [11]), and introducing visual conflicts that lead to simulator sickness (Carnegie and Rhee [12]).

This motivated the study described in this paper, where the effects of changing both focal distance and visual context were examined. Here, *focal distance switching* is defined as the *change in accommodation* when switching between one focal distance and another, and *context switching* is defined as the switching in *visual and cognitive attention* between real world and virtual information. These two concepts are distinct yet often tightly coupled. For example, consider a surgeon looking between a patient and flat panel display. If the display is positioned about the same distance from the surgeon as the patient, then there would be context switching but no focal distance switching. However, if the display was positioned much farther than the patient, then there would be both context switching and focal distance switching. If the surgeon's current task did not require looking at the display, then there would be no context switching.

These facts motivated the experiment reported here, which examined switching context and focal distance between a real and an AR environment. The experiment used a text-based visual search task that required integrating information presented in both the real world and AR, and was presented on a monocular AR display.

- *J.L. Gabbard is with the Grado Department of Industrial and Systems Engineering, Virginia Tech, VA 24061. E-mail: jgabbard@vt.edu.*
- *D.G. Mehra is with HERE Technologies, Chicago, IL 60606. E-mail: digupta@gmail.com.*
- *J.E. Swan II is with the Department of Computer Science and Engineering, Mississippi State University, MS 39762. E-mail: swan@acm.org.*

Author version; date of current version 25 Apr. 2018.

2 Background and Related Work

2.1 Accommodation and Vergence

The ability of the eye to adjust its focal length is known as *accommodation* (Goss and West [13]). When the eye accommodates to a nearby object, the ciliary muscles contract, causing the lens to assume a more convex shape, while when a distant object is viewed, the ciliary muscles relax, causing the lens to become thinner and flatter. The four primary stimuli that drive the accommodation response are blur, proximity, binocular disparity, and an empty visual field. When viewing an empty field, accommodation comes to rest at an intermediate level. This level varies per person, and is termed their *dark focus*—the distance at which their eyes focus when the ciliary muscles are in a relaxed state (Iavecchia et al. [14]). In addition, with increasing age the lens becomes less flexible and the depth range at which accommodation is possible decreases, and by the 40s to 50s, the ability to focus on near objects and accommodate to different depths is significantly limited (Hoffman et al. [15]).

Because the eyes accommodate to focus on objects in the real world, a long-held assumption was that when users viewed virtual content in a head-mounted or head-up display, their eyes would also accommodate to the distance of the virtual symbology. Therefore, for example, head-up displays used in aviation are *collimated*—focused at optical infinity—under the assumption that pilots would not have to refocus the eyes when shifting gaze between the symbology and the real world beyond the cockpit (Iavecchia et al. [14]). However, multiple studies have shown that when viewing virtual content, the eyes generally do not accommodate to the display's focal demand; instead, they accommodate to an intermediate distance between the display's focal demand and the individual's dark focus (Owens [16], Iavecchia et al. [14]). This incorrect accommodation biases judgments of size and distance (Roscoe [17]), and is implicated in many aviation crashes caused by pilot error (Roscoe [18], Iavecchia et al. [14]). When viewing virtual symbology, the acuity demand of the visual task determines how far from an individual's dark focus the eye accommodates: symbology is a weak accommodation cue, while reading text causes more accommodation (Iavecchia et al. [14], Schwerdtfeger et al. [4]).

While accommodation focuses the eyes, viewing objects at different distances also requires the eyes to rotate. These rotational eye movements are termed *vergence* movements—when shifting gaze to a nearby object, the eyes *converge*, while when shifting gaze to a distant object, the eyes *diverge*. The primary stimulus for vergence eye movements is stereo disparity between the two retinal images: the human visual system rotates the eyes to reduce this disparity until the images can be fused into a single image (Krishnan and Stark [19]). Moreover, similar to *dark focus*, there is also a resting point of vergence termed *dark vergence*: the distance at which the eyes verge when the rotational muscles are relaxed.

In addition to the independent stimuli to accommodate and verge, accommodation and vergence are linked so each response also drives the other: a stimulus to accommodate also causes the eyes to verge, and a stimulus to verge also causes the eyes to accommodate (Mon-Williams and Tresilian [20], Krishnan and Stark [19]). Both accommodation and vergence are also linked to pupil diameter, which controls the eye's focal depth of field; the three responses all drive and are driven by each other. Because of the strong linkage between vergence, accommodation, and pupil diameter, using monocular AR displays, such as the one used in the current experiment, can be expected to drive vergence eye movements (Myers and Stark [21]).

Although viewing real world objects typically requires vergence and accommodation responses that co-vary in depth, it is relatively easy for the human visual system to override the vergence-accommodation linkage. This is why almost all users can successfully fuse virtual objects presented at disparity depths that are different from the fixed focal depth of an augmented or virtual reality display. However, this *vergence-accommodation* conflict distorts perceived depth and size (Kruijff, Swan, and Feiner [22]), and over time can also disrupt normal binocular vision (Wann and Mon-Williams [23]). A number of experiments have shown that vergence-accommodation conflicts reduce performance when fusing stereo images and other visual tasks (Akeley et al. [24], MacKenzie et al. [25], Hoffman et al. [15]).

In addition to driving the accommodative response, blur is an important pictorial depth cue. Blur has been found to affect perceived size and depth, both directly as a depth cue, and indirectly through interaction with other perspective cues. This has been found in experiments using sparse laboratory scenes (Watt et al. [26]), as well as photos of natural objects (Nefs [27]), photos of natural textures (Vishwanath and Blaser [28]), and photos of both urban and natural scenes (Held et al. [29]).

2.2 Visual Fatigue

Visual fatigue has been defined as "physiological strain or stress resulting from excessive exertion of the visual system" (Lambooij et al. [7]). It is a concern for any visual task, and in extreme cases can make a visual display unusable. For example, Schwerdtfeger et al. [4] report a user study of an industrial order-picking AR application, where users pick numbered parts out of bins and load them into a wheeled basket. In order to verify that a part was correct, users had to compare a long part number printed on the item with the same part number presented on a monocular AR display. After 2 hours, 20% of the participants experienced significant visual fatigue, described as "pressure in the eyes", and had to stop using the AR display. Note that, in the context of the experiment reported here, this task requires both context switching and focal distance switching.

Therefore, understanding the causes of visual fatigue is an important topic. Lambooij et al. [7] provide a comprehensive review, primarily concerned with flat panel autostereoscopic displays used for 3D television. In this context, they find that visual discomfort is primarily driven by stereo disparity greater than 1 degree of visual angle, quickly changing depth cues resulting in demands on the accommodation-vergence linkage, insufficient depth information, and unnatural blur. Vergence-accommodation conflicts also cause visual fatigue and discomfort (MacKenzie et al. [25], Hoffman et al. [15]), and viewing nearby objects exerts a greater strain on the vergence muscles than distant objects (Davson [30]). Repeated accommodation causes eyestrain (Miller et al. [8]), and because it requires relatively large amounts of muscle constriction, also leads to eye fatigue (Ostberg [31], Lambooij et al. [7]). And, while depth-of-field blur alone does not directly lead to discomfort (O'Hare et al. [32]), the visual system may attempt to accommodate to relieve the blur, which can lead to discomfort (O'Hare et al. [32]). Several researchers have concluded that accommodative strain can be lowered if work is performed at a viewing distance near the dark focus (Johnson [33], Ostberg [31], Roscoe [9]).

2.3 AR Head-Up Displays

Head-up displays (HUDs) superimpose virtual images on a window; HUDs are typically mounted in vehicles to provide instrument symbology in the forward view of the operator (Weintraub and Ensing [34]). HUDs share many properties with optical see-through head-mounted AR displays: both typically display at a fixed focal depth, both can suffer from visual interference caused by ambient lighting and real-world backgrounds, and both can present conformal AR graphics. Given these similarities, relevant findings from aviation research and, more recently, from surface transportation research should also be considered.

A major design issue for HUDs is the focal distance of the symbology: the distance should reduce focal distance switching between the real-world view and the symbology. Previous findings are divergent, with some recommending that the symbology be displayed at approximately 2 meters from the user's eye (Inuzuka et al. [35], Kato et al. [36]), while others recommend a distance closer to optical infinity (Okabayashi et al. [37], Weintraub et al. [38]). However, when the visual processing demands of air traffic controllers using a large HUD in an airport tower were examined (Albrecht [39]), better performance was found for uncollimated HUDs. In this case, because the unattended information was blurred, operators were able to easily attend to either the HUD or the background scene. Overall, there is no consensus on the optical distance at which to display HUD symbology (Weintraub and Ensing [34], Wolffsohn et al. [40]).

In aviation, HUDs used to annotate the horizon and runway are often collimated at optical infinity. However, the distances at which pilots work are significantly greater than drivers, and in the surface transportation community, it is generally accepted that head-up displays should be focused between 2.0 and 2.5 meters (Tufano [41]). However, Tufano also argues that the justification for this distance is based on prior studies [35, 36] that examined extracting information from the HUD, but Tufano points out that these studies did not examine the effect of the HUD on the perception of real objects in the driver's forward view. This is an important observation, because a driver's primary task is to attend to real world objects.

Indeed, as we move towards tasks that more tightly integrate real and virtual information, there will be cases, such as conformal text labels, where we want the graphics to be at the same focal distance as real-world objects, and other cases, such as notifications, where we might prefer that the graphics be at a standard focal distance. Gabbard et al. [42] note the need for an AR display capable of generating graphics at multiple focal depths simultaneously. With such as a display, conformal AR graphics would not require switching focal distance. Along these lines, there are a handful of companies producing AR displays to support multiple focal depths, one example being Magic Leap. In another example, Bark et al. [43] report their work with a prototype volumetric head-up display with a focal plane that can be moved between 5 meters and infinity, which resulted in significantly better depth judgments than a fixed focal plane distance.

