An Extended Analysis on the Benefits of Dark Mode User Interfaces in Optical See-Through Head-Mounted Displays

AUSTIN ERICKSON, University of Central Florida
KANGSOO KIM, University of Central Florida
ALEXIS LAMBERT, University of Central Florida
GERD BRUDER, University of Central Florida
MICHAEL P. BROWNE, SA Photonics
GREGORY F. WELCH, University of Central Florida

Light-on-dark color schemes, so-called "Dark Mode," are becoming more and more popular over a wide range of display technologies and application fields. Many people who have to look at computer screens for hours at a time, such as computer programmers and computer graphics artists, indicate a preference for switching colors on a computer screen from dark text on a light background to light text on a dark background due to perceived advantages related to visual comfort and acuity, specifically when working in low-light environments.

In this paper, we investigate the effects of dark mode color schemes in the field of optical see-through head-mounted displays (OST-HMDs), where the characteristic "additive" light model implies that bright graphics are *visible* but dark graphics are *transparent*. We describe two human-subject studies in which we evaluated a normal and inverted color mode in front of different physical backgrounds and different lighting conditions. Our results indicate that dark mode graphics displayed on the HoloLens have significant benefits for visual acuity, and usability, while user preferences depend largely on the lighting in the physical environment. We discuss the implications of these effects on user interfaces and applications.

CCS Concepts: • **Human-centered computing → Empirical studies in visualization**; • **Computing methodologies → Mixed / augmented reality**.

Additional Key Words and Phrases: Augmented Reality; Optical See-Through Display; Dark Mode; Eye Fatigue; Visual Acuity; User Experience

ACM Reference Format:
Austin Erickson, Kangsoo Kim, Alexis Lambert, Gerd Bruder, Michael P. Browne, and Gregory F. Welch. 2020. An Extended Analysis on the Benefits of Dark Mode User Interfaces in Optical See-Through Head-Mounted Displays. 1, 1 (March 2020), 23 pages. https://doi.org/10.1145/1122445.1122456??

Authors' addresses: Austin Erickson, University of Central Florida, 4000 Central Florida Blvd. Orlando, Florida, 32816, ericksona@knights.ucf.edu; Kangsoo Kim, University of Central Florida, 4000 Central Florida Blvd. Orlando, Florida, 32816, kangsoo.kim@ucf.edu; Alexis Lambert, University of Central Florida, 4000 Central Florida Blvd. Orlando, Florida, 32816, lamberta2@knights.ucf.edu; Gerd Bruder, University of Central Florida, 4000 Central Florida Blvd. Orlando, Florida, 32816, bruder@ucf.edu; Michael P. Browne, SA Photonics, 120 Knowles Dr. Los Gatos, California, 95032, m.browne@saphotonics.com; Gregory F. Welch, University of Central Florida, 4000 Central Florida Blvd. Orlando, Florida, 32816, welch@ucf.edu.

XXXX-XXXX/2020/3-ART $15.00
https://doi.org/10.1145/1122445.1122456??

1 INTRODUCTION

A large body of literature in the field of human-computer interaction focused on colors and light in user interfaces among different display technologies, and evaluated their benefits and drawbacks for different types of tasks. Dark themed user interfaces and so-called "Dark Modes" are gaining popularity in recent years in that they are characterized by a reversal of the prevalent color choices in most user interfaces, i.e., the light-on-dark color scheme that light (e.g., white) font colors are shown on a dark (e.g., black) background. Normal and inverted color choices were investigated over a wide range of display technologies and environments, and were linked to effects of legibility, aesthetics, energy savings, semantic effects, and emotions [6, 40, 41].

Effects of different color modes depend on the display technology used [6]. In this scope, augmented reality (AR) displays present an under-explored domain [60]. In particular, optical see-through (OST) displays and their corresponding light models change the way light and dark colors are perceived compared to traditional screens [26].

OST displays can be classified as *additive*, which means that they add light to the user's view, or *subtractive*, meaning that they can selectively dim areas in the user's view. The most popular current OST head-mounted displays (HMDs), such as the Microsoft HoloLens, use an additive design, which is characterized by "color blending" as they mix (a) light reaching the user's eyes from the physical environment with (b) light emitted by the display [23]. The practical effect is that these displays can only add light but not "take light away" from the user's view, which means that dark colors can not be induced but depend entirely on a dark background in the physical environment when looking through the display. Technological solutions to the issue of selectively dimming light from the user's view are difficult to realize and will not be available to consumers in the foreseeable future [33, 54, 59]. Hence, with AR annotations on OST-HMDs, such as black text on a white background, the white pixels match the brightness of the display and the black pixels match the color of the physical background behind the pixels [29].

It is our hypothesis that common AR annotations on OST-HMDs, involving light/dark foreground colors on a dark/light background, can have a significant impact on users' visual acuity and fatigue. In this journal paper, we extend a previously published conference paper presented at ACM SUI 2019 [32] by discussing and investigating these challenges on the example of a Microsoft HoloLens, while reporting two human-subject studies with an extended analysis. We asked participants in the two studies to read AR annotations using different color modes and to complete visual acuity tests and adjust font sizes of AR annotations in front of different backgrounds and with two different lighting conditions. We assessed participants' visual acuity, and preferences.

In particular, we investigated the following research questions:

RQ1 Are there subjective or objective benefits of dark mode color schemes with OST-HMDs with respect to visual acuity?
RQ2 Do users subjectively prefer normal or dark mode color schemes with OST-HMDs?
RQ3 Do users' preferences match the objective benefits or drawbacks of the modes?

This paper is an extension of our previous work in dark mode style color schemes in AR OST-HMDs, and extends upon this original work by describing a second user study, its results, and a discussion of its implications [32]. The paper is structured as follows. Section 2 presents an overview of related work. Section 3 describes the first human-subject study. The results of the first study are presented in Section 4 and discussed in Section 5. The second study is presented in Section 6, with results presented in Section 7 and discussion in Section 8. Section 9 concludes the paper and discusses future research.

2 RELATED WORK

Computer displays usually strive to present information with a high signal-to-noise ratio, in particular when presenting text to readers, which emphasizes the benefits of strong luminance differences instead of chromatic differences between the foreground and background. In the early age of electronic display technology when cathode-ray tube (CRT) monitors were prevalent, light-on-dark color scheme interfaces, i.e., light text on a dark background, were common because the text on the monitors was displayed by the electron beam hitting the phosphorous material for luminescence that is normally dark in the normal state. However, as the dark-on-light color scheme, i.e., dark text on a light background, was introduced in WYSIWYG editing systems to simulate ink on paper in the real world, it has been dominant in many computer user interfaces. Presenting dark text on a light background is usually referred to as *positive contrast*, which goes back to the signal processing theory, where the peak-to-peak contrast (or Michelson contrast [44]) measures the ratio between the spread and the sum of two luminances. This ratio is defined as $c = \frac{L_b - L_t}{L_b + L_t}$ with text luminance L_t and background luminance L_b, which is negative if $L_b < L_t$. While both positive and negative contrast conditions can provide the same theoretical peak-to-peak contrast ratio, a large body of literature has focused on identifying benefits of one of them over the other for different display technologies and use cases.

Multiple studies have found that *positive contrast* has benefits when the goal is to read text on computer screens [2, 5, 11, 47, 49, 58]. More recent studies investigated the causes of these benefits. Taptagaporn and Saito observed that participants developed a smaller *pupil diameter* when they used a positive contrast display compared to a negative contrast display [56]. This was also later confirmed to be the case by Piepenbrock et al. [48]. A small pupil diameter is known to increase the quality of the retinal image with greater depth of field and less spherical aberration, and it is largely affected by the amount of light reaching the observer's eyes. Buchner et al. investigated the display luminance in positive and negative contrast modes, showing that it is usually *higher* in positive contrast modes, e.g., when dark text is presented on a light background [6], which can be traced back to the ratio of screen space filled by (dark) letters or the (light) background. They further performed a study showing that, indeed, the amount of luminance had a dominant effect on performance while reading, but there was no difference between positive and negative contrast modes if the overall luminance was equivalent. In other words, by increasing the lightness of letters on a dark background they created the same effect that dark letters had on a moderately light background.

While the benefits of the positive contrast mode originate in the increased display luminance, this is not always desirable. For instance, the automotive industry has a long history of designing in-car displays and illumination for daytime and nighttime use. While positive contrast could be beneficial in terms of reading text on in-car displays independently of the environment lighting conditions, increasing the amount of light reaching the driver's eyes can have negative effects during the night, since it reduces the dark adaptation of the driver's eyes and thus their ability to perceive obstacles or people in low light road conditions [42]. Modern in-car displays thus usually switch to a "night mode" when it gets dark outside, which is characterized by a switch from a positive contrast (daytime) to a negative contrast (nighttime) mode, and a shift toward longer wavelength red colors that do not effect the dark adaptation of rods on the user's retina. Also, human circadian physiology and cognitive performance can be influenced by different displays [8]. Higuchi et al. found that performing a task with a bright display influences the nocturnal melatonin concentration and other physiological indicators of the human biological clock [27].

In modern life, people spend an increasing amount of time in front of computer screens, and experience various ocular symptoms, such as eyestrain, tired eyes, and sensitivity to bright lights and eye discomfort, which are referred to as computer vision syndrome (CVS) [3]. Various recommendations have been made with regard to luminance values for background and characters. Campbell and Durden emphasized that individual users should be able to adjust the brightness of the computer devices to adjust the luminance and contrast depending upon the time and the ambient lighting of the workplace [10]. Such features are now widespread, and many companies have adopted the dark mode interface design scheme in their hardware and software. For example, Apple included the feature of a dark mode setting that could be applied to adjust the coloration and brightness of all core applications on the device to a darker format with the release of their operating system Mojave[1].

There have been many studies about the effects of different displays on visual fatigue and acuity, e.g., 3D displays and virtual reality (VR) headsets [15, 35, 36, 38, 62]. Even in the domain of AR research, researchers investigated the effects of real background patterns and focal distance on visual fatigue and acuity [22, 43]. However, we are not aware of any work on positive or negative contrast modes in the field of AR, specifically with respect to text-based AR annotations. We see parallels between in-car heads-up displays and current-state OST-HMDs in the use of additive display designs (see Section 1), and the overall desire not only to ensure legibility of the displayed text but also to retain natural viewing of the physical environment behind the display without inducing severe visual fatigue.

3 EXPERIMENT I

In this section we describe the experiment that we conducted to investigate the three research questions stated in Section 1. Participants were asked to read text in AR on a HoloLens OST-HMD under different vision modes, and we asked them to complete visual acuity tests as well as rate their subjective experience and preference of these vision modes.

