
Electronic Theses and Dissertations, 2020-

2023

Understanding, Modeling, and Simulating the Discrepancy Between Intended and Perceived Image Appearance on Optical See-Through Augmented Reality Displays

Austin Erickson
University of Central Florida

 Part of the [Computer Sciences Commons](#)

Find similar works at: <https://stars.library.ucf.edu/etd2020>

University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2020- by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Erickson, Austin, "Understanding, Modeling, and Simulating the Discrepancy Between Intended and Perceived Image Appearance on Optical See-Through Augmented Reality Displays" (2023). *Electronic Theses and Dissertations, 2020-*. 1555.
<https://stars.library.ucf.edu/etd2020/1555>



UNDERSTANDING, MODELING, AND SIMULATING THE DISCREPANCY BETWEEN
INTENDED AND PERCEIVED IMAGE APPEARANCE ON OPTICAL SEE-THROUGH
AUGMENTED REALITY DISPLAYS

by

AUSTIN ERICKSON

M.Sc. Computer Science, University of Central Florida, 2021

B.Sc. Computer Science, University of South Florida, 2018

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2023

Major Professor: Gregory F. Welch

© 2023 Austin Erickson

ABSTRACT

Augmented reality (AR) displays are transitioning from being primarily used in research and development settings, to being used by the general public. With this transition, these displays will be used by more people, in many different environments, and in many different contexts. Like other displays, the user's perception of virtual imagery is influenced by the characteristics of the user's environment, creating a discrepancy between the intended appearance and the perceived appearance of virtual imagery shown on the display. However, this problem is much more apparent for optical see-through AR displays, such as the HoloLens. For these displays, imagery is superimposed onto the user's view of their environment, which can cause the imagery to become transparent and washed out in appearance from the user's perspective. Any change in the user's environment conditions or in the user's position introduces changes to the perceived appearance of the AR imagery, and current AR displays do not adapt to maintain a consistent perceived appearance of the imagery being displayed. Because of this, in many environments the user may misinterpret or fail to notice information shown on the display.

In this dissertation, I investigate the factors that influence user perception of AR imagery and demonstrate examples of how the user's perception is affected for applications involving user interfaces, attention cues, and virtual humans. I establish a mathematical model that relates the user, their environment, their AR display, and AR imagery in terms of luminance or illuminance contrast. I demonstrate how this model can be used to classify the user's viewing conditions and identify problems the user is prone to experience when in these conditions. I demonstrate how the model can be used to simulate changes in the user's viewing conditions and to identify methods to maintain the perceived appearance of the AR imagery in changing conditions.

TABLE OF CONTENTS

LIST OF FIGURES	xiv
LIST OF TABLESxviii
CHAPTER 1: INTRODUCTION	1
1.1 Research Methodology	6
1.2 Thesis Statements	11
1.3 Outline	12
CHAPTER 2: LITERATURE REVIEW	14
2.1 Contrast	14
2.1.1 Contrast of Additive Imagery	16
2.1.2 Contrast Polarity	19
2.2 Visual Attention Cues	22
2.2.1 Dichoptic Visual Cues	23
2.2.2 Virtual Humans	25
CHAPTER 3: PERCEPTION OF USER INTERFACES	27

3.1	Overview	28
3.2	Experiment I	29
3.2.1	Participants I	30
3.2.2	Materials I	30
3.2.2.1	AR Stimuli and Vision Modes	30
3.2.2.2	Physical Environment and Background	34
3.2.2.3	Physical Lighting	36
3.2.3	Methods I	36
3.2.3.1	Procedure I	37
3.2.3.2	Measures I	39
3.2.4	Hypotheses I	40
3.3	Results I	40
3.3.1	Visual Acuity	41
3.3.2	Usability	43
3.3.3	Subjective Preferences	44
3.4	Discussion I	45
3.4.1	Dark Mode Improves Visual Acuity	45

3.4.2	Dark Mode Improves Usability and Preference in Dark Environments . . .	47
3.4.3	Limitations	48
3.5	Experiment II	48
3.5.1	Participants II	49
3.5.2	Material II	50
3.5.2.1	Physical Setup	50
3.5.2.2	Virtual Environment	51
3.5.2.3	Task Stimuli	52
3.5.3	Methods II	53
3.5.3.1	Procedure II	53
3.5.4	Hypotheses II	56
3.6	Results II	56
3.6.1	Visual Acuity	58
3.6.2	Visual Fatigue	60
3.6.3	Usability	60
3.6.4	Subjective Preferences	60
3.7	Discussion II	62

3.7.1	Dark Mode Improves Visual Acuity Only in Dim Lighting Conditions . . .	62
3.7.2	Dark Mode Decreases Visual Fatigue	65
3.7.3	Preference of Dark Mode over Light Mode	65
3.7.4	Perimeter Lighting Showed no Significant Effects	66
3.8	Summary	66
CHAPTER 4: PERCEPTION OF ATTENTION CUES		69
4.1	Overview	70
4.2	Experiment 1	70
4.2.1	Participants I	71
4.2.2	Materials I	71
4.2.3	Methods I	76
4.2.3.1	Study Design	76
4.2.3.2	Measures I	77
4.2.3.3	Procedure I	78
4.2.3.4	Hypotheses I	79
4.3	Results 1	79
4.3.1	Hue	82

4.4	Experiment II	84
4.4.1	Materials II	84
4.4.2	Methods II	84
4.4.2.1	Study Design	84
4.4.2.2	Measures II	85
4.4.2.3	Procedure II	86
4.4.2.4	Hypotheses II	86
4.5	Results II	87
4.6	General Discussion	88
4.6.1	Saliency of Color-Based Dichoptic Cues	89
4.6.2	Effect of Eye Dominance	93
4.6.3	Effects of Background Appearance	94
4.6.4	Subjective Qualities of Dichoptic AR Visual Cues	95
4.6.5	Using Dichoptic Cues on OST AR Displays	97
4.7	Summary	98
CHAPTER 5: PERCEPTION OF VIRTUAL HUMANS		100
5.1	Overview	101

5.2	Subjective Qualities of Transparent AR Virtual Humans	102
5.2.1	Methods	103
5.2.1.1	Image Generation	103
5.2.1.2	Experimental Procedure	107
5.2.1.3	Measures	108
5.2.1.4	Participants	110
5.2.2	Results	110
5.2.2.1	Validation Analyses	111
5.2.2.2	Human Faces	115
5.2.2.3	Avatar Faces	117
5.2.2.4	Humanness Implications	118
5.2.3	Discussion	118
5.2.3.1	Implications	121
5.2.3.2	Limitations & Future Directions	122
5.3	Subjective Qualities of Self-Avatar Appearance in OST AR	124
5.3.1	Experiment	125
5.3.1.1	Participants	125

5.3.1.2	Materials	125
5.3.1.3	Methods	128
5.3.1.4	Measures	128
5.3.1.5	Procedure	128
5.3.2	Results	130
5.3.3	Discussion	131
5.3.3.1	Implications	133
5.3.3.2	Limitations	133
5.4	Summary	134
CHAPTER 6: ESTABLISHING A MODEL		136
6.1	Modeling Contrast for OST Displays	138
6.1.1	Use with Luminance Units	145
6.1.2	Limitations of the Model	146
6.2	Photometric Measurements on Optical See-Through AR Displays	148
6.2.1	Apparatus	149
6.2.2	Methods	151
6.2.3	Procedure	152

6.2.4	Results and Discussion	153
6.2.4.1	Illuminance Measurements	153
6.2.4.2	Contrast Calculations	158
6.3	Effects of Contrast on AR Search Task Performance