2.4 Context Switching in AR Head-Worn Displays

The only published work we found that explicitly examines AR context switching is Huckauf et al. [44]. They examined context switching in the same industrial order-picking AR application as Schwerdtfeger et al. [4], and found that context switching between AR and real world displays resulted in decreased visual performance. We are not aware of any existing work that has systematically examined both context and focal switching in the same experiment, as we report here.

3 Experiment

The purpose of the experiment was to examine the effect of both context switching and focal distance switching, while performing a task that required integrating information presented in both the real world and AR. As the reviewed literature indicated that both context switching and focal distance switching affect human performance (e.g., Schwerdtfeger et al. [4], Neveu et al. [6], Lambooij et al. [7], Huckauf et al. [44]), it was anticipated that similar effects would occur in AR, but this hypothesis needed to be verified. In addition, the literature indicated that previous AR user studies had not employed tasks that required participants to repeatedly attend to both the real world and AR. Therefore, an experimental task was developed that requires participants to integrate information from both locations. Finally, because reading tasks are commonly used when studying accommodation (e.g., Schwerdtfeger et al. [4], Iavecchia et al. [14], Owens and Wolf-Kelly [45]), the experimental task involves scanning text strings for letters.

The initial motivation for the experiment was that we obtained a pre-production version of a Microvision Nomad optical see-through AR display. This display has an adjustable focal distance, and this feature makes it *unique*: in ~20 years of experience in the AR field, the authors have not encountered any other commercially available AR display that allows the focal distance to be adjusted. Outside of building ones' own AR display, the ability to adjust the Nomad's focal distance makes it an appropriate, and perhaps exclusive, display for AR research involving accommodation.

The Nomad is a monocular display. In the history of head-mounted displays, there is a long history of monocular designs, in particular for military pilot tasks, dating from the 1960's onwards (Hughes et al. [46]). Monocular Nomad displays have been used for a comprehensive industrial order-picking project at Volkswagen AG (Huckauf et al [44], Schwerdtfeger et al. [4], Tümler et al. [47]), and newer monocular displays, including the Google Glass and Vuzix M100, are being used for similar industrial tasks (UBiMAX [48]). When applied to AR applications, monocular displays have often been intended for tasks where the augmented information is

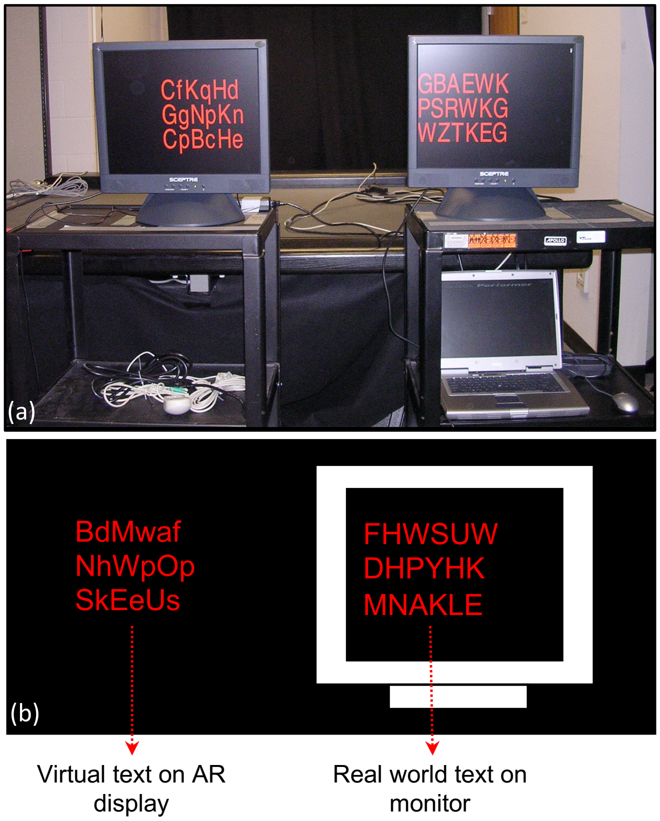


Fig. 1. The experimental task and visual stimuli. (a) Participant's view of the real-world to real-world condition, where participants identified the doubled target letter "G" in the left text, and then counted the number of target occurrences in the right text; the correct answer is "3". Note that the lights are on to show the equipment; the lights were off during the experiment. (b) Conceptual depiction of the AR to real-world condition, where participants identified the doubled target letter "E" in the left text, and then counted the occurrences in the right text; the correct answer is "1".

context- or location-aware, but is not visually conformal to a specific object at the location, such as, for example, the maps application in Google Glass.

3.1 Experimental Subtask and Task

A structured visual search task was developed that forced participants to switch both context and focal distance between real and virtual text. The task was inspired by reading AR text labels and comparing them to text in the real world, similar to the industrial task described by Schwerdtfeger et al. [4] and discussed above in Section 2.2.

Fig. 1 illustrates the task. Participants saw three different strings of random characters in two separate blocks, herein referred to as the *left text* and the *right text*. Each block consisted of 3 lines of text, and each line consisted of 6 characters, so there were 18 characters in each block. The left text consisted of alternating upper- and lower-case letters, while all characters in the right text were upper case. Participants were instructed to examine the left text to locate a pair of identical letters, where the first letter was upper case and the second lower case. This was called the *target letter*: "Gg" in Fig. 1a; "Ee" in Fig. 1b. Placement of the target letter in the left text was randomized. Letters "i", "l", and "j" were not included, because they looked too similar to each other: "I", "l", and "J" look almost identical in the san-serif font that was used. After locating the target letter in the left text, participants were instructed to look at the right text and count how many times the target letter appeared. The number and placement of the target letter in the right text was also randomized. The target letter could appear 0, 1, 2, or 3 times. Target letters that appeared 2 times always appeared in two different lines of the right text, and target letters that appeared 3 times always appeared in all three lines of the right text. Participants responded by pressing a number key on a numeric keypad. In Fig. 1a, the correct answer is "3", as "G" appears in all three lines of the right text. In Fig. 1b, the correct answer is "1", as "E" appears in the bottom line of the right text. Depending on the experimental condition, the left text could be presented on a monitor (Fig. 1a), or in an AR display (Fig. 1b). In all conditions, the right text was presented on a monitor.

Each participant response constituted a subtask. After completion of a subtask, the right text remained the same, while the left text changed and presented a new target letter. Participants had 25 seconds to complete a maximum of five subtasks, which constituted one task. On completion of either 5 subtasks, or after 25 seconds, both screens blanked for two seconds, and then a new set of left text and right text was presented, which began the next task.

Fig. 2. The experiment used a Microvision Nomad optical see-through AR display. The adjustable focal distance slider is outlined in green.

3.2 Apparatus

Participants used a Microvision Nomad monocular optical see-thorough AR display (Fig. 2). A movable slider on the Nomad allows the focal distance to be adjusted from 1 foot to optical infinity (Fig. 2, green outline). As the focal distance slider of the Nomad was not calibrated, we used a dioptometer[1] with a –5/+5 diopter range to determine slider positions that resulted in the tested focal distances. The Nomad has a resolution of 800x600 pixels.

The field of view is 23 × 17 degrees, which is approximately equivalent to a 17-inch monitor at arm's length. The luminance of the Nomad ranges from 1 to 800 foot-lambert. The only supported display color is red. The total head-worn weight of the Nomad is 226.8 grams. An Intersense IS 900 tracking system receiver was mounted to the display to provide head tracking.

Participants used the Nomad display with both eyes open (Fig. 2), and therefore they saw the virtual text monocularly, and the real-world text binocularly. However, because the experimental room was dark, room items had low visual salience compared to the text, and therefore when viewing the

[1]A *dioptometer* is a precision optical laboratory instrument that precisely measures the focal distance of an optical system.

virtual text participants were not distracted by seeing different information with their unused eye.

Depending on the experimental condition, either one or two 19-inch LCD monitors were used to display real world text (Fig. 1). Each monitor was placed on a portable cart that allowed the distance to the real-world text to be easily adjusted. The monitors were set to 800×600 resolution to match the resolution of the AR display. Text was rendered in a red color that closely matched the color of the AR display. During the experiment the room lights were switched off, and virtual text was viewed against a black curtain background, which eliminated any color shifting of the AR text from background color (Gabbard, Swan, and Zarger [49]). This lighting design made the virtual and real-world text appear very similar in brightness, color, and appearance.

The experiment used a custom AR system developed by the authors and additional collaborators. The code was written in C++ and ran on Linux OS, and was built using the Diverse Toolkit [50].

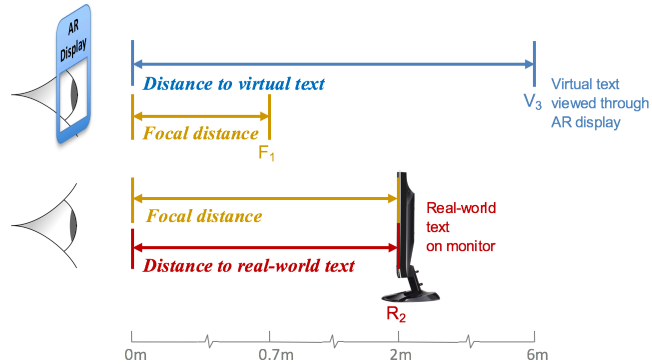
3.3 Independent Variables

The study manipulated 4 independent variables (Figs. 1 and 3): *context switching* (real-real, AR-real), *distance to real world text* (near, medium, far), *distance to virtual text* (near, medium, far), and *focal distance* (near, medium, far). For the distances, *near* = 0.7 meters, *medium* = 2 meters, and *far* = 6 meters.