3.1 Participants

We recruited 19 participants for our experiment; ten male and nine female (ages 18 to 41, M=25.47, SD=5.93). The participants were members of the local university community. All of the participants had normal or corrected-to-normal vision; four participants wore glasses during the experiment, and four wore contact lenses. None of the participants reported known visual or vestibular disorders, such as color or night blindness, dyschromatopsia, or a displacement of balance. We ensured the normal condition of the participants' eyes by measuring the Ocular Surface Disease Index (OSDI) [50], which consists of 12 questions evaluating the frequency of dry eye disease symptoms over the preceding week. All 19 participants were categorized as normal with an OSDI score that is less than 12 in the range of 0–100. 18 participants reported that they had used a VR or AR HMD in the past, and four of them rated themselves as frequent users, having used HMDs on more than ten separate occasions. We asked participants to rate their current preference and usage of dark mode and inverted color schemes on their computers and mobile devices before the experiment. Two participants used these features whenever they were available, seven participants used these modes frequently, nine used these modes occasionally, and one never made use of these modes.

3.2 Materials

3.2.1 AR Stimuli and Vision Modes. For the presentation of the visual stimuli, we used a Microsoft HoloLens 1 so that participants could see the AR visual stimuli, which were displayed in front of them (Figure 1 and Figure 2). As a widely-used OST-HMD, the HoloLens provides an augmented field of view of circa 30 degrees horizontally by 17 degrees vertically in the center of the total human visual field. The resolution is 1268×720 pixels per eye. The HoloLens 1 leverages SLAM-based tracking [7] to localize itself with respect to the physical environment. For all study conditions, the HoloLens 1 display was set to maximum brightness.

For the rendering of the visual stimuli, we used the Unity game engine and its integration with the HoloLens 1 in order to present AR annotations in stereoscopic 3D. We chose AR textual annotations registered as planar objects ("holograms") in the laboratory space that consisted of either black text on a white background or white text on a black background. All virtual imagery used in the study was world fixed as opposed to head fixed, meaning that the virtual imagery would always be displayed in a fixed position relative to the study environment.

[1]Apple, "How to Use Dark Mode on your Mac" (https://support.apple.com/en-us/HT208976).

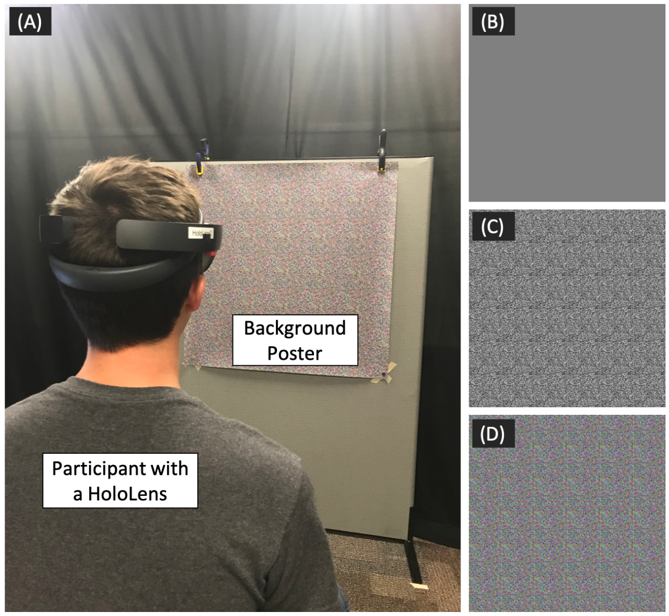


Fig. 1. Experimental setup and the images for the background posters: (A) A participant wearing the HoloLens, seated in front of the background poster. (B) The image used for the solid grey background poster. (C) The image used for the lightness distortion background poster. (D) The image used for the chromatic distortion background poster.

Participants in the study were positioned 1.52 meters away from the annotations, which were presented at the same depth as a physical background poster and were presented at the participant's eye height to avoid inclination conflicts. This distance was chosen because our virtual visual acuity chart was modeled after a physical chart where the size of its stimuli were calibrated for a distance of five feet (1.52m) from the user.

While the focal distance of the HoloLens 1 is two meters, we chose to keep the virtual acuity chart at 1.52 meters so that rescaling of our virtual assets could be avoided. Because of this chosen distance, it is possible that vergence-accommodation conflict may have limited the potential acuity of our participants. It was possible that the users' perception of virtual content could be influenced by non-uniformities in the display of the HMD [39], because of this we allowed users to change the orientation of their head as they needed, which should ensure that no user was stuck trying to distinguish imagery in a portion of the display prone to poor image quality.

We prepared four reading passages extracted from Pearson Test of English Read Aloud Practice Questions[2] in a 2×2 grid text board in a size of 72 by 72 centimeters (Figure 2). We further developed an AR version of a common visual acuity test chart similar to a Golovin–Sivtsev Table with Landolt C characters [20, 21, 57], which are characterized by circles with a missing piece on either of four sides (Figure 3). We implemented a randomized version of this test, where each trial resulted in different orientations of these circles. The chart had a physical (registered) size of 36 by 36 centimeters. The size of letters on the virtual chart were measured in Unity to range from 6 to 38 millimeters, with the opening on the Landolt C being 1/5 of the size of the letter. When converted into units of visual angle, this means participants had to resolve a feature that ranged in size from 0.0452 to 0.286 degrees. The minimum discernible feature size, measured in degrees of visual angle and corresponding to a visual acuity score of 20/20, is one arcminute, or 0.0167 degrees [28]. However, since the visual feature is

[2]Pearson Test of English (PTE) Read Aloud Practice Questions (https://pteacademicexam.com/pte-academic-speaking-read-aloud-practice-test-1-sample-exercises/).

being shown on an OST-HMD, there is a limitation imposed on the minimum size that a feature can be drawn based on the resolution of the device. The HoloLens 1 has a display resolution of 1268x720 pixels per eye, and a reported holographic density of 2500 light points per radian. Using these parameters along with the field of view of the device, we can calculate an approximate minimum visual angle that the device can achieve, which is 0.0238 degrees for a single pixel or 0.023 degrees for a single light point. Again, the chart either consisted of black Landolt C's on a white background or white Landolt C's on a black background. The chart was placed at the same distance as the AR annotations.

The two considered vision mode conditions were as follows:

- **Light Mode**: We used a *positive* contrast mode in which a Landolt C character in the foreground was presented as black and the background as white on the HoloLens.
- **Dark Mode**: We used a *negative* contrast mode in which a Landolt C character in the foreground was presented as white and the background as black.

The illuminance of the user interfaces was measured in a manner similar to Erickson et al. where the HMD was positioned in front of a dimmable light configured to illuminance values that were measured from the study environment [16]. Illuminance values of 240 lux and 10 lux were chosen based on the study environment conditions for the high light and low light conditions described below. Five sequential illuminance measurements were then made from the user's left eye position on the HMD directly facing the light source with the display on while rendering black, and then again in the same manner with the display on and rendering either the dark mode or light mode style UI. Contrast values were then calculated from these illuminance measurements using Michelson's contrast equations, which can be seen in table 1.

Table 1. This table shows the measured illuminance from the point of view of the user's left eye for both UI vision modes and both environment lighting conditions. Measurements for the 'Environment illuminance' column were made against a uniformly lit background without the HMD, measurements for the 'HMD Illuminance' column were made through the HMD facing the uniformly lit background with the display powered on and rendering black, and measurements for the 'UI Illuminance' column were made through the HMD rendering the UI. The illuminance values from the latter two were used in Michelson's Contrast equation to calculate a contrast value between 0 and 1 for the virtual imagery. Due to our light meter being incapable of making measurements at less than 1 lux, a value of 0.99 lux is used in Michelson's equation in the first two rows to provide a lower bound on the measured contrast.

Vision Mode	**Environment Illuminance**		**HMD Illuminance**		**UI Illuminance**		**Contrast**
	mean	std_dev	mean	std_dev	mean	std_dev	
Light Mode	9.92 lux	0.146 lux	<1 lux	~	4.5 lux	0.566 lux	>0.6695
Dark Mode	9.92 lux	0.146 lux	<1 lux	~	2.48 lux	0.376 lux	>0.3378
Light Mode	240.76 lux	4.631 lux	48.14 lux	1.572 lux	48.68 lux	2.636 lux	0.0103
Dark Mode	240.76 lux	4.631 lux	48.14 lux	1.572 lux	49.7 lux	1.213 lux	0.0204

3.2.2 Physical Environment and Background. We prepared an isolated room, which was surrounded by black curtains, in our laboratory space so that participants were not exposed to other visual stimuli during the study (Figure 1 A). We created different backgrounds for the experiment by mounting large-scale printed posters on a partition wall in front of the participants (Figure 1 B–D). The posters were made of 36" × 36" Premium Archival Matte papers. The three considered background conditions were as follows (Figures 1 and 3):

- **Uniform**: Participants perceived a uniform gray background (printed using a pixel intensity of 128 in the range of 0–255).
- **Lightness Distortions**: The background consisted of a mixture of randomly generated gray-scale pixels, impacting the apparent luminance of the stimuli being presented on the OST-HMD.

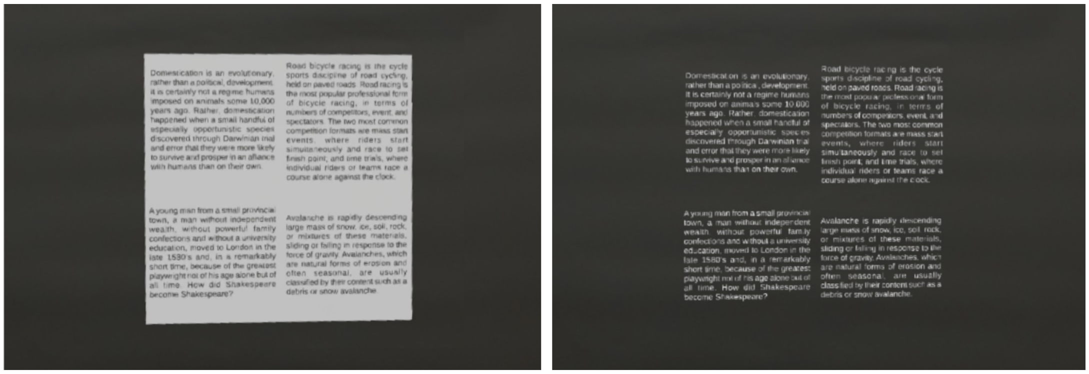


Fig. 2. Illustrations showing the AR text-reading task participants had to read during the experiment: the light mode (left) and the dark mode (right). While all virtual content was presented upright to study participants, these screenshots may appear slightly skewed or tilted due to the user's head position.