Context Switching: In the real-real condition, when both the left and right text blocks were shown on monitors (Fig. 1a), *context switching* was not required to complete the task. However, in the AR-real condition, when the left text was shown in the AR display and the right text on a monitor (Fig. 1b), context switching was required.

Distance to Real World Text: For both real and AR objects, the perception of distance is driven by *depth cues* (Cutting and Vishton [51]); the ones used in this experiment are occlusion, binocular disparity, motion parallax, vergence angle, height in the visual field, relative size, and relative density. Real world text was presented on the monitors, and *distance to real world text* was the distance of the monitors from the participant. Because of the physical design of the experiment, the distance to the real-world text was encoded by the depth cues of binocular disparity, motion parallax, vergence angle, and relative size.

Distance to Virtual Text: Virtual text was presented with the AR display conditions only, and *distance to virtual text* was the distance encoded by the depth cues available from the display's head-tracked, monocular AR graphics: motion parallax and relative size. When objects are viewed monocularly, motion parallax is a very powerful depth cue that largely takes the place of binocular disparity (Rogers and Graham [52]). However, observers did not sway in a controlled manner, as they have done for other experiments that systematically used motion parallax to encode monocular depth (e.g., McCandless, Ellis, and Adelstein [53]), and distance to virtual text is not a primary independent variable—the primary independent variables are context and focal distance switching.



LEVELS OF INDEPENDENT VARIABLES

Manipulated Independently for Virtual Text on AR Display		
	Manipulated in Concert for Real-world Text on Monitor	
Distance to virtual text	*Focal distance*	*Distance to real-world text*
Near – 0.7 m	Near – 0.7 m	Near – 0.7 m
Medium – 2 m	Medium – 2 m	Medium – 2 m
Far – 6 m	Far – 6 m	Far – 6 m

Fig. 3. The independent distance variables. During the experiment, each was manipulated to distances of 0.7 m, 2 m and 6 m. Focal distance was manipulated separately for the virtual text (independent of distance to virtual text) and real-world text (in concert with distance to real-world text). Note the configuration depicted above represents condition $V_3F_1R_2$ in Table 2.

Focal Distance: As discussed in detail in Section 2.1, *focal distance* is the distance at which the eye must accommodate in order for the text string to be seen in sharp focus. For the real-world text strings, by definition the focal distance was the same as the distance to the physical monitor (Fig. 3). However, for the virtual text strings, the AR display's optical system determined the focal distance (Fig. 3). Because the AR display's focal distance could be adjusted with a physical slider (Fig. 2, green box), for the virtual text strings it was possible for the distance to the text string (encoded by motion parallax and relative size), and the focal distance (encoded by the optical system), to vary independently. Thus, Fig. 3 depicts both context switching (the left text is seen in the AR display, and the right text is seen on a monitor) and focal distance switching (the AR display focal distance is set to 0.7 meters, and the real-world text focal distance is 2 meters). Also, in Fig. 3 the virtual text distance (6 meters) differs from the focal distance (0.7 meters).

3.4 Rationale for Levels of Independent Variables

Selection of the three distance levels—near = 0.7 meters, medium = 2 meters, and far = 6 meters—was based on inconclusive literature findings that claim better performance when the eye is focused at these distances. For the near distance, researchers have found that when the eye focuses at approximately arm's distance, accommodation is minimal and the eye is in a relaxed position (Iavecchia et al. [14], Jaschinski-Kruza [54], Norman and Ehrlich [55], Tufano [41]). An arm's distance has been approximated as 0.7 meters. Rationale for the medium distance was based on head-up displays literature that finds a distance of approximately 2 meters leads to faster performance on visual detection tasks (Inuzuka et al.

TABLE 1
REAL-WORLD TO REAL-WORLD CONDITIONS

Real-world distance to left text (R)	*Real-world distance to right text (R)*		
	R_1: near (0.7 m)	R_2: medium (2 m)	R_3: far (6 m)
R_1: near (0.7 m)	**R_1R_1**		
R_2: medium (2 m)		**R_2R_2**	
R_3: far (6 m)			**R_3R_3**

TABLE 2
REAL-WORLD TO AUGMENTED REALITY CONDITIONS

Virtual distance to left text (V)	*Focal distance to left text* (F)	*Real-world distance to right text* (R)		
		R_1: near (0.7m)	R_2: medium (2m)	R_3: far (6m)
V_1: near (0.7 m)	F_1: near (0.7 m)	**$V_1F_1R_1$**	$V_1F_1R_2$	$V_1F_1R_3$
	F_2: medium (2 m)	$V_1F_2R_1$	$V_1F_2R_2$	$V_1F_2R_3$
	F_3: far (6 m)	$V_1F_3R_1$	$V_1F_3R_2$	$V_1F_3R_3$
V_2: medium (2 m)	F_1: near (0.7 m)	$V_2F_1R_1$	$V_2F_1R_2$	$V_2F_1R_3$
	F_2: medium (2 m)	$V_2F_2R_1$	**$V_2F_2R_2$**	$V_2F_2R_3$
	F_3: far (6 m)	$V_2F_3R_1$	$V_2F_3R_2$	$V_2F_3R_3$
V_3: far (6 m)	F_1: near (0.7 m)	$V_3F_1R_1$	$V_3F_1R_2$	$V_3F_1R_3$
	F_2: medium (2 m)	$V_3F_2R_1$	$V_3F_2R_2$	$V_3F_2R_3$
	F_3: far (6)	$V_3F_3R_1$	$V_3F_3R_2$	**$V_3F_3R_3$**

[35], Kato et al. [36]). However, other research has reported that display symbology at optical infinity is better for visual tasks (Okabayashi et al. [37], Weintraub et al. [38]). According to Simonelli [56], optical infinity is roughly 6 meters, and therefore 6 meters was selected for the far distance. In addition, 0.7, 2, and 6 meters encompass a range of distances that are relevant for many indoor AR applications. In particular, working distance for personal space applications like surgery, maintenance, and repair is roughly an arm's length, whereas indoor navigation often involves distances between 2 and 6 meters.

To ensure that the text was legible at all distances, we set the text size so that at the far distance the text met the ANSI standard for minimum character height of 22 arc minutes of visual angle (ANSI [57]). Both the authors and pilot participants determined that the text was easily legible at all distances.

3.5 Experimental Design and Counterbalancing

The experimental design is shown in Tables 1 and 2. For the *real-real* condition, Table 1 shows the 9 possible distance combinations. Out of these, participants saw the three conditions where the distance to each monitor matched; these are labeled and highlighted in Table 1. For the *AR-real* condition, Table 2 shows the 27 possible distance combinations; participants saw all 27 of these conditions. Therefore, participants saw 3 + 27 = 30 conditions in total. The cells in Tables 1 and 2 uniquely label each condition, so for example, condition R_2R_2 means that participants saw real world left and right text at a distance of 2 meters, while condition $V_3F_1R_2$ means that participants saw virtual left text at a distance of 6 meters and a focal distance of 0.7 meters, and real world right text at a distance of 2 meters; this is the experimental condition depicted in Fig. 3.

As each participant saw all 30 conditions, a within-subjects design was used. Between participants, we counterbalanced the presentation order of context switching: half the participants saw the AR-real condition followed by real-real, and the remaining saw real-real followed by AR-real. A 2×2 Latin square controlled the presentation order of context switching. Within this square, a 6×3 Latin rectangle controlled the presentation order of distance to real world text. Therefore, the presentation order of context switching and real-world text distance were evenly counterbalanced over each group of 12 participants. Within each participant, the presentation order of the remaining independent variables, focal distance and virtual text distance, was randomly permuted.

Each experimental condition was repeated 4 times. Therefore, each participant experienced 3 real world distances × 4 repetitions (Table 1) + 3 real world distances × 3 focal distances × 3 virtual text distances × 4 repetitions (Table 2) = 120 tasks. The conditions varied in the following order: (1) Context switching varied the slowest, switching once during the experiment. When it switched, the experimenter removed (or added) the left text monitor (Fig. 1). Participants wore the AR display throughout the experiment, in both the real-real and AR-real conditions. (2) Real world distance varied next; each setting required the experimenter to reposition the real-world monitor(s). (3) Within each real-world distance setting, the focal distance varied; changing this required the experimenter to adjust the focus slider on the AR display (Fig. 2). (4) Then, within each focal distance setting, the virtual distance changed. This was managed by the experimental control software and did not require any activity by the experimenter. (5) Finally, the current setting was repeated 4 times.

3.6 Dependent Variables

The study measured five dependent variables: number of subtasks completed (0–5), number of subtasks correct (0–5), undercount errors (1–3), overcount errors (1–3), and eye fatigue (1–7).

Number of Subtasks Completed: As discussed in Section 3.1 above, during each task participants had 25 seconds to complete up to 5 subtasks. The *number of subtasks completed* (or, subtask completion) depended on the experimental conditions and served as a dependent variable. The 25-second time limit was based on an informal pilot study, which showed that, over all experimental conditions, the mean time to complete 5 subtasks was 30 seconds. Therefore, the time limit imposed some performance pressure, which we believed would better differentiate the experimental conditions. In particular, a similar augmented reality study conducted by Gabbard [58] yielded an error rate of only 1.5%. Because we believed that errors would further differentiate the experimental conditions, we wanted a larger error rate for this experiment.

Number of Subtasks Correct: For each subtask, error was calculated as *error = participant target count – correct target count*, where each target count ranged from 0 to 3. When *error* = 0, the subtask was correct. The *number of subtasks correct* (also, *accuracy*) depended on the experimental conditions, and served as a complementary dependent variable to the number of subtasks completed.

Undercount and Overcount Errors: When error ≠ 0, the subtask was incorrect. Error < 0 meant that the participant counted fewer targets than what was shown in the right text block; this *undercount error* ranged from –1 to –3. Likewise, error > 0 meant that the participant counted more targets than what was shown; this *overcount error* ranged from 1 to 3.

Eye Fatigue: After completing each task, the participant was asked to subjectively rate eye fatigue by answering the following question, which was displayed as virtual text on the AR display: "Please rate the condition of your eyes". A 7-point bi-polar rating scale ranging from "very rested" to "very fatigued" was used. Participants responded by pressing the appropriate number key on the numeric keypad.

3.7 Participants

Twenty-four graduate and undergraduate students (14 male and 10 female) from Virginia Tech participated in the study. The mean age was 22.6 years, with a standard deviation of 2.5 years. Participants were compensated at the rate of $10 per hour or secured course credit for their participation. Participants were required to be less than 35 years in age, to have at least normal or corrected 20/25 far vision, to have at least normal or corrected 20/25 near vision, and to not have been previously diagnosed with attention deficit disorder. These requirements were included in the call for participation flyers and emails. Questions related to attention deficit disorder symptoms were also included in the pre-experiment questionnaire.

3.8 Procedure

Each participant was first administered the Runge Near Point Card [59] and Snellen Chart [60], which measured near and far visual acuity. Next, the Porta test [61] screened for right-eye dominance. Participants who did not satisfy the 20/25 near and far vision criteria or who were left-eye dominant were not allowed to participate in the study. Eligible participants next completed a pre-experiment questionnaire, which collected information regarding age, vision, depth perception, and attention deficit disorder symptoms. Finally, the experimental procedures were described.

The Nomad eyepiece was then positioned in front of the participant's right eye. Practice trials were performed at each of the three distances, for a total of 12 practice trials. Following the practice trials, participants then performed 120 tasks, as discussed in Section 3.5. To minimize carryover fatigue effects, after completing 4 repetitions of each experimental cell (Tables 1 and 2), participants were provided a rest break of 45 seconds. During this break, they were instructed to close their eyes and relax. This rest break also allowed experimenters to adjust the positions of the monitor(s) and/or change the AR display's focal distance.

Upon completion of the experiment, an informal interview was conducted to gather additional information. Each experimental session lasted about two hours.

4 Results & Discussion

Unless otherwise noted, analyses discussed below used one-tailed, paired *t*-tests, with a significance value of 0.05.

4.1 Post-Experiment Questionnaire and Interviews

After completing the experiment, feedback was gathered from participants through a post-experiment questionnaire and informal interviews. Participants were uncomfortable wearing the display, and 87.5% of the participants complained of some form of eye discomfort, including blurry vision, eyestrain and fatigue, dizziness, and scratchy eyes. They felt that their right eye, which wore the monocular display, became more fatigued than their left eye. Participants also complained of physical discomfort, such as headaches, from wearing the head-mounted display, and 41.6% of participants found the display too heavy to be worn for extended periods of time. Other problems included discomfort due to the display lens touching the nose, neck ache, difficulty adjusting the display to fit properly on the head, and awkwardness while using the display.

Participants also complained of problems focusing on the virtual text at the far distance. While all participants indicated that they could see the far distance text properly in the beginning, they perceived it getting blurry sometime after the start of the experiment, and had trouble reading the far distance text. Three participants said that they struggled to discern the different letters. All participants considered the far distance to virtual text to be more fatiguing than the near or medium distances.

4.2 Context Switching

4.2.1 Results

We hypothesized that task performance would be better when context switching was not required. We examined context switching by comparing cells where context switching occurred, but focal distance was held constant: we compared the real-real conditions R_1R_1, R_2R_2, and R_3R_3 (Table 1) to the AR-real conditions $V_1F_1R_1$, $V_2F_2R_2$, and $V_3F_3R_3$ (Table 2).

There was a significant effect of context switching on subtask completion at the far distance, $t(23) = 5.83$, $p < 0.001$ (Fig. 4, upper panel). Participants completed a greater number of subtasks when information was presented in the real-real control condition, as compared to the AR-real treatment condition. Also, there was a significant effect of context switching on subtask accuracy at the far distance, $t(23) = 5.32$, $p < 0.001$ (Fig. 4, bottom panel). Participants were more accurate when information was presented in the real-real control condition, as compared to the AR-real treatment condition. Therefore, context switching had a negative impact on per-

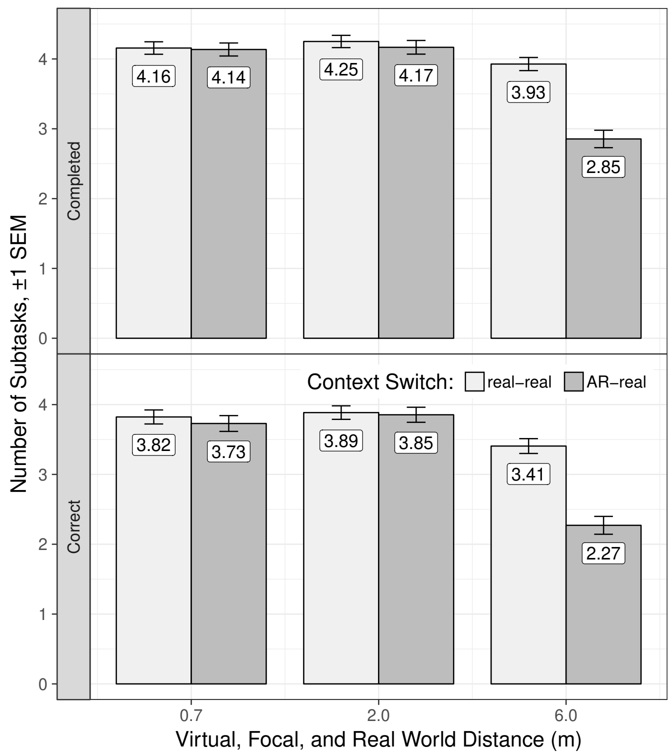


Fig. 4. Context switching between AR and real-world visual information at 6 m resulted in significantly fewer completed subtasks (upper panel) and less accurate performance (lower panel), as compared to the 0.7 m and 2 m viewing conditions. Data from Tables 1 and 2: *real-real* = R_1R_1, R_2R_2, R_3R_3; *AR-real* = $V_1F_1R_1$, $V_2F_2R_2$, $V_3F_3R_3$.

formance when information was presented at the far distance, but not when it was presented at either the near and medium distances.

There was also a significant effect of context switching on eye fatigue at all distances (Fig. 5): 0.7 meters: $t(23) = 3.32$, $p = 0.001$; 2 meters: $t(23) = 2.69$, $p = 0.006$; and 6 meters: $t(23) = 2.48$, $p = 0.011$. At all distances, participants found task performance less fatiguing when information was presented in the real-real control condition, as compared to the AR-real treatment condition. In addition, there was a positive correlation between trial number and eye fatigue: Fig. 6 shows the eye fatigue ratings by trial number, split according to whether participants saw the AR-real trials followed by real-real (upper panel), or the real-real trials followed by AR-real (lower panel). In this graph, it is easy to see that participants performed 4 repetitions of each experimental condition, and generally entered the same fatigue rating for each repetition, which leads to almost all data points appearing in groups of 4. In addition, note that participants began the experiment by rating their eyes as "1 (very rested)"; the lack of error bars for the first 4 data points indicates that this was the unanimous rating from all 24 participants. Likewise, after the transition from AR-real to real-real, and vice-versa (red dotted lines in Fig. 6), participants unanimously rated their eyes as "1 (very rested)". During this transition, equipment was moved (Fig. 1), which paused the experiment for several minutes. In both the upper and lower panels of Fig. 6, for both the AR-real and real-real trials, there was an overall increase in reported eye fatigue over trial number. A linear model over the AR-real trials shows an average eye fatigue

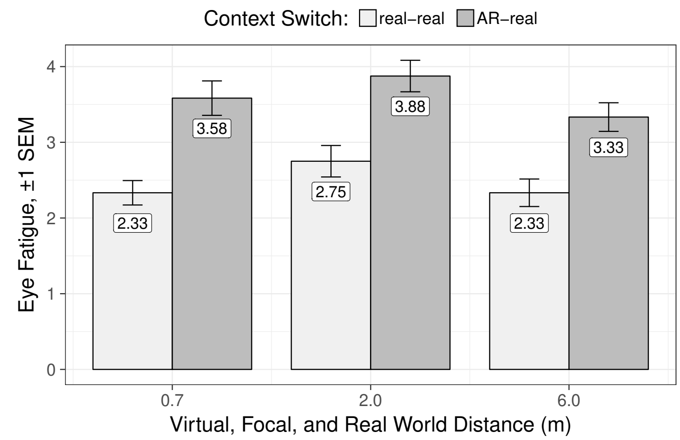


Fig. 5. Context switching between AR and real-world visual information resulted in significantly higher levels of eye fatigue at all distances. Data from Tables 1 and 2: *real-real* = R_1R_1, R_2R_2, R_3R_3; *AR-real* = $V_1F_1R_1$, $V_2F_2R_2$, $V_3F_3R_3$.

rating increase of 1.6 for the upper panel, explaining $R^2 = 5.5\%$ of the observed variation, and an increase of 1.8 for the lower panel, explaining $R^2 = 9.5\%$ of the observed variation.