- **Chromatic Distortions**: The background consisted of a mixture of randomly generated RGB pixels, creating chromatic differences in line with an exaggerated simulation of using an OST-HMD in a cluttered environment.

Illuminance measures were made on each of the background posters and a repeated measures ANOVA was run. Post-hoc testing showed that there was no significant difference in illuminance between all possible pairs of background posters except for when comparing between the uniform background and the chromatic distortion background, where it was found that there was a significant difference of roughly eight lux between the two with the uniform background having the higher illuminance.

For the lightness distortion and chromatic distortion background, the individual pixels were sized to be 1/16 of an inch, which corresponds to 16.756 pixels per degree of visual angle from the user's position. The individual color of these pixels was determined by an online random pixel image generator, where pixel color was limited to greyscale for the lightness distortion poster, and pixel color was not otherwise limited for the chromatic aberration poster. This online tool generated a 100x100 pixel image, that was placed into a repeating 6x6 grid to form the final proof for poster printing.

3.2.3 Physical Lighting. To evaluate the differences between the amount of light in the physical environment, we controlled the overall lighting in the experimental setup. We created a well-lit environment that illuminated the room, and we compared it to a reduced-light environment. The two considered lighting conditions in this experiment were as follows (Figure 3):

- **High Light**: The environment was well-lit due to diffuse indirect ceiling lighting in the room (with 200–270 lux[3]).
- **Low Light**: The environment was reasonably dark due to dimmed lighting (with 10–12 lux).

3.3 Methods

We used a full-factorial within-subjects design in this experiment. As described in Section 3.2, the independent variables were as follows:

- **Vision Mode** (*Light Mode*, *Dark Mode*),
- **Physical Background** (*Uniform*, *Lightness Distortions*, *Chromatic Distortions*), and

[3]Measured by a URCERI Light Meter Digital Illuminance Meter

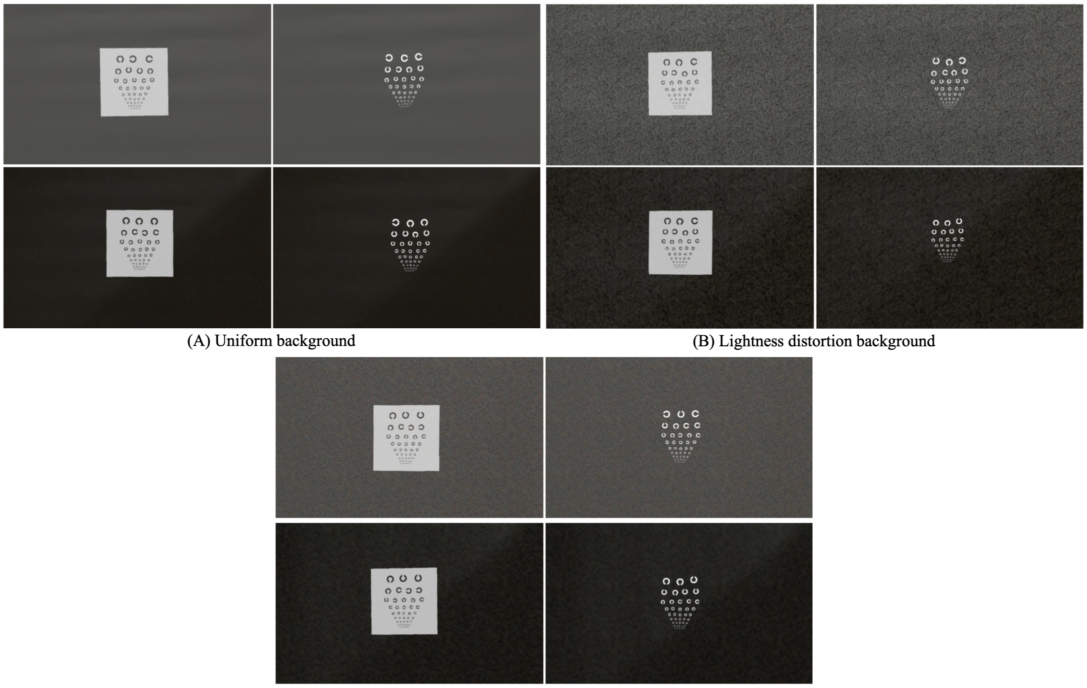


Fig. 3. Illustrations of the different experimental conditions. The left column shows the light mode AR stimuli and the right column is the dark mode stimuli. From top to bottom, the rows show the different backgrounds: (A) uniform, (B) lightness distortion, (C) chromatic distortion. Each in the high-light and low-light physical lighting conditions. These pictures were taken from the HoloLens via its built in device portal, which allows for screen captures of both physical and virtual content.

- **Physical Lighting** (*High Light*, *Low Light*).

Each participant completed all twelve possible configurations of the above-listed conditions. In order to avoid wear and tear on our archival matte background posters as well as to conserve time, the conditions were presented to the user in three groups of four conditions, grouped by background poster. The ordering of the groups was determined using a separate 3x3 Latin square design, while the physical lighting and vision mode conditions within each group followed a counter-balanced format using a 4x4 Latin square. Since we had 19 participants, our results are somewhat prone to ordering and sequencing effects.

3.3.1 Procedure. Prior to the experiment trials, participants first were asked to give their informed consent. Afterwards, they received task instructions and the experimenters made sure that they understood the task. Participants performed the interpupillary distance (IPD) calibration on the HoloLens before the experiment, so that the virtual content was rendered correctly in their view. Participants further completed a demographics questionnaire and then started the experimental trials.

At the beginning of each trial, participants were instructed to sit on the designated chair positioned directly in front of the wall which supported the posters (Figure 1). Participants were asked to verify that the positioning of the chart was correct before observing a set of four paragraphs that would be displayed for one minute (Figure 2). During this time, participants were asked to read the paragraphs silently (which were the same for each trial), and observe how easy or difficult the text was to read while sensing their general preference. After one minute had passed, the participant performed the (randomly generated) visual acuity test, where their accuracy and response time was recorded (Figure 3). Participants were encouraged to read as far down the acuity chart as they could go, and were not incentivized by time. Following each trial, participants were asked to complete a short usability questionnaire. After completing the questionnaire, the participant immediately moved to the condition with no other break taken in between.

After completing four trials associated with the same physical background poster, participants were asked to further compare the light mode and dark mode AR annotations and choose their preferred option for the displayed lighting and background combination. They were also asked to choose which option they found to be most comfortable, which option they found to be easiest to read, and which option they thought that they performed better on. After answering these questions for both lighting conditions, the background poster was changed for the next set of four trials. Testing resumed immediately after changing out the background posters, with no other breaks taken in between.

After completing all trials, participants had a brief interview with the experimenter on their overall perception or feeling about the conditions. Finally they received monetary compensation and finished the study.

3.3.2 Measures. We collected both objective and subjective measures to understand the benefits or drawbacks of the vision modes under the different background and lighting conditions in AR.

We considered the following dependent variables:

- **Visual Acuity**: As explained in Section 3.2.1, we used a visual acuity test based on a Golovin–Sivtsev Table with Landolt C characters [20, 21, 57]. The acuity is computed by the number of mistakes that a participant makes when reading from the chart. The choice of visual acuity as a measure differs from the measures of text legibility used in several similar studies [24, 25], where search tasks are employed and performance measures such as response time and error rate are used. The reason for this change was because of the limited field of view typically found on OST HMDs. Since screen space is very valuable in these type of devices, it may be beneficial to choose a UI configuration that can feasibly be displayed in a smaller size.
- **Usability**: We asked participants to rate the usability of the AR annotations after each condition using the short user experience questionnaire (UEQ-S) [52]. While the original UEQ is a semantic differential with 26 items, the UEQ-S consists of only eight items. The UEQ-S focuses on the measurement of the two meta-dimensions, *pragmatic quality*, which measures the perceived utility and practical qualities of the interface and *hedonic quality*, which measures the enjoyment or boredom experienced by the user when interacting with the interface. The overall usability score is based on those two quality aspects.
- **Preferences**: We asked participants several questions to indicate their subjective preferences and rank the two vision modes for each of the background and lighting conditions. We asked users which of the two UIs they preferred, which was more comfortable, which was easier to read, and which UI they thought they performed better with.

We further debriefed the participants and asked them to verbalize additional observations and impressions.

3.4 Hypotheses

Based on the related work, and our study design, we formulated the following hypotheses for the objective and subjective results:

H1 Participants will show higher visual acuity with the light mode AR annotations than using the dark mode.
H2 Participants will indicate higher subjective ratings of usability and preference for the dark mode AR annotations in dark physical environments.