4.2.2 Discussion

There was no difference between real-real and AR-real task performance when information was displayed at 0.7 and 2 meters. However, results revealed that at 6 meters, participants completed fewer tasks and were also less accurate when context switching (Fig. 4). For both AR and real text, the text size at 6 meters, although smaller than the text size at 0.7 and 2 meters, was well above the legibility standard of 22 arc minutes (ANSI [57]), and hence it seems likely that the size itself was not the only cause for the performance degradation at 6 meters. In examining the 6-meter context-switching case, we note that participants reported clear vision at the beginning of the experiment, and then blurry vision as the experiment progressed. We expect that this blurry vision was caused by visual fatigue, resulting from repeatedly changing eye focus to accommodate to 6 meters. For most participants, this distance was likely different from the resting point of accommodation (Iavecchia et al. [14]), and hence should have caused the greatest strain of the eye's oculomotor mechanism (Lambooij et al. [7]).

Compared to the 6-meter real-real condition, the additional performance decrements in the 6-meter AR-real condition could be caused by the monocular, laser-based virtual retinal display. Specifically, laser-based displays exhibit *speckle*, or spot-to-spot intensity fluctuations, that degrades image quality and thus can be troublesome for text and graphics with small visual footprints (Chellappan, Erden, and Urey [62]). Generally speaking, we empirically noted the effect of speckle in the AR display, but did not find it debilitating. However, we did not anticipate the potentially compounding effect of fatigue and speckle. Thus, as participants became fatigued, the effect of speckle on the legibility of 6-meter text could have become more pronounced. In addition, the monocular nature of the display removes the human visual system's ability to leverage binocular disparity to drive accommodation. In the real-real condition, binocular disparity drives vergence, which in turn drives accommodation.

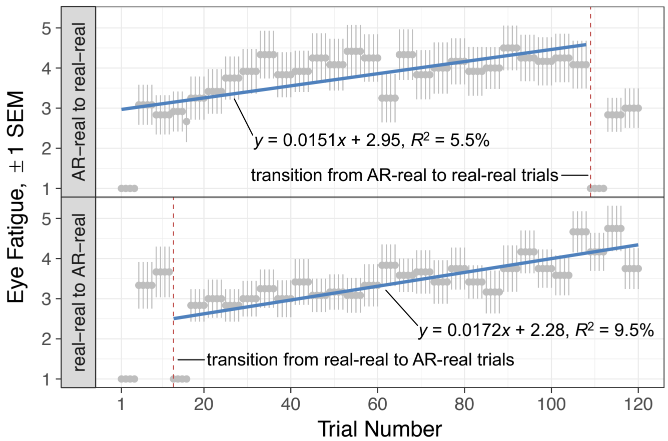


Fig 6. Context switching during the AR-real trials resulted in a linear increase in reported eye fatigue. Half of the observers saw the AR-real trials followed by the real-real trials (upper panel), while the remaining observers saw real-real trials followed by AR-real trials (lower panel). Data from Tables 1 and 2.

However, in the monocular AR-real condition, there is no AR-based binocular disparity to drive vergence, and thus no binocular vergence to drive accommodation. Instead, the accommodation itself would drive any observed vergence (Semmlow and Hung [63]). Over time, this effect could have also interacted with fatigue.

Therefore, the AR display produced text that was easily read at closer distances, but the 6-meter distance resulted in relatively poor legibility, despite our efforts to make the text large enough to be easily readable at all distances. However, context switching did have a negative effect on eye fatigue ratings at *all three distances*: participants rated the AR-real condition more fatiguing than the real-real condition (Fig. 5). This result is explained by the steady increase in reported eye fatigue over the AR-real trials (Fig. 6): Although participants were given a 45-second rest break between experimental conditions, which occurred between the 4-trial blocks shown in Fig. 6, and although participants were instructed to close their eyes and relax during this 45 second break, Fig. 6 illustrates that this was not enough time for their eyes to return to a completely rested state.

4.3 Focal Distance Switching

4.3.1 Results

As discussed in Sections 1 and 2, in most optical see-through AR use cases, users must switch focal distance from real world to virtual objects, and it was hypothesized that this switching would result in lower performance. This was tested by comparing conditions where focal distance switching was not required to conditions where it was required. Therefore, we compared the mean of the hatched cells in Table 2 ($V_1F_1R_1$, $V_1F_2R_2$, $V_1F_3R_3$, $V_2F_1R_1$, $V_2F_2R_2$, $V_2F_3R_3$, $V_3F_1R_1$, $V_3F_2R_2$, $V_3F_3R_3$), where focal distance and real distance matched, to the mean of the remaining 18 cells, where focal distance and real distance did not match (Fig. 7). There was a significant effect of focal distance switching on both subtask completion, $t(23) = 2.06$, $p = 0.026$ (Fig. 7, left), and subtask accuracy, $t(23) = 2.17$, $p = 0.021$ (Fig. 7, right). Participants completed more subtasks, and were more accurate, when focal distance switching was not required.

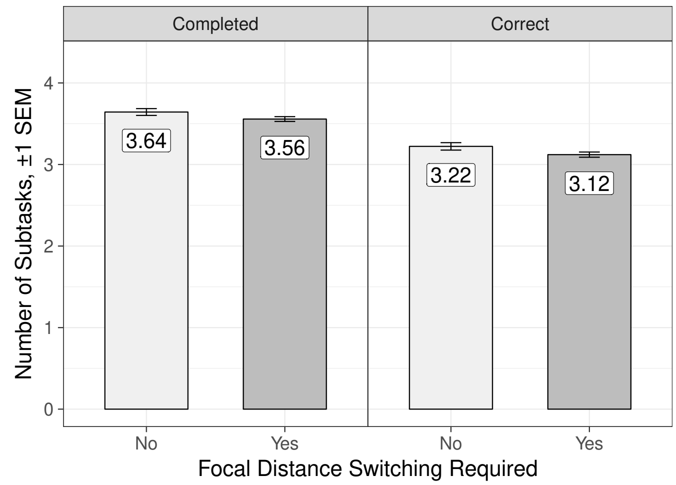


Fig. 7. Focal distance switching between AR and real world visual information resulted in significantly fewer completed subtasks (left) and less accurate performance (right). These findings suggest there is a cost for focal distance switching when visually integrating AR and real-world information. Data from Table 2: Focal Distance Switching Required: *No* = hatched cells; *Yes* = remaining cells.

4.3.2 Discussion

In the experiment, focal distance switching required accommodating to one distance to read real text, and then accommodating to a different distance to read virtual text. Focal distance switching reduced both the number of completed subtasks, and the subtask accuracy (Fig. 7).

Changing accommodation takes time, and this can explain the reduced number and accuracy of subtasks. Up to the age of 20, with repetitive visual stimuli, a human eye takes about 360 milliseconds to accommodate from far to near distances, and about 380 milliseconds to accommodate from near to far distances (Stark, Takahashi, and Zames [64], Campbell and Westheimer [65]). For non-repetitive stimuli, accommodation times can be longer, up to 425 milliseconds (Carter [66]). Beyond the age of 20, far-to-near accommodation time remains constant, but near-to-far accommodation becomes slower (Kasthurirangan and Glasser [67]). In the experiment, participants had to first accommodate to the distance of the left text, and then change accommodation to the distance of the right text. It is safe to assume that most of these accommodation changes took more than 350 milliseconds.

In addition, because the task was time-pressured, participants were likely still accommodating as they began to scan the right text. Furthermore, because the participants were all native readers of written English, they likely scanned for target letters using the standard reading direction of left-to-right within each line, and top-to-bottom between the lines. Therefore, participants likely began scanning the right text before the letters were in sharp focus. If these assumptions are correct, they predict that, when focal distance switching was required, participants were more likely to miss target letters in the first line of the right text, as opposed to target letters in the second or third lines.

This hypothesis was tested by analyzing the error subtasks, were error ≠ 0 (Section 3.6). Among the 9295 collected subtasks, 1118 involved errors, of which 119 were overcounts (error > 0), and 999 were undercounts (error < 0). When there

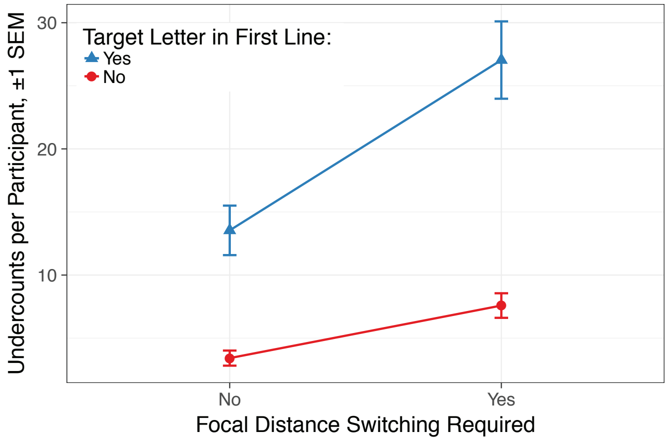


Fig 8. Participants undercounted more letters when a target letter appeared in the first line of text, and when focal distance switching was required. This indicates that when participants had to switch focal distances, they began scanning the first line for a target letter before their eyes had finished accommodating. This made the text blurry, and therefore they were more likely to miss the target letter. Data from Table 2: Focal Distance Switching Required: *No* = hatched cells; *Yes* = remaining cells.

is focal distance switching and a target letter in the first line of text, the hypothesis predicts that undercount errors will be higher. Fig. 8 shows the mean undercounts per participant, analyzed according to the interaction between whether focal distance switching was required and whether there was a target letter in the first line of text. An ANOVA of this data reveals strong main effects of distance switching, $F(1,23) = 46.22$, $p < 0.001$, target letter in the first line, $F(1,23) = 68.29$, $p < 0.001$, and the interaction between the two, $F(1,23) = 24.95$, $p < 0.001$. These results strongly support the hypothesis: Because of the time pressure to complete as many subtasks as possible, participants were more likely to miss a target letter in the first line. Furthermore, when distance switching was required, participants began scanning the first line of text for target letters before their eyes had time to fully accommodate to that distance, which made them even more likely to miss a target letter.

4.4 Matched vs. Mismatched Distances

4.4.1 Results

We anticipated improved performance when all the distances matched, as opposed to when any distance differed from the others. We examined this hypothesis with two analyses.

First, we conducted a 3×3×3 repeated-measures analysis of variance (ANOVA) across all conditions in Table 2 (Fig. 9). There were no significant main effects or interactions with focal distance (F's vary between 0.096 and 1.4), and therefore Fig. 9 shows the results plotted by real world distance and virtual distance. There was a strong main effect of virtual distance on both subtask completion, $F(2,46) = 80.20$, $p < 0.001$, and accuracy, $F(2,46) = 139.68$, $p < 0.001$. This effect was analyzed using Tukey's Honestly Significant Difference Test at a significance level of 0.05. As clearly shown in Fig. 9, for both subtask completion and accuracy, Tukey's Test revealed significantly worse performance for virtual distance at 6 meters than at 0.7 or 2 meters, and no significant difference between virtual distances of 0.7 and 2 meters. In addition, there was a

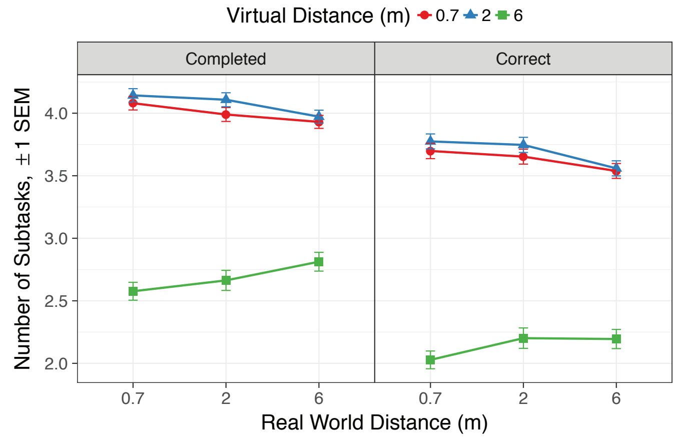


Fig. 9. Observers completed more subtasks with higher accuracy at the 0.7 m and 2 m virtual distances than they did at 6 m. There was also an interaction between virtual and real world distance. Data from Table 2.

significant interaction between real world distance and virtual distance for subtask completion, $F(4,92) = 2.64$, $p = 0.0384$, and the same interaction approached significance for subtask accuracy, $F(4,92) = 1.99$, $p = 0.102$. However, these interactions are weak compared to the main effect of virtual distance, and we doubt they are reliable. There were no other main effects or interactions.

For the second analysis, the conditions where all distances matched were directly compared to the conditions where the distances did not match. Therefore, in Table 2, the conditions where all distances matched—cells $V_1F_1R_1$, $V_2F_2R_2$, $V_3F_3R_3$—were compared to the remaining 24 cells, where at least one distance differed from the others. Fig. 10 gives the result. There is a significant effect of matched vs. mismatched distances on both subtask completion, $t(23) = 2.02$, $p = 0.028$ (Fig. 10, left, distances = all), and subtask accuracy, $t(23) = 1.783$, $p = 0.043$ (Fig. 10, right, distances = all). When the distance to real world text, distance to virtual text, and focal depth matched, participants performed better than when the distances did not match.

However, because of the very powerful effect of virtual text distance at 6 meters (Fig. 9), we were interested in determining if this effect still held when we excluded the 6-meter distance from the analysis. Therefore, in Table 2, the cells $V_1F_1R_1$ and $V_2F_2R_2$, which represent matched near and medium distances, were compared to the remaining 16 near and medium cells with mismatched distances. There is still a significant effect of matched vs. mismatched conditions on task completion, $t(23) = 1.73$, $p = 0.048$ (Fig. 10, left, distances = near, medium). Participants completed a greater number of tasks in the matched condition than in the mismatched condition. However, the effect of matched vs. mismatched conditions on accuracy was only close to significance, $t(23) = 1.79$, $p = 0.076$ (Fig. 10, right, distances = near, medium).

4.4.2 Discussion

We found improved task completion and accuracy when all distances matched, as compared to when distances did not match (Fig. 10, distances = all). There was also a very large negative effect on performance for virtual text at 6 meters (Fig. 9). However, when the 6-meter distance was eliminated, task completion and accuracy were still improved when all

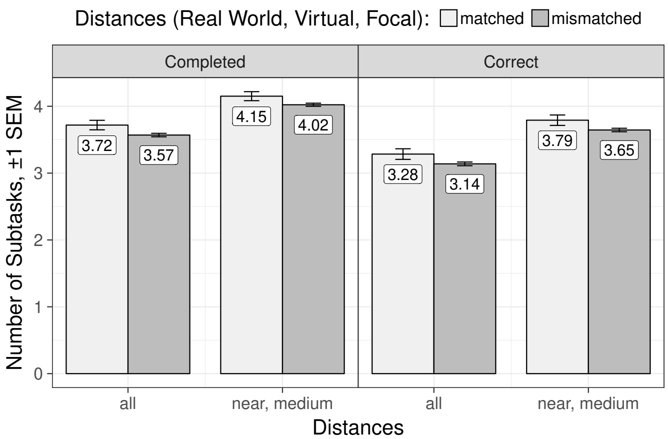


Fig 10. Observers completed more tasks with higher accuracy when distances to the virtual, focal, and real-world distances matched. These findings suggest that costs for focal distance switching may be exasperated by incongruent virtual distance cues. Data from Table 2: Distances (Virtual, Focal, Real World): *matched* = $V_1F_1R_1$, $V_2F_2R_2$, $V_3F_3R_3$; *mismatched* = remaining cells.

distances matched (Fig. 10, distances = near, medium).

As discussed in Section 4.2.2 above, at 6 meters there was a significant effect of context switching, where participants performed better in the real-real condition than the AR-real condition (Fig. 4). In addition, participants complained that at the far distance, virtual text became blurry. The analysis here further measures the poor performance of virtual text at 6 meters. As previously discussed, the combined effects of relatively poor display resolution, smaller text size as compared to the near and medium distances, the excessive need to accommodate, increasing visual fatigue, and possibly also the fact that the display was monocular and exhibited laser speckle, are the likely causes of the poor performance for virtual text at 6 meters.

In addition, in Section 4.3 above, we found that focal distance switching significantly reduced performance (Fig. 7). The analysis in Section 4.3 is consistent with the hypothesis that the reason for this reduced performance is that it took additional time to accommodate between focal distances when distance switching was required. Furthermore, the distance switching resulted in increasing fatigue throughout the experiment (Fig. 6). This analysis also likely explains the decreased performance for mismatched distances in Fig. 10: focal distance switching is not required for the matched distances, while it is required for the mismatched distances. Furthermore, the size of the effects in Figs. 10 and 7 are similar. Therefore, the analysis in Fig. 10 is likely another measurement of the effect in Fig. 7.

5 LESSONS LEARNED

Based on the results of this study, we offer the following lessons learned, which may be of use to AR researchers, application developers, and display producers:

Font Legibility at Optical Infinity: There was no difference in user task performance between the near (0.7 meter) and medium (2 meters) virtual text distances, but poor performance at the far distance of 6 meters, which corresponds to optical infinity. If virtual text will be displayed at distances greater than or equal to 6 meters, then sufficiently large font sizes should be used, larger than the 22 arc minutes of visual angle called for by the ANSI standard [57]. Given the Nomad's display resolution of 800x600 pixels, smaller fonts sizes at 6 meters result in poor text legibility. Hence displays with higher resolution and high quality spatial light modulation may be required for text displayed at far distances, such as in outdoor augmented reality, where real world objects of interest may be far from the user.

In addition, a motivation for using laser scanning AR displays is that it is one of the few, perhaps the only, display technology that can match the extreme luminance changes that occur when operating outdoors in conditions that might range from starless, moonless nighttime to bright sunshine on snow or white sand (Gabbard, Swan, and Hix [68]). Although laser speckle may well have been a problem in this experiment, and potentially limits the technology's use in AR displays, the laser-based display community continues to develop methods for reducing speckle (Chellappan, Erden, and Urey [62]).

For comparison with newer monocular AR displays, the Google Glass has a resolution of 640×360 pixels, and the Vuzix M100 432×240 pixels. Generally, unlike binocular displays, which are designed for wide fields of view and visually conformal graphics, monocular displays are designed for industrial applications, where an unobstructed view of the real world is important, and the graphics, while relevant to the location and task, are not necessarily presented conformally (Schwerdtfeger et al. [4], Huckauf et al [44], UBiMAX [48]). Therefore, they have a lower resolution and cover a smaller field of view.

Conformal Text at Same Focal Distance as Real-World Objects: Performance was better when the focal distance to virtual text matched the distance to real text. Therefore, if virtual text is conformal to a real-world object, and the user's task requires scanning both the real-world object and the virtual text, then the virtual text should be presented at the same focal distance as the real-world object. This will minimize changing accommodation and convergence, and therefore put the least amount of strain on the eye's oculomotor mechanism.

Switching Context Between Real and Virtual Information: Performance was better when all information was available in the real world, as opposed to when users had to integrate information between the real and virtual spaces. Although the purpose of AR is to provide virtual information in real-world contexts, this study suggests there will still be a performance penalty when using AR.