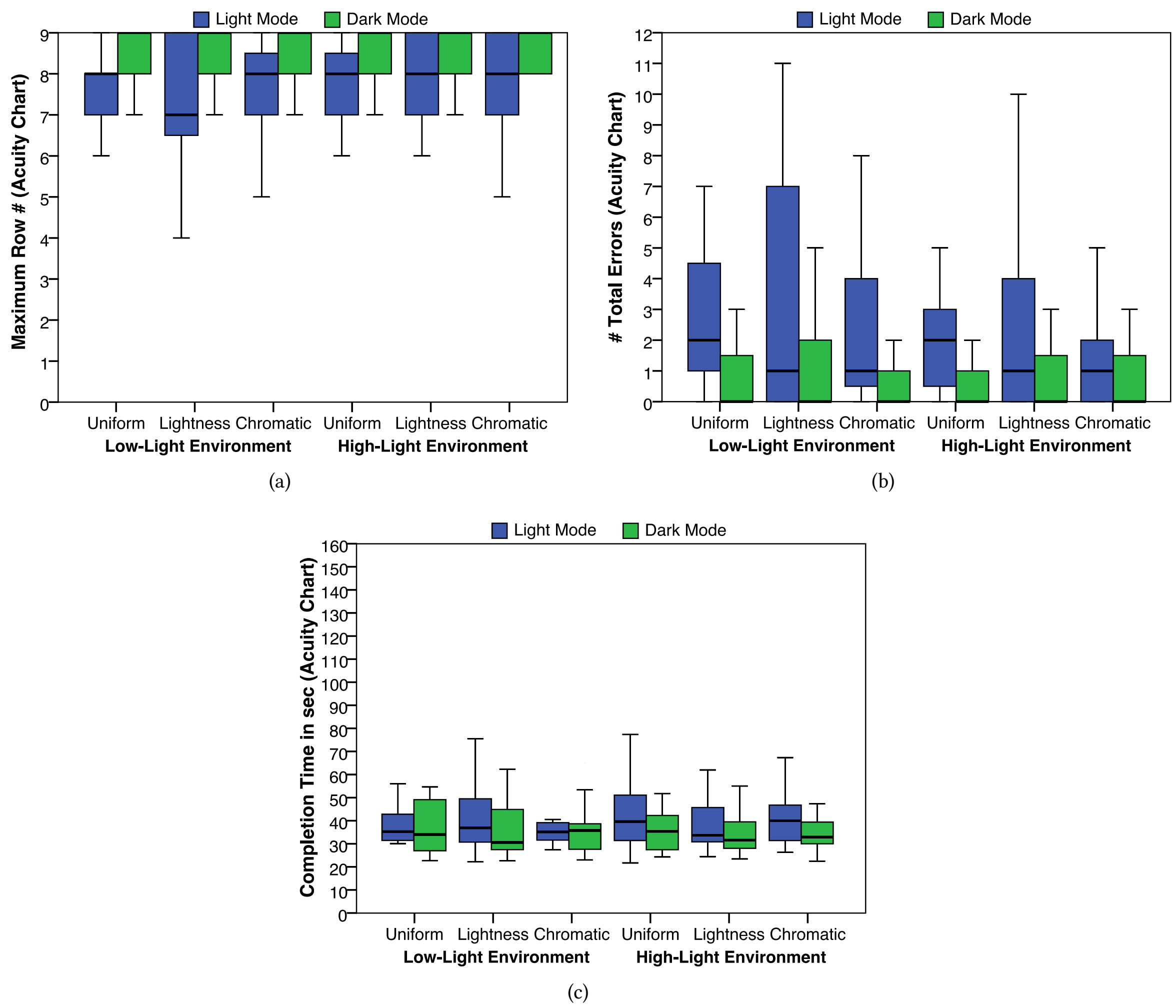


Fig. 4. Results for the *visual acuity* tests: (a) maximum row on the acuity chart that could be completed without errors (between 0 and 9; higher is better), (b) total number of errors on acuity chart (lower is better), and (c) completion time for the acuity chart (lower is better).

4 RESULTS I

We used parametric statistical tests to analyze the responses in line with the ongoing discussion in the field of psychology indicating that parametric statistics can be a valid and more informative method for the analysis of combined experimental questionnaire scales with individual ordinal data points measured by questionnaires or coded behaviors [34, 37]. We analyzed the responses with repeated-measures ANOVAs and Tukey multiple comparisons with Bonferroni correction at the 5% significance level. We confirmed the normality with Shapiro-Wilk tests at the 5% level and QQ plots. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated.

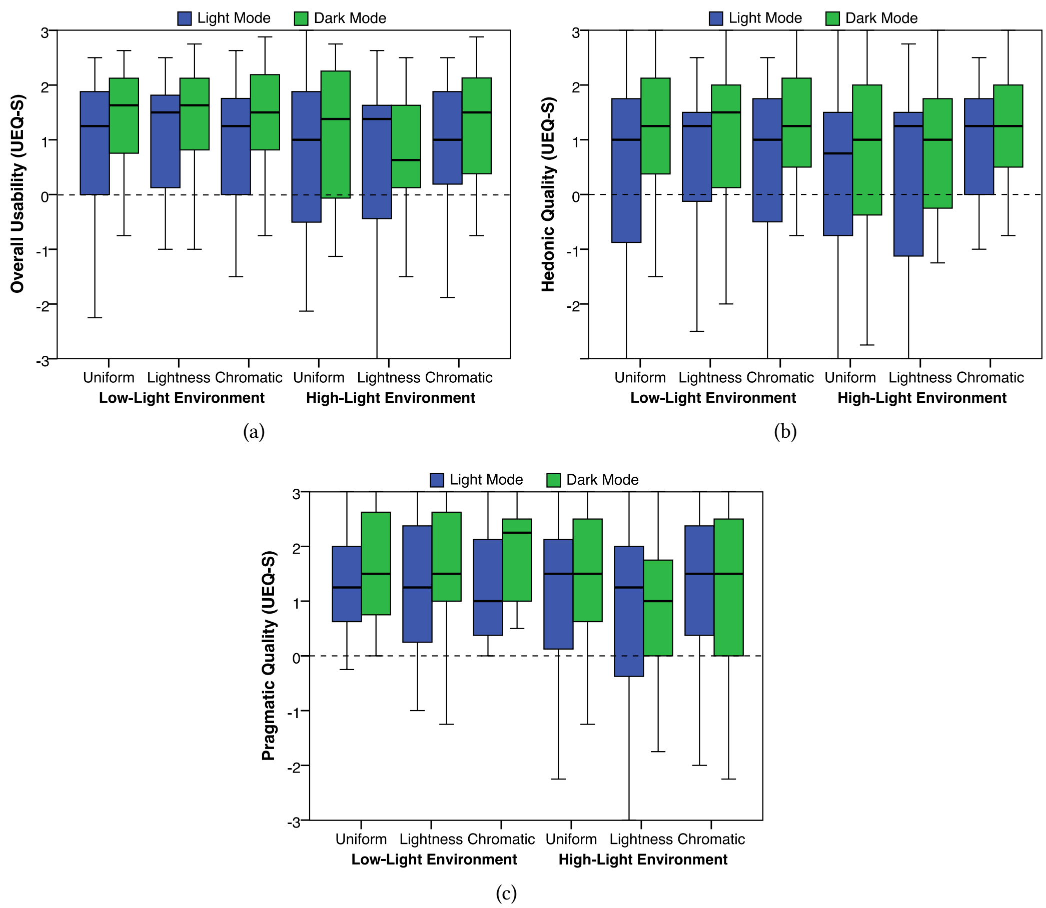


Fig. 5. Results for the *usability* estimates using the UEQ-S questionnaire (between -3 and 3; higher is better): (a) overall usability score, (b) hedonic quality, and (c) pragmatic quality.

4.1 Visual Acuity

The results for the visual acuity test are shown in Figure 4.

We found a significant main effect for **vision mode** between the light mode ($M = 7.62$, $SD = 1.31$) and the dark mode ($M = 8.32$, $SD = 1.27$) on the maximum row on the visual acuity chart that could be completed without errors, $F(1, 18) = 9.20$, $\boldsymbol{p = 0.007}$, $\eta_p^2 = 0.338$, indicating that participants had a significantly higher visual acuity for the dark mode than the light mode.

We found a significant main effect for **vision mode** between the light mode ($M = 2.52$, $SD = 2.95$) and the dark mode ($M = 1.10$, $SD = 1.94$) on the numbers of errors made in the visual acuity tests, $F(1, 18) = 12.65$, $\boldsymbol{p = 0.002}$, $\eta_p^2 = 0.413$, indicating that participants completed the tests with significantly fewer errors for the dark mode than the light mode.

We found a significant main effect for **physical lighting** between the low-light ($M = 2.04$, $SD = 2.84$) and high-light ($M = 1.57$, $SD = 2.30$) environment on the numbers of errors made in the visual acuity tests, $F(1, 18) = 11.68$,

p = 0.003, $\eta_p^2 = 0.394$, indicating that participants completed the tests with significantly fewer errors in the high-light environment than in the low-light environment.

We also found a significant main effect for **vision mode** between the light mode (M = 40.89 sec, SD = 16.64 sec) and the dark mode (M = 37.18 sec, SD = 14.20 sec) on the completion time of the visual acuity tests, $F(1, 17) = 16.04$, ***p = 0.001***, $\eta_p^2 = 0.485$, indicating that participants completed the tests significantly faster for the dark mode than the light mode.

4.2 Usability

The results for usability (UEQ-S) are shown in Figure 5. This data was initially analyzed using the UEQ-S data analysis tool found on the ueq-online website, after which it was transferred into SPSS to calculate P values and generate figures.

We found a significant main effect for **vision mode** on *overall usability*, $F(1, 17) = 9.17$, ***p = 0.008***, $\eta_p^2 = 0.350$, specifically on *hedonic quality*, $F(1, 17) = 10.22$, ***p = 0.005***, $\eta_p^2 = 0.375$, however its effect of *pragmatic quality* was not significant, $F(1, 17) = 4.17$, $p = 0.057$, $\eta_p^2 = 0.197$. The results indicate that participants rated usability as significantly higher for the dark mode than the light mode.

We found a significant main effect for **physical lighting** on *overall usability*, $F(1, 17) = 7.00$, ***p = 0.017***, $\eta_p^2 = 0.292$, specifically on *pragmatic quality*, $F(1, 17) = 6.25$, ***p = 0.023***, $\eta_p^2 = 0.269$, and a but not for *hedonic quality*, $F(1, 17) = 3.72$, $p = 0.071$, $\eta_p^2 = 0.180$. The results indicate that participants rated usability as significantly higher for the low-light physical environment than the high-light environment.

We found a significant interaction effect between **physical lighting** and **vision mode** on *overall usability*, $F(1, 17) = 6.85$, ***p = 0.018***, $\eta_p^2 = 0.287$, specifically on *pragmatic quality*, $F(1, 17) = 4.89$, ***p = 0.041***, $\eta_p^2 = 0.223$, and on *hedonic quality*, $F(1, 17) = 6.97$, ***p = 0.017***, $\eta_p^2 = 0.291$. Multiple comparisons (all $p < 0.05$) showed that *overall usability*, *pragmatic quality*, and *hedonic quality* were all significantly higher for the dark mode than the light mode in the low-light environment. We further found that the three measures were significantly higher for the low-light environment than the high-light environment for the dark mode. Last but not least, we found that *overall usability* and *hedonic quality* were significantly higher for the dark mode in the low-light environment than the light mode in the high-light environment, while pragmatic quality was non-significant only achieving $p = 0.082$.

4.3 Subjective Preferences

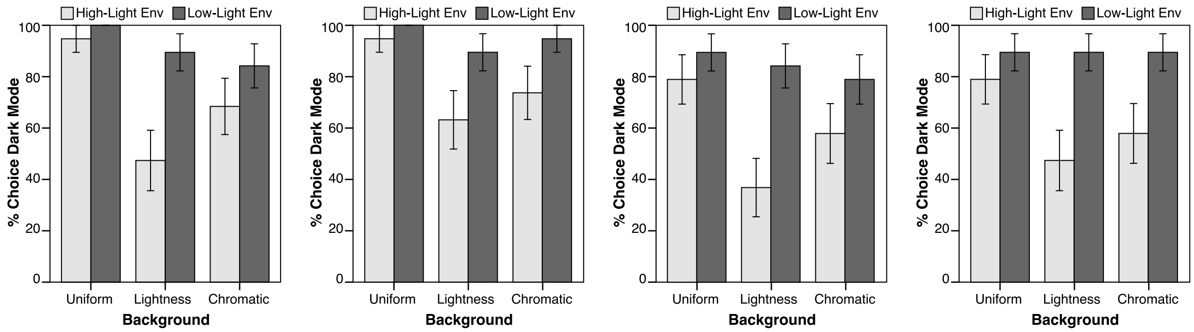
The subjective preferences of our participants are shown in Figure 6.