Head-Mounted Display Ergonomics: The Nomad's weight of 226.8 grams is too heavy to be worn for long periods of time. In addition, the Nomad did not properly fit all participants, which resulted in additional physical discomfort and fatigue. Better adjustments are required, so that head-mounted displays can more comfortably accommodate different head sizes. We note that the weight and ergonomics of the Google Glass and Vuzix M100, both newer displays, are significantly improved.

6 Conclusions and Future Work

In conclusion, this research contributes one of the first known empirical user studies to examine switching context and focal distance between a real and an AR environment. Both context and focal distance switching resulted in significantly reduced performance. In addition, repeatedly performing the task caused visual fatigue to steadily increase. Performance was poor for virtual text presented at optical infinity. Overall, the results show that context switching and focal distance switching are important AR user interface design issues.

Of course, any single laboratory experiment has limitations, and addressing these leads to the following ideas for future work:

Ecologically Valid Tasks: This study used a visual scanning task consisting of random character strings. Although the task was inspired by reading AR text labels, and was carefully designed to meet the study's research goals, it does not replicate an actual AR use case. Future studies should seek to replicate these findings with more ecologically valid tasks.

Interaction Between Context and Distance Switching: Because the experimental design took about two hours to complete, and involved significant fatigue effects, the design did not fully cross the conditions of context and distance switching; note the blank cells in Table 1. However, measuring the interaction between context and distance switching would add important additional information to what is reported here. A future study should use a design that completely crosses context and distance switching.

Virtual Text Distance: This study used a monocular display, and therefore virtual text distance was encoded by the depth cues of relative size and motion parallax. However, the study found poor performance at a virtual text distance of 6 meters (Fig. 9), where the relative size cue made the text small, and, when combined with the other factors discussed above in Section 4.2.2, difficult to read. In addition, although motion parallax is a powerful depth cue (Rogers and Graham [52]), the task did not involve systematic user movement. Future studies should therefore examine virtual text strings that are the same size in visual angle, but appear at different focal and virtual distances.

Binocular AR Displays: This study used a monocular Nomad AR display, because it provided—uniquely—a mechanism to adjust the display's focal distance. As discussed above in Section 3, there is a long history of using monocular displays, especially for military and other specialized tasks (e.g., Hughes et al. [46], Huckauf et al [44], Schwerdtfeger et al. [4], Tümler et al. [47], UBiMAX [48]). Therefore, although the results are relevant for monocular displays, it is not clear how relevant the results are for binocular displays. In particular, binocular displays also provide a stereo disparity cue, which was necessarily lacking here. Therefore, future similar studies should use a binocular AR display with an adjustable focal distance[2].

Acknowledgments

This material is based upon work supported by the Office of Naval Research, under awards to J.L. Gabbard, D.S. Hix, and J.E. Swan II. It is also based upon work supported by the National Science Foundation, under awards IIS-1320909 and IIS-1018413, to J.E. Swan II. The authors would like to thank Dr. Deborah S. Hix, Dr. Robert S. Schulman, and Mr. John Lucas, for their help with this research.

References

[1] M. A. Livingston, L. J. Rosenblum, D. G. Brown, G. S. Schmidt, S. J. Julier, Y. Baillot, J. E. Swan II, Z. Ai, P. Maassel, "Military applications of augmented reality," in *Handbook of augmented reality*, ed: Springer, 2011, pp. 671-706.

[2] H. Regenbrecht, G. Baratoff, and W. Wilke, "Augmented reality projects in the automotive and aerospace industries," *Computer Graphics and Applications, IEEE,* vol. 25, pp. 48-56, 2005.

[3] E. Barsom, M. Graafland, and M. Schijven, "Systematic review on the effectiveness of augmented reality applications in medical training," *Surgical Endoscopy,* pp. 1-10, 2016.

[4] B. Schwerdtfeger, R. Reif, W. A. Günthner, G. Klinker, D. Hamacher, L. Schega, *et al.*, "Pick-by-Vision: A First Stress Test," in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2009*, Orlando, FL, USA, 2009, pp. 115–124.

[5] S. Webel, U. Bockholt, T. Engelke, N. Gavish, M. Olbrich, and C. Preusche, "An augmented reality training platform for assembly and maintenance skills," *Robotics and Autonomous Systems,* vol. 61, pp. 398-403, 2013.

[6] C. Neveu, T. Blackmon, and L. Stark, "Evaluation of the effects of a head-mounted display on ocular accommodation," *Presence: Teleoperators and Virtual Environments,* vol. 7, pp. 278-289, 1998.

[7] M. Lambooij, M. Fortuin, I. Heynderickx, and W. IJsselsteijn, "Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review," *Journal of Imaging Science and Technology,* vol. 53, p. 030201, 2009.

[8] R. Miller, R. G. Pigion, M. F. Wesner, and J. G. Patterson, "Accommodation fatigue and dark focus: The effects of accommodation-free visual work as assessed by two psychophysical methods," *Perception & psychophysics,* vol. 34, pp. 532-540, 1983.

[9] S. Roscoe, "Training good eyes to see better," *Displays,* vol. 6, pp. 218-220, 1985.

[10] G. K. Edgar, J. C. Pope, and I. R. Craig, "Visual accomodation problems with head-up and helmet-mounted displays?," *Displays,* vol. 15, pp. 68-75, 1994.

[11] W. Wu, I. Tošić, K. Berkner, and N. Balram, "Depth-disparity calibration for augmented reality on binocular optical see-through displays," in *Proceedings of the 6th ACM Multimedia Systems Conference*, 2015, pp. 120-129.

[12] K. Carnegie and T. Rhee, "Reducing Visual Discomfort with HMDs Using Dynamic Depth of Field," *Computer Graphics and Applications, IEEE,* vol. 35, pp. 34-41, 2015.

[13] D. A. Goss and R. W. West, *Introduction to the Optics of the Eye*: Butterworth-Heinemann Medical, 2001.

[14] J. H. Iavecchia, H. P. Iavecchia, and S. N. Roscoe, "Eye accommodation to head-up virtual images," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 30, pp. 703-712, 1988.

[2]However, this will require a binocular, optical see-through AR display with adjustable focal distance. The authors have never seen a commercially available example of such a device.

[15] 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*, vol. 8, pp. 33-33, 2008.

[16] D. A. Owens, "The Mandelbaum Effect: Evidence for an Accommodative Bias Toward Intermediate Viewing Distances," *Journal of the Optical Society of America,* vol. 69, pp. 646–652, 1979.

[17] S. Roscoe, "Bigness Is in the Eye of the Beholder," *Human Factors,* vol. 27, pp. 615–636, 1985.

[18] S. Roscoe, "When Day Is Done and Shadows Fall, We Miss the Airport Most of All," *Human Factors,* vol. 21, pp. 721–731, 1979.

[19] V. V. Krishnan and L. Stark, "A Heuristic Model for the Human Vergence Eye Movement System," *IEEE Transactions on Biomedical Engineering,* vol. 24, pp. 44–49, 1977.

[20] M. Mon-Williams and J. R. Tresilian, "Ordinal Depth Information from Accommodation?," *Ergonomics,* vol. 43, pp. 391–404, March 2000.

[21] G. A. Myers and L. Stark, "Topology of the Near Response Triad," *Ophthalmic and Physiological Optics,* vol. 10, pp. 175–181, 1990.

[22] E. Kruijff, J. E. Swan II, and S. Feiner, "Perceptual Issues in Augmented Reality Revisited," in *Technical Papers, IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2010)*, Seoul, Korea, 2010, pp. 3–12.

[23] J. P. Wann, S. Rushton, and M. Mon-Williams, "Natural Problems for Stereoscopic Depth Perception in Virtual Environments," *Vision Research,* vol. 35, pp. 2731–2736, 1995.

[24] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A Stereo Display Prototype with Multiple Focal Distances," *ACM Transactions on Graphics,* vol. 23, pp. 804–813, 2004.

[25] K. J. MacKenzie, R. A. Dickson, and S. J. Watt, "Vergence and Accommodation to Multiple-Image-Plane Stereoscopic Displays: "Real World" Responses with Practical Image-Plane Separations?," *Journal of Electronic Imaging,* vol. 21, 2012.

[26] S. J. Watt, K. Akeley, M. O. Ernst, and M. S. Banks, "Focus Cues Affect Perceived Depth," *Journal of Vision,* vol. 5, pp. 834–862, 2005.

[27] H. T. Nefs, "Depth of Field Affects Perceived Depth-width Ratios in Photographs of Natural Scenes," *Seeing and Perceiving,* vol. 25, pp. 577–595, 2012.

[28] D. Vishwanath and E. Blaser, "Retinal Blur and the Perception of Egocentric Distance," *Journal of Vision,* vol. 10, pp. 1–16, 2010.

[29] R. T. Held, E. A. Cooper, J. F. O'Brien, and M. S. Banks, "Using Blur to Affect Perceived Distance and Size," *ACM Transactions on Graphics,* vol. 29, pp. 1–16, March 2010.

[30] H. Davson, *Physiology of the Eye*: Elsevier, 2012.

[31] O. Ostberg, "Accommodation and visual fatigue in display work," *Displays,* vol. 2, pp. 81-85, 1980.

[32] L. O'Hare, T. Zhang, H. T. Nefs, and P. B. Hibbard, "Visual Discomfort and Depth-of-Field," *i-Perception,* vol. 4, pp. 156–169, 2012.

[33] C. A. Johnson, "Effects of luminance and stimulus distance on accommodation and visual resolution," *Journal of the Optical Society of America,* vol. 66, pp. 138-142, 1976.

[34] D. J. Weintraub and M. Ensing, *Human Factors Issues in Head-Up Display Design: the Book of HUD*. Wright-Patterson AFB, Dayton, OH, USA: Crew System Ergonomics Information Analysis Center (CSERIAC), 1992.