We found significant main effects for **physical lighting** on *overall preference*, $F(1, 18) = 6.09$, ***p = 0.024***, $\eta_p^2 = 0.253$, on *visual comfort*, $F(1, 18) = 7.44$, ***p = 0.014***, $\eta_p^2 = 0.292$, on *easy to read*, $F(1, 18) = 9.21$, ***p = 0.007***, $\eta_p^2 = 0.338$, and on *perceived performance*, $F(1, 18) = 19.36$, ***p < 0.001***, $\eta_p^2 = 0.518$. The results show that the dark mode is mainly the preferred choice in the low-light environment and less so in the high-light environment.

We also found a significant main effect for **background** on *overall preference*, $F(1.51, 27.17) = 5.78$, ***p = 0.013***, $\eta_p^2 = 0.243$. Post-hoc tests (all $p < 0.05$) showed a significantly higher preference for the dark mode for the uniform background than the lightness/chromatic backgrounds for *overall preference*. They further showed a significantly higher preference for the dark mode for the uniform background than the lightness background for *visual comfort* and for *easy to read*.

5 DISCUSSION I

In this section, we summarize the main findings of experiment one and discuss implications for the use of AR user interfaces with current-state OST-HMDs.



(a) Overall Preference (b) Visual Comfort (c) Easy to Read (d) Perceived Performance

Fig. 6. Subjective preferences: Percentages indicating how often participants chose the dark mode over the light mode for the different *physical lighting* and *background* conditions. The plots indicate (a) which of them were overall preferred, (b) perceived as more visually comfortable, (c) perceived as easier to read, and (d) perceived as resulting in higher performance (fast and accurate reading). Error bars represent ± one standard error.

5.1 Dark Mode Improves Visual Acuity

In contrast to our Hypothesis **H1**, we found that participants had a significantly higher visual acuity for the dark mode than the light mode. They were able to complete significantly more rows on the visual acuity test chart without errors for the dark mode. Moreover, they also made significantly fewer overall errors on that test for the dark mode conditions, and it also took them significantly less time to complete the test.

This result is interesting as it implies that visual details such as text are easier to see on OST-HMDs if they are presented in light colors over the background (i.e., dark mode) instead of the traditional approach on computer screens, where the details are dark and the space around them is illuminated (i.e., light mode). Since the light mode approach illuminates a screen area on an OST-HMD with gaps of non-illuminated pixels (dark transparent pixels) for the details on the OST-HMD, we assumed that these might be prone to an influence of complex physical backgrounds shining through those gaps.

This difference between the results from the two vision modes could be related to optical scattering resulting from the refractive and diffractive properties of the materials of the HMD, where light is unintentionally deviated from its intended position on the display due to irregularities in the lens or waveguide through which the light travels or the surface which is being projected upon [53]. As our stimuli consisted solely of black and white colored user interfaces, and the OST-HMD cannot render black due to the additive nature of the display, it is likely that the white stimuli were prone to optical scatter, while the dark stimuli were not. In the dark mode conditions, the white coloring is found only on the text, which means that if this portion of the visual stimuli is affected by scatter, then the perimeter of the text may appear to bleed into the background, as the white from the projection partially illuminates areas around the perimeter of the text. Conversely, in the light mode conditions, the white coloring is found in the background behind the text and no light is being projected in areas where the letters are located, so the inner perimeter of the letters may be partially illuminated by the white background, which causes the perimeter of the letters to appear to bleed in with the background.

We also found that participants made significantly fewer errors on the visual acuity test in the high-light physical environment compared to the low-light environment. However, this is in line with the literature and not surprising as it can be explained by the well-known relationship between environment luminance and pupil diameter (see Section 2).

Erickson et al. performed a comparison of results between this study and among similar studies, and showed that there are variations between the results of studies where UI configuration colors were investigated in terms of human performance [17]. They show that there are many different OST displays that have been tested in previous literature, both monoscopic and stereoscopic displays, and displays of varying focal distance and luminance capability, which may account for some of the differences between studies. Negative contrast UI configurations tended to perform well in several of the previous works [13, 61] However, there were some works that directly contradict those that were obtained here [19]. Because of this, more research is needed in this area to understand the best practice guidelines for UI design in OST-HMDs, and how they are impacted by factors such as device luminance and focal distance.

5.2 Dark Mode Improves Usability and Preference in Dark Environments

In line with our Hypothesis **H2**, we found that participants indicated a clear overall preference for the dark mode over the light mode. Participants' responses suggest that this goes back to subjective impressions of higher visual comfort with the dark mode, an overall sense of making it easier to read, and the impression that the dark mode increased their performance in AR during the experiment. They further indicated that the dark mode significantly increased the overall usability of AR annotations as well as their hedonic and pragmatic qualities.

We found that participants preferred the dark mode mainly in the low-light environment and less so in the high-light environment. In particular, as shown in Figure 6, participants indicated a balanced preference for either light or dark modes in the high-light environment with a more complex background (with chromatic or lightness distortions), which, arguably, might be more ecologically valid than a uniform background. We would also like to point out that these preferences might further shift towards the light mode in situations with even more environmental light. Both physical lighting setups in our study were designed for typical room interior lighting found in office environments.

5.3 Limitations

As discussed in Section 1, our results are specific to additive light model AR OST-HMDs with similar optical scattering properties to the HoloLens. Due to the peculiarities of OST displays, we are making no claims that these results would transfer to video see-through or immersive VR displays, or that these findings would still hold true if technological solutions to the color blending of OST displays become feasible [33, 54, 59]. In particular, a recent study performed using the Oculus Rift S virtual reality HMD has shown that the effects found on visual acuity were different in that users showed higher acuity with light mode interfaces when the virtual environment lighting was bright, and higher acuity with dark mode interfaces when virtual environment lighting was dim [15].

It is likely that strong subjective preferences or a user's sense of gained benefits of the light or dark mode might transcend different display technologies. As listed in Section 3.1, our participants were roughly split into one half who used dark mode graphics occasionally in their daily life and the other half who used them frequently. We observed no effects of these general tendencies on the results in this study.

Our methods employed a non-conventional counterbalanced design to reduce study fatigue and avoid wear and tear on our background posters. As mentioned above, the conditions were placed into three groups based on the background poster (that were determined through use of a 3x3 Latin Square.) Within these groups, a separate 4x4 Latin square was used to counterbalance the remaining conditions (lighting and vision mode). Because of this non-conventional approach, it is possible that our results are prone to some ordering and sequencing effects.

While this initial study successfully showed benefits of dark mode style user interfaces, all virtual stimuli were presented at the same depth. As AR HMDs have a designated focal depth at which there is no vergence-accommodation conflict, it is possible that when the depth of text being displayed is changed to be either in front or behind this optimal distance that the benefits and drawbacks of certain color schemes may vary. Further, it is also possible that as the distance at which text is displayed increases away from a user, that the minimum visual angle required to resolve small features of the text may not stay as consistent as it otherwise would with a physical visual stimulus. For these reasons, we designed a second experiment which addresses some of these concerns, which is described in the following section.

6 EXPERIMENT II

In this section we describe the design and implementation of a second experiment, in which we investigate the relationship between minimum discernible feature size, distance from user, vision mode, physical lighting, and background appearance.

6.1 Participants

We recruited a total of 12 additional participants from the university community; eight male and four female (ages 19 to 29, M=23.6, SD=3.0). All of the participants had normal or corrected-to-normal vision; six participants wore glasses during the experiment, and one wore contact lenses. None of the participants reported known visual or vestibular disorders, such as color or night blindness, dyschromatopsia, or a displacement of balance. Nine participants reported that they had used a VR or AR HMD in the past, and two of these participants rated their self as a frequent user, having used HMDs on more than ten separate occasions.

6.2 Materials

The materials used for experiment two were largely the same as those used during experiment one, with all study stimuli being displayed on the Microsoft HoloLens 1 in stereoscopic 3D. The same set of three background posters (uniform, lightness distortions, and chromatic distortions) were used within the same testing environment as the previous experiment, where lighting levels with lights on and with lights dimmed were measured with an Urceri MT-912 light meter and were confirmed to fall within the same ranges tested during experiment one (see Section 3).

The study implementation was constructed in the Unity engine and was run using holographic remoting on a laptop computer, where imagery was streamed to the HoloLens.

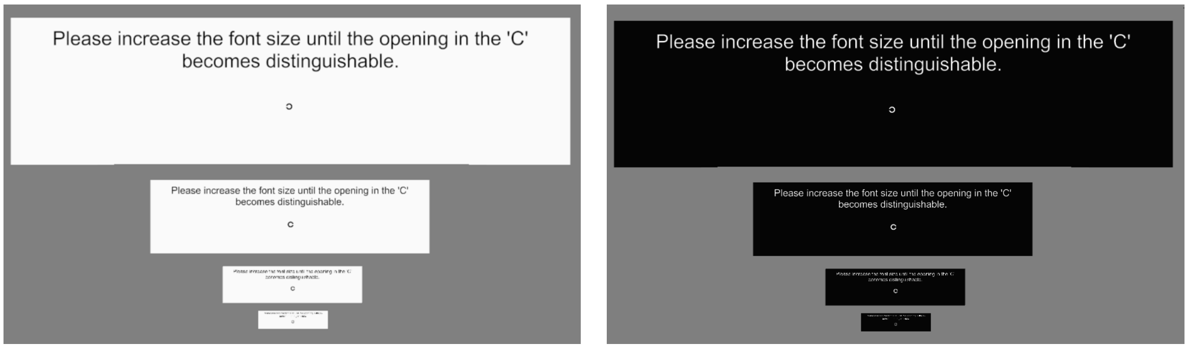
6.3 Methods

We used a full-factorial within-subjects design in this experiment. The independent variables were as follows:

- **Vision Mode** (*Light Mode*, *Dark Mode*),
- **Physical Background** (*Uniform*, *Lightness Distortions*, *Chromatic Distortions*)
- **Physical Lighting** (*High Light*, *Low Light*).
- **Depth of Stimuli** (*1, 2, 4, or 8 meters from the user*).

These depth values were chosen since the HoloLens has a focal depth of 2 meters. Because of this, stimuli depths at levels other than 2 meters may introduce effects of accommodation-vergence conflict, where the users' eyes converge at a distance that is different than the distance at which the graphics are rendered. The procedure and dependent variables in this study are described in the following sections.

In the same manner as experiment 1, the conditions in this experiment are grouped based on the three background posters, which are presented using a separate 3x3 Latin square from the rest of the conditions, which were counter balanced using a 4x4 Latin square design based on the physical background and physical lighting variables. Depth of the virtual stimuli was not included in the counterbalancing of conditions, and was presented in order from nearest depth value to furthest. Because of this approach, it is possible that the results of the study are subject to ordering and sequencing effects.



(a) (b)

Fig. 7. Illustration of the sizing of the user interface panels for the different depth levels for the (a) light mode and (b) dark mode, shown over the uniform grey physical background.

6.3.1 Procedure. Upon arrival, participants were asked to read through an IRB-approved protocol and provide verbal consent to participate in the study. Following this, participants were instructed on how to properly don and doff the HoloLens, and were asked to complete the integrated HoloLens IPD calibration. The participants were then shown a floating user interface panel at a distance of one meter directly forward of the user. This user interface can be seen in Figure 7, which was displayed in front of one of the three background posters described in Section 3.2. The background posters were displayed to participants in a counter-balanced order using a Latin Square, and would be traded out as the user progressed through the study conditions.

Following this initial setup, the participant was asked to use an Xbox controller to increase the size of a Landolt C in size increments of one font point, (which was calculated to correspond to a one millimeter change in the height of the letter). The letter first appeared as one-point font, and users were tasked to increase the size until the opening in the C was just barely discernible. The participant then pressed a button to confirm their selection, and again pressed the button as a final confirmation before moving on to the next condition. As participants progressed, the depth of the user interface panel changed from one, to two, to four, to eight meters away, after which the color schemes would be changed to be displayed in the opposite vision mode. The participants then complete the size adjustment task for the four depths again before finishing the condition.

After completing a condition of four depths for each of the two vision modes, the lighting of the room was configured to the opposite of what the participant had just experienced. After both lighting configurations were experienced for each of the vision modes and depths, the background poster was traded out to be the next poster in the counter-balanced order. This same process was repeated for each of the remaining background posters until the participant had experienced all possible combinations of conditions. No breaks were taken between conditions besides small amounts of time (10-15 seconds) to switch out background posters or adjust the lighting.

Once all conditions were experienced, the participant was instructed to take off the HoloLens and answer a post-questionnaire which consisted of subjective questions and gathered demographics information. Participants then received monetary compensation for their time.

6.3.2 Measures. We collected both objective and subjective measures to complement the results obtained in experiment one. We considered one dependent variable, which was a user-selected minimum readable font size for each set of conditions. This font size is converted into feature size in units of degrees of visual angle for analysis in our results section. We additionally gathered subjective feedback from the users with a post-questionnaire and asked users to choose their preferred color scheme that they experienced during the course of the study.

6.4 Hypotheses

Do to the background literature and results of experiment one, we had the following hypotheses regarding the results of experiment two:

- **H4** Participants will select smaller minimum readable feature sizes for AR annotations displayed in dark mode than they will for annotations displayed in light mode.
- **H5** Participants will indicate a preference for the dark mode over the light mode.
- **H6** Participants will select the smallest feature sizes at the depth at which there is no VAC (2 meters).

Whereas H4 and H5 directly stem from the results of experiment one, H6 is motivated by the previous work in the AR domain involving VAC. We could not find any existing studies comparing visual acuity, or other similar acuity measures, at varying depths in AR. However, there has been much work that has investigated the interesting effect of VAC on distance estimations, where it is thought that the conflicting visual depth cues cause users to overestimate distances that are closer than the focal depth of the display, and underestimate distances that are further than the focal depth of the display [46, 55]. VAC has also been shown to have effects on the visual fatigue and comfort level of the user [1].

7 RESULTS II

Similarly to experiment one, we used parametric statistical tests to analyze the responses in line with the ongoing discussion in the field of psychology indicating that parametric statistics can be a valid and more informative method for the analysis of combined experimental questionnaire scales with individual ordinal data points measured by questionnaires or coded behaviors [34, 37]. We analyzed the responses with repeated-measures ANOVAs and Tukey multiple comparisons with Bonferroni correction at the 5% significance level. We confirmed the normality with Shapiro-Wilk tests at the 5% level and QQ plots. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated.

7.1 Minimum Discernible Feature Size

For all figures in this section, the depth variable shown on the x-axis is converted into focal power with units of diopters (1/meters) for better comparison with the existing literature in vision sciences and vergence-accommodation conflict. Additionally, the dependent variable shown on the y-axis is the feature size measured in visual angle rather than the character size. The feature size was calculated by dividing the character size by five, as the opening on the Landolt C is one fifth the size of the character.

As mentioned in section 3, the minimum discernible feature size, measured in degrees of visual angle and corresponding to a visual acuity score of 20/20, is one arcminute, or 0.0167 degrees [28]. However, since the visual feature is being shown on an OST-HMD, there is a limitation imposed on the minimum size that a feature can be drawn based on the resolution of the device. The HoloLens 1 has a display resolution of 1268x720 pixels per eye, and a reported holographic density of 2500 light points per radian. Using these parameters along with the field of view of the device, we can calculate an approximate minimum visual angle that the device can achieve, which is 0.0238 degrees for a single pixel or 0.023 degrees for a single light point.

The figure depicting the mean feature sizes across all tested conditions and broken down into separate plots by participant number is shown in Figure 8 a. As shown in the figure, participant one had a mean minimum discernible feature size at one meter of approximately 0.24 degrees, which is an order of magnitude worse than expected, although their value at eight meters (0.125 D) falls within a reasonable range of the other participants. This participant alone was shown to have a significant effect on our analysis, and was thus removed from the data set for the ANOVA analysis and all acuity-based figures.

The results comparing the minimum discernible feature sizes separated by the independent variables of vision mode, physical background, and lighting are respectively displayed in subfigures b, c, and d of Figure 8.

For main effects, we did not find significant results for vision mode ($F(1, 10) = 1.830$, $p = 0.206$, $\eta_p^2 = 0.155$), for physical background ($F(2, 20) = 0.475$, $p = 0.629$, $\eta_p^2 = 0.045$), or for lighting ($F(1, 10) = 0.172$, $p = 0.687$, $\eta_p^2 = 0.017$). However, we did find a significant main effect of depth on the minimum discernible feature size, ($F(1.094, 10.943) = 9.540$, $\boldsymbol{p = 0.009}$, $\eta_p^2 = 0.488$), indicating smaller minimum discernible feature sizes at higher depth values. Pairwise comparisons between participant means at each depth value revealed a significant difference in minimum discernible feature size between the depths of one meter and 2 meters (**p=0.042**), indicating significantly larger feature sizes at one meter than at 2. No other pairwise comparisons between depth values were significant.

While no significant effects were found on feature size, physical background, and physical lighting, the partial eta squared values indicate a large effect, small effect, and small effect respectively as interpreted by Cohen [12]. This indicates that our sample size of 11 participants, since we excluded participant 1 in the analysis, was not a sufficient sample size to produce significant results. A power analysis run in G-Power indicated that significant results for vision mode would likely be achieved with a minimum of 16 participants, while 54 or 174 participants would be required for significant results for physical background or physical lighting respectively.

We found two significant interaction effects. One was between the physical lighting and the physical background on the minimum discernible feature size, $F(2, 20) = 6.599$, $\boldsymbol{p = 0.006}$, $\eta_p^2 = 0.398$. The other was between the vision mode, physical lighting and the depth on the minimum discernible feature size, $F(3, 30) = 3.421$, $\boldsymbol{p = 0.030}$, $\eta_p^2 = 0.255$.

7.2 Subjective Preferences

Participants were asked in a post-questionnaire to choose their overall preference between the two vision modes. Ten participants indicated a preference for dark mode and two participants indicated a preference for light mode. This was analyzed using a binomial test with a test value of 0.5, where it was found that there was a statistically significant preference in favor of the dark mode, $\boldsymbol{p = 0.039}$.

8 DISCUSSION II

In this section, we summarize the main findings of experiment two and discuss implications for the use of AR user interfaces with current-state OST-HMDs.

8.1 Dark Mode Yields Smaller Feature Sizes

Based on the results of experiment one, we predicted in hypothesis **H4** that participants would be capable of resolving smaller features with the dark mode style UI than with the light mode style UI. While we did not see a significant main effect of vision mode on feature size, we did achieve a large effect via Cohen's interpretation of the partial eta squared value. Additionally, the graph of our results in Figure 8 b indicates that on average, smaller feature sizes were resolved for all tested depths in the dark mode style vision mode compared to the light mode style. Based on these findings we partially accept hypothesis **H4**. As mentioned in the previous discussion section, it is possible that this benefit of dark mode style UIs is related to optical scatter.

8.2 Dark Mode is Preferred Over Light Mode

In this second experiment, we again see a significant preference of our participants for the dark mode over the light mode, therefore we accept hypothesis **H5**. These results align well with what was obtained from the first study, where factors of decreased minimum readable feature size combined with subjective aesthetics likely led our participants to choose their preference.