[35] Y. Inuzuka, Y. Osumi, and H. Shinkai, "Visibility of head up display (HUD) for automobiles," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1991, pp. 1574-1578.

[36] H. Kato, H. Ito, J. Shima, M. Imaizumi, and H. Shibata, "Development of hologram head-up display," SAE Technical Paper 0148-7191, 1992.

[37] S. Okabayashi, M. Sakata, and T. Hatada, "How an automotive head-up display affects a driver's ability to recognize the forward view," *Proceedings of the society for information display,* vol. 32, pp. 31-37, 1991.

[38] D. J. Weintraub, R. F. Haines, and R. J. Randle, "The utility of Head-Up Displays-Eye-focus vs decision times," 1984.

[39] T. Albrecht, "On the use of transparent rear projection screens to reduce head-down time in the air-traffic control tower," *Human Performance, Situation Awareness and Automation: Current Research and Trends: HPSAA II,* vol. 2, p. 195, 2004.

[40] J. Wolffsohn, G. Edgar, and N. McBrien, "The effect of viewing a car head-up display on ocular accommodation and response times," *Vision in Vehicles,* vol. 6, pp. 143-151, 1998.

[41] D. R. Tufano, "Automotive HUDs: The overlooked safety issues," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 39, pp. 303-311, 1997.

[42] J. L. Gabbard, G. M. Fitch, and H. Kim, "Behind the Glass: Driver Challenges and Opportunities for AR Automotive Applications," *Proceedings of the IEEE,* vol. 102, pp. 124-136, 2014.

[43] K. Bark, C. Tran, K. Fujimura, and V. Ng-Thow-Hing, "Personal Navi: Benefits of an Augmented Reality Navigational Aid Using a See-Thru 3D Volumetric HUD," in *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2014, pp. 1-8.

[44] A. Huckauf, M. H. Urbina, I. Böckelmann, L. Schega, R. Mecke, J. Grubert, F. Doil, J. Tümler, "Perceptual issues in optical-see-through displays," in *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, 2010, pp. 41-48.

[45] D. A. Owens and K. Wolf-Kelly, "Near work, visual fatigue, and variations of oculomotor tonus," *Investigative Ophthalmology & Visual Science,* vol. 28, pp. 743-749, 1987.

[46] R. L. Hughes, L. R. Chason, and J. C. Schwank, "Psychological considerations in the design of helmet-mounted displays and sights: Overview and annotated bibliography," ed: Aerospace Medical Research Laboratory, report AD-770 993, 1973.

[47] J. Tumler, F. Doil, R. Mecke, G. Paul, M. Schenk, E. A. Pfister, A. Huckauf, I. Böckelmann, A. Roggentin, "Mobile augmented reality in industrial applications: Approaches for solution of user-related issues," in *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*, 2008, pp. 87-90.

[48] UBiMAX. (2017, March 24, 2017). *UBiMAX Home Page*. Available: http://www.ubimax.de/index.php/en/

[49] J. L. Gabbard, J. E. Swan II, and A. Zarger, "Color Blending in Outdoor Optical See-through AR: The Effect of Real-world Backgrounds on User Interface Color," in *IEEE Virtual Reality,* Orlando, FL, 2013, pp. 157–158.

[50] J. Kelso, L. E. Arsenault, S. G. Satterfield, and R. D. Kriz, "Diverse: A framework for building extensible and reconfigurable device independent virtual environments," in *IEEE Virtual Reality*, Orlando, Florida, USA, 2002, pp. 183-190.

[51] J. E. Cutting and P. M. Vishton, "Perceiving layout and knowing distances: The integration, relative potency and contextual use of different information about depth," in *Handbook of Perception and Cognition.* vol. 5: Perception of Space and Motion, W. Epstein and S. Rogers, Eds., 1995, pp. 69-117.

[52] B. J. Rogers and M. Graham, "Similarities between motion parallax and stereopsis in human depth perception," *Vision Research,* vol. 22, pp. 261-270, 1982.

[53] J. W. McCandless, S. R. Ellis, and B. D. Adelstein, "Localization of a time-delayed, monocular virtual object superimposed on a real environment," *Presence,* vol. 9, pp. 15-24, 2000.

[54] W. Jaschinski-Kruza, "Eyestrain in VDU users: viewing distance and the resting position of ocular muscles," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 33, pp. 69-83, 1991.

[55] J. Norman and S. Ehrlich, "Visual accommodation and virtual image displays: Target detection and recognition," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 28, pp. 135-151, 1986.

[56] N. M. Simonelli, "Apparent size and visual accommodation under day and night conditions," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1979, pp. 374-378.

[57] Human Factors Society and American National Standards Institute, *American National Standard for Human Factors Engineering of Visual Display Terminal Workstations*: Human Factors Society, 1988.

[58] J. L. Gabbard, "Usability Engineering of Text Drawing Styles in Augmented Reality User Interfaces," Ph.D. Dissertation, Department of Computer Science, Virginia Tech, Blacksburg, VA, 2008.

[59] P. E. Runge, "Eduard Jaeger's Test-Types (Schrift-Scalen) and the Historical Development of Vision Tests," *Transactions of the American Ophthalmological Society,* vol. 98, p. 375, 2000.

[60] H. Snellen, "Optotypi ad Visum Determinandum (Letterproeven tot Bepaling der Gezichtsscherpte; Probebuchstaben zur Bestimmung der Sehschaerfe)," *Utrecht, The Netherlands: Weyers,* 1862.

[61] G. Della Porta, *De refractione optices parte: libri novem*: Ex officina Horatii Salviani, apud Jo. Jacobum Carlinum, & Antonium Pacem, 1593.

[62] K. V. Chellappan, E. Erden, and H. Urey, "Laser-Based Displays: A Review," *Applied Optics,* vol. 49, pp. F79-F98, 2010.

[63] J. Semmlow and G. K. Hung, "The Near Response: Theories of Control," in *Vergence Eye Movements: Basic & Clinical Aspects*, S. M. Schor and K. J. Ciuffreda, Eds., Butterworth, 1983, pp. 175-195.

[64] L. Stark, Y. Takahashi, and G. Zames, "Nonlinear servoanalysis of human lens accommodation," *IEEE Transactions on Systems Science and Cybernetics,* vol. 1, pp. 75-83, 1965.

[65] F. Campbell and G. Westheimer, "Dynamics of accommodation responses of the human eye," *The Journal of physiology,* vol. 151, pp. 285-295, 1960.

[66] J. H. Carter, "A Servoanalysis of the Human Accomodative Mechanism," Indiana University, 1962.

[67] S. Kasthurirangan and A. Glasser, "Age Related Changes in Accommodative Dynamics in Humans," *Vision Research,* vol. 46, pp. 1507-1519, 2006.

[68] J. L. Gabbard, J. E. Swan, II, and D. Hix, "The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality," *Presence: Teleoperators & Virtual Environments,* vol. 15, pp. 16–32, 2006.

Joseph L. Gabbard received the B.S., M.S., and Ph.D. degrees in computer science from Virginia Tech, Blacksburg, VA, USA, in 1994, 1997, and 2008, respectively. He also holds the B.A. degree in sociology from Virginia Tech. From 2005 to 2011, he was a Research Associate with the Virginia Tech Center for Human–Computer Interface, and from 2011 to 2013, he held an appointment as a Research Assistant Professor at the Virginia Bioinformatics Institute. Since 2013, he has been an Associate Professor with the Grado Department of Industrial and Systems Engineering, specializing in human factors. He is the coauthor of four book chapters and 45 articles. His research interests include perception, cognition, and usability in augmented and virtual reality; usability and cognitive engineering; human–computer interaction; user interface and visualization challenges for emerging technologies; and empirical methods (experimental design, user-centered design and evaluation, and statistical analysis). Dr. Gabbard was a recipient of the Alan Berman Publication Award for the publication “Resolving multiple occluded layers in augmented reality,” and was awarded the IEEE VR “Best Paper Award” in 1999 for the publication “User-centered design and evaluation of a real-time battlefield visualization virtual environment.”

Divya Gupta Mehra received the B.E. in Chemical Engineering from the University of Mumbai in 2002, and an M.S. degree in Industrial and Systems Engineering from Virginia Tech, in 2004. She started her professional career at Nokia, where she worked as a Usability Scientist from 2004–2008. She subsequently went into Product Marketing and worked with some of the largest mobile phone carriers to improve the user experience of mobile devices. From 2010–2012, she was a part of the sales team in the U.K. working on establishing relationships with the European operators. In 2013, she joined HERE (formerly Navteq), and is currently working in the mapping and location intelligence space as manager of customer and market development.

J. Edward Swan II is Professor and Interim Department head of Computer Science and Engineering at Mississippi State University. He holds a B.S. (1989) degree in computer science from Auburn University and M.S. (1992) and Ph.D. (1997) degrees in computer science from Ohio State University, where he studied computer graphics and human-computer interaction. Before joining Mississippi State University in 2004, Dr. Swan spent seven years as a scientist at the Naval Research Laboratory in Washington, D.C. Dr. Swan’s research has been broad-based, centering on the topics of augmented and virtual reality, perception, human-computer interaction, human factors, empirical methods, computer graphics, and visualization. Currently, Dr. Swan is studying perception in augmented and virtual reality, including depth and layout perception and depth presentation methods, as well as empirical techniques for evaluating and validating visualizations. His research has been funded by the National Science Foundation, the Department of Defense, the National Aeronautics and Space Administration, the Naval Research Laboratory, and the Office of Naval Research. Dr. Swan is a member of ACM, IEEE, the IEEE Computer Society, and ASEE.