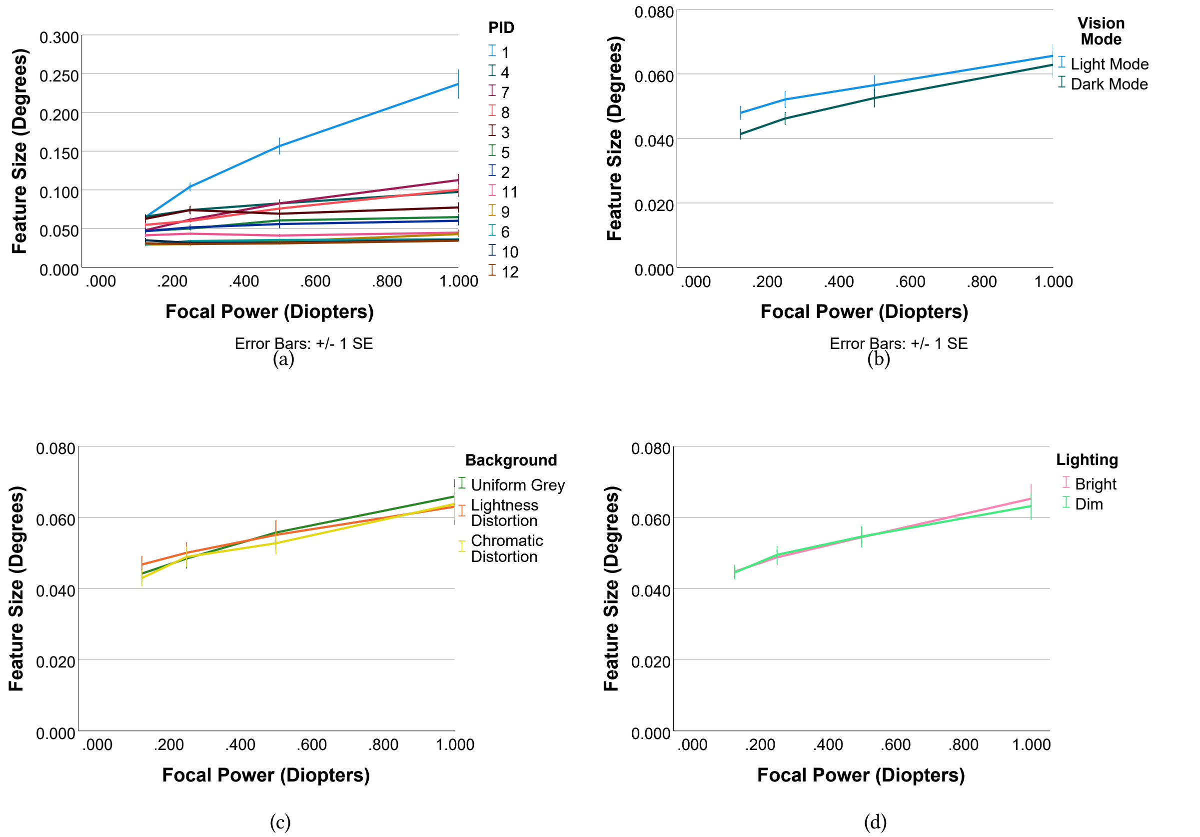


Fig. 8. Figure depicting the smallest discernible feature sizes (measured in degrees of visual angle) for the given focal powers based on the different tested depths between (a) participants, (b) vision modes, (c) physical backgrounds, and (d) physical lighting. Error bars represent ± one standard error.

8.3 Smaller Feature Sizes with Increased Distance

Perhaps the most surprising result of experiment two was the significant main effect of depth on the minimum discernible feature size. In general, our users were able to distinguish smaller sized features as the distance between user and UI increased, which can clearly be seen in subfigures b-d of Figure 8. However, when we examine subfigure a, we see that there is quite a bit of deviation in the shape of the plots between the different participants, with some having a near-flat horizontal shape, while others show the shape where size decreases with increased distance. We expected to see this flatter shape on these plots, since the y-axis is essentially a measure of visual acuity. This means that, barring effects that arise from accommodation, convergence, and peculiarities of the display, the user should theoretically have a fixed minimum visual angle that they are capable of resolving. While this seems to occur for some of our participants (see participants 6, 10, and 12 in Figure 8 a), there are many others that deviate from this expected pattern. Due to these results we cannot accept hypothesis **H6**, which raises the question of what may be causing such an effect.

One strong possibility is that the vergence cues given to the user by the OST-HMD cause vergence-accommodation, where accommodation happens as an involuntary reflex of vergence. This was previously examined by Inoue and Ohzu, where they measured accommodative responses to 3D imagery shown at distances closer to the user than the focal distance of the display [30]. Their work showed that users under-accommodate to imagery shown in this manner, which would account for some of the increase in feature sizes that we see from the 0.5 (2 meter) diopter distance to the 1 diopter (1 meter) distance. Their work additionally showed that pupil size decreases with decreased distance at a quicker rate for 3D imagery than it does for a real object. As pupil size is linked to visual acuity, it is possible that this had additional effects on the feature sizes for all tested depths [9]. The work by Inoue and Ohzu however, does not show experimental results of what occurs when the 3D image is shown at depths greater than the focal depth of the display, although their figures suggest that users would over accommodate and pupil size would further increase with increased depth [30].

This relationship between feature size and depth also bears a striking resemblance to a similar phenomena found in distance perception research, where in general users overestimate distances when the task object is closer than the focal depth of the display, and underestimate distances when it is further than the focal depth of the display [4, 55]. It is thought that this effect comes as a result of vergence-accommodation conflict. Schmidt et al. also investigated this phenomena and found that there was a relationship between the accuracy of depth estimations with the level of experience of the user with stereoscopic 3D displays [51]. They showed that users tended to have more accurate depth estimations if they had more experience with these type of displays. It is possible that a similar effect may account for some of the deviation of results between participants in our study, with less experienced users having a steeper plot in Figure 8 a, and more experienced users having a flatter plot. This would also suggest that more experienced users are capable of overcoming the vergence-accommodation reflex in order to better resolve visual features on the OST-HMD.

8.4 Limitations

There were several limitations in the design of this second experiment. We had originally intended to use font size in points as our dependent variable as opposed to the feature size in degrees of visual angle, and because of this, our implementation had users scale the size of the font by adding or subtracting one point of font size per click of input. It was calculated from our Unity implementation that changing the font size by one point was the equivalent of changing the height of the Landolt C by one millimeter. However since the user interface is presented at varying depths, and visual angle is dependent on depth, one millimeter of change has a more significant effect on the perceived size of the letter at closer distances than it does at further distances. This allows users to be more precise with their scaling task for further distances than closer distances, and it is possible that feature sizes are slightly overestimated at near distances because of this.

There is also a limitation that stems from the age demographics of our participants, where due to their young age, our participants are not representative of the general population. A user's ability to accommodate is linked to their age, where people generally become presbyopic and start to lose their ability to accommodate in their early 40's. We intentionally recruited participants younger than this to avoid making comparisons between groups of users based on their age, as it is likely that this same procedure performed on presbyopic participants would yield significantly different results. We would hypothesize that feature sizes would be more consistent regardless of depth for presbyopic participants, but this should be confirmed in future work.

9 CONCLUSION AND FUTURE WORK

In this paper, we presented an analysis of positive and negative contrast modes, also known as light and dark modes, with OST-HMDs based on additive light models. We presented two human-subject studies that we conducted to understand objective benefits and participants' subjective preferences of these vision modes under different physical lighting and background conditions. We showed in our first experiment that dark mode color schemes used for AR annotations significantly increased visual acuity, and were overall preferred by participants, especially in low-light physical environments and with complex backgrounds. We further demonstrated in our second experiment that minimum readable feature sizes at varying distance are smaller in dark mode than in light mode, and that users tend to choose significantly smaller feature sizes as the distance to the UI increases.

In future work, we believe that it is important to investigate related differences with OST-HMDs, and to investigate vision modes also with video see-through and immersive VR displays as well as projection-based displays, night vision goggles and ambient light displays [31]. It will also be necessary to study older demographics of participants so that the effects of presbyopia on the user's interactions with AR 3D UIs can be further understood. The significant effect of depth on feature size also opens up interesting avenues for future research, where perhaps the vergence-accommodation reflex can be exploited via dynamic UIs to provide further benefits to user perception. Further, more research is needed to transfer the results and evaluate the effects for non-textual stimuli requiring luminance differences, such as gaze rays [18, 45], as well as to investigate secondary perceptual effects [14], such as night vision changes, affect, etc. Future work in this direction could lead to a better understanding of augmented human vision across technologies.

ACKNOWLEDGMENTS

This material includes work supported in part by the National Science Foundation under Collaborative Award Numbers 1800961, 1800947, and 1800922 (Dr. Ephraim P. Glinert, IIS) to the University of Central Florida, University of Florida, and Stanford University respectively; the Office of Naval Research under Award Numbers N00014-17-1-2927 and N00014-18-1-2964 (Dr. Peter Squire, Code 34); and the AdventHealth Endowed Chair in Healthcare Simulation (Prof. Welch). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the supporting institutions.

REFERENCES

[1] M. S. Banks, J. Kim, and T. Shibata. 2013. Insight into Vergence/Accommodation Mismatch. In *Head- and Helmet-Mounted Displays XVIII: Design and Applications*, Peter L. Marasco and Paul R. Havig (Eds.), Vol. 8735. International Society for Optics and Photonics, SPIE, 59 – 70.

[2] D. Bauer and C. R. Cavonius. 1990. *Ergonomic Aspects of Visual Display Terminals.* London: Taylor & Francis, Chapter Improving the legibility of visual display units through contrast reversal.

[3] C. Blehm, S. Vishnu, A. Khattak, S. Mitra, and R. W. Yee. 2005. Computer Vision Syndrome: A Review. *Survey of Ophthalmology* 50, 3 (2005), 253–262.

[4] G. Bruder, F. A. Sanz, A. Olivier, and A. Lecuyer. 2015. Distance estimation in large immersive projection systems, revisited. In *2015 IEEE Virtual Reality (VR).* 27–32.

[5] A. Buchner and N. Baumgartner. 2007. Text – background polarity affects performance irrespective of ambient illumination and colour contrast. *Ergonomics* 50, 7 (2007), 1036–1063.

[6] A. Buchner, S. Mayr, and M. Brandt. 2009. The advantage of positive text-background polarity is due to high display luminance. *Ergonomics* 52, 7 (2009), 882–886.

[7] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. Leonard. 2016. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics* 32, 6 (2016), 1309–1332.

[8] C. Cajochen, S. Frey, D. Anders, J. Späti, M. Bues, A. Pross, R. Mager, A. Wirz-Justice, and O. Stefani. 2011. Evening exposure to a light-emitting diodes (LED)-backlit computer screen affects circadian physiology and cognitive performance. *Journal of Applied Physiology* 110, 5 (2011), 1432–1438.

[9] FW Campbell and AH Gregory. 1960. Effect of size of pupil on visual acuity. *Nature* 187, 4743 (1960), 1121–1123.

[10] F. W. Campbell and K. Durden. 1983. The Visual Display Terminal Issue: A Consideration of Its Physiological, Psychological and Clinical Background. *Ophthalmic and Physiological Optics* 3, 2 (1983), 175–192.

[11] A. H. S. Chan and P. S. K. Lee. 2005. Effect of display factors on Chinese reading times, comprehension scores and preferences. *Behaviour and Information Technology* 24 (2005), 81–91.

[12] J. Cohen. 2013. *Statistical Power Analysis for the Behavioral Sciences.* Academic Press. Google-Books-ID: rEe0BQAAQBAJ.

[13] S. Debernardis, M. Fiorentino, M. Gattullo, G. Monno, and A. E. Uva. 2014. Text Readability in Head-Worn Displays: Color and Style Optimization in Video versus Optical See-Through Devices. *IEEE Transactions on Visualization and Computer Graphics* 20, 1 (2014), 125–139.
[14] A. Erickson, G. Bruder, P. J. Wisniewski, and G. Welch. 2020. Examining Whether Secondary Effects of Temperature-Associated Virtual Stimuli Influence Subjective Perception of Duration. In *Proceedings of IEEE Virtual Reality (VR)*. 493–499.
[15] A. Erickson, K. Kim, G. Bruder, and G. Welch. 2020. Effects of Dark Mode Graphics on Visual Acuity and Fatigue with Virtual Reality Head-Mounted Displays. In *Proceedings of IEEE Virtual Reality (VR)*. 434–442.
[16] A. Erickson, K. Kim, G. Bruder, and G. Welch. 2020. Exploring the Limitations of Environment Lighting on Optical See-Through Head-Mounted Displays. In *Proceedings of ACM Conference on Spatial User Interfaces (SUI)*. 1–8.
[17] A. Erickson, K. Kim, G. Bruder, and G. Welch. 2020. A Review of Visual Perception Research in Optical See-Through Augmented Reality. In *Proceedings of the International Conference on Artificial Reality and Telexistence & Eurographics Symposium on Virtual Environments*. 1–8.
[18] A. Erickson, N. Norouzi, K. Kim, J. LaViola Jr., G. Bruder, and G. Welch. 2020. Effects of Depth Information on Visual Target Identification Task Performance in Shared Gaze Environments. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (2020), 1934–1944.
[19] M. Fiorentino, S. Debernardis, A. E. Uva, and G. Monno. 2013. Augmented Reality Text Style Readability with See-Through Head-Mounted Displays in Industrial Context. *Presence: Teleoperators and Virtual Environments* 22, 2 (2013), 171–190.
[20] International Organization for Standardization (ISO). 1994. ISO 5725-2:1994 - Accuracy (Trueness and Precision) of Measurement Methods and Results. Basic Methods for the Determination of Repeatability and Reproducibility of a Standard Measurement Method.
[21] International Organization for Standardization (ISO). 2009. 8597:1994 and 8596:2009 - Ophthalmic Optics. Visual Acuity Testing. Standard Optotype and its Presentation.
[22] J. Gabbard, D. Mehra, and J. E. Swan. 2019. Effects of AR Display Context Switching and Focal Distance Switching on Human Performance. *IEEE Transactions on Visualization and Computer Graphics* 25, 6 (2019), 2228–2241.
[23] J. Gabbard, J. E. Swan, J. Zedlitz, and W. W. Winchester. 2010. More than meets the eye: An engineering study to empirically examine the blending of real and virtual color spaces. In *Proceeding of IEEE Virtual Reality (VR)*. 79–86.
[24] J. L. Gabbard, J. E. Swan, D. Hix, S. Kim, and G. Fitch. 2007. Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications. In *Proceedings of the IEEE Virtual Reality Conference*. 35–42.
[25] M. Gattullo, A. E. Uva, M. Fiorentino, and G. Monno. 2015. Effect of Text Outline and Contrast Polarity on AR Text Readability in Industrial Lighting. *IEEE Transactions on Visualization and Computer Graphics* 21, 5 (2015), 638–651.
[26] J. Harris. 2014. *Sensation and Perception*. SAGE.
[27] S. Higuchi, Y. Motohashi, Y. Liu, M. Ahara, and Y. Kaneko. 2015. Effects of VDT tasks with a bright display at night on melatonin, core temperature, heart rate, and sleepiness. *Journal of Applied Physiology* 94, 5 (2015), 1773–1776.
[28] J. T. Holladay. 2004. Visual acuity measurements. *Journal of Cataract & Refractive Surgery* 30, 2 (Feb. 2004), 287–290. https://doi.org/10.1016/j.jcrs.2004.01.014
[29] W. Huang, L. Alem, and M. A. Livingston. 2012. *Human Factors in Augmented Reality Environments*. Springer Science & Business Media.
[30] Tetsuri Inoue and Hitoshi Ohzu. 1997. Accommodative responses to stereoscopic three-dimensional display. *Applied Optics* 36, 19 (July 1997), 4509–4515. https://www.osapublishing.org/ao/abstract.cfm?uri=ao-36-19-4509 Publisher: Optical Society of America.
[31] K. Kim, M. Billinghurst, G. Bruder, H. Duh, and G. Welch. 2018. Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008–2017). *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 24, 11 (2018), 2947–2962.
[32] K. Kim, A. Erickson, A. Lambert, G. Bruder, and G. Welch. 2019. Effects of Dark Mode on Visual Fatigue and Acuity in Optical See-Through Head-Mounted Displays. In *ACM Symposium on Spatial User Interaction (SUI)*. 9:1–9:9.
[33] K. Kiyokawa, M. Billinghurst, B. Campbell, and E. Woods. 2003. An occlusion-capable optical see-through head mount display for supporting co-located collaboration. In *Proceeding of the International Symposium on Mixed and Augmented Reality (ISMAR)*. 133–141.
[34] T. R. Knapp. 1990. Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing research* 39, 2 (1990), 121–123.
[35] F. L. Kooi and A. Toet. 2004. Visual comfort of binocular and 3D displays. *Displays* 25, 2–3 (2004), 99–108.
[36] K. Krösl, D. Bauer, M. Schwärzler, H. Fuchs, G. Suter, and M. Wimmer. 2018. A VR-based user study on the effects of vision impairments on recognition distances of escape-route signs in buildings. *Visual Computer* 34, 6-8 (2018), 911–923.
[37] W. M. Kuzon Jr, M. G. Urbanchek, and S. McCabe. 1996. The seven deadly sins of statistical analysis. *Annals of plastic surgery* 37, 3 (1996), 265–272.
[38] M. Lambooij, M. Fortuin, W. Ijsselsteijn, B. Evans, and I. Heynderickx. 2010. Measuring visual fatigue and visual discomfort associated with 3-D displays. *Journal of the Society for Information Display* 18, 11 (2010), 931–943.
[39] Y.H. Lee, T. Zhan, and S.T. Wu. 2019. Prospects and challenges in augmented reality displays. *Virtual Real. Intell. Hardw.* 1, 1 (2019), 10–20.
[40] D. Löffler, L. Giron, and J. Hurtienne. 2017. Night Mode, Dark Thoughts: Background Color Influences the Perceived Sentiment of Chat Messages. In *Proceeding of INTERACT*. 184–201.
[41] L. W. MacDonald. 1999. Using color effectively in computer graphics. *IEEE Computer Graphics and Applications* 19, 4 (1999), 20–35.

[42] S. Mayr and A. Buchner. 2010. After-effects of TFT-LCD display polarity and display colour on the detection of low-contrast objects. *Ergonomics* 53, 7 (2010), 914–925.
[43] C. Merenda, M. Smith, J. Gabbard, G. Burnett, and D. Large. 2016. Effects of real-world backgrounds on user interface color naming and matching in automotive AR HUDs. In *IEEE VR 2016 Workshop on Perceptual and Cognitive Issues in AR (PERCAR)*. 1–6.
[44] A. A. Michelson. 1927. *Studies in optics*. University Press, Chicago.
[45] N. Norouzi, A. Erickson, K. Kim, R. Schubert, J. LaViola Jr., G. Bruder, and G. Welch. 2019. Effects of Shared Gaze Parameters on Visual Target Identification Task Performance in Augmented Reality. In *ACM Symposium on Spatial User Interaction (SUI)*. 12:1–12:11.
[46] E. Peillard, Y. Itoh, G. Moreau, J. M. Normand, A. Lécuyer, and F. Argelaguet. 2020. Can Retinal Projection Displays Improve Spatial Perception in Augmented Reality?. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 80–89.
[47] C. Piepenbrock, S. Mayr, and A. Buchner. 2014. Positive Display Polarity Is Particularly Advantageous for Small Character Sizes: Implications for Display Design. *Human Factors* 56, 5 (2014), 942–951.
[48] C. Piepenbrock, S. Mayr, and A. Buchner. 2014. Smaller pupil size and better proofreading performance with positive than with negative polarity displays. *Ergonomics* 57, 11 (2014), 1670–1677.
[49] C. Piepenbrock, S. Mayr, I. Mund, and A. Buchner. 2013. Positive display polarity is advantageous for both younger and older adults. *Ergonomics* 56, 7 (2013), 1116–1124.
[50] R. M. Schiffman, M. D. Christianson, G. Jacobsen, J. D. Hirsch, and B. L. Reis. 2000. Reliability and validity of the ocular surface disease index. *Archives of Ophthalmology* 118, 5 (2000), 615–621.
[51] S. Schmidt, G. Bruder, and F. Steinicke. 2017. Moving towards Consistent Depth Perception in Stereoscopic Projection-Based Augmented Reality. In *Proceedings of the 27th International Conference on Artificial Reality and Telexistence and 22nd Eurographics Symposium on Virtual Environments (ICAT-EGVE '17)*. Eurographics Association, Goslar, DEU, 161–168.
[52] M. Schrepp, A. Hinderks, and J. Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 6 (2017), 103.
[53] S. Soãres. 2019. *Optical Scatter—Techniques and Analysis*. Springer, Cham, 1–31.
[54] S. K. Sridharan, J. D. Hincapie-Ramos, D. R. Flatla, and P. Irani. 2013. Color Correction for Optical See-Through Displays Using Display Color Profiles. In *Proceeding of the ACM Symposium on Virtual Reality Software and Technology (VRST)*. 231–240.
[55] J Edward Swan, Gurjot Singh, and Stephen R Ellis. 2015. Matching and reaching depth judgments with real and augmented reality targets. *IEEE transactions on visualization and computer graphics* 21, 11 (2015), 1289–1298. Publisher: IEEE.
[56] S. Taptagaporn and S. Saito. 1990. How display polarity and lighting conditions affect the pupil size of VDT operators. *Ergonomics* 33 (1990), 201–208.
[57] H. Walker, W. Hall, and J. Hurst. 1990. *Clinical Methods: The History, Physical, and Laboratory Examinations*. Butterworths, Boston.
[58] A.-H. Wang, J.-J. Fang, and C.-H. Chen. 2003. Effects of VDT leading-display design on visual performance of users in handling static and dynamic display information dual-tasks. *International Journal of Industrial Ergonomics* 32 (2003), 93–104.
[59] C. Weiland, A.-K. Braun, and W. Heiden. 2009. Colorimetric and photometric compensation for optical see-through displays. *Universal Access in Human-Computer Interaction* 5615 (2009), 603–612.
[60] G. Welch, G. Bruder, P. Squire, and R. Schubert. 2019. *Anticipating Widespread Augmented Reality: Insights from the 2018 AR Visioning Workshop*. Technical Report. University of Central Florida and Office of Naval Research.
[61] Y. Zhao, M. Hu, S. Hashash, and S. Azenkot. 2017. Understanding Low Vision People's Visual Perception on Commercial Augmented Reality Glasses. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. 4170–4181.
[62] Y. Zhao, S. Szpiro, and S. Azenkot. 2015. ForeSee: A Customizable Head-Mounted Vision Enhancement System for People with Low Vision. *Proceedings of the International ACM SIGACCESS Conference on Computers & Accessibility* (2015), 239–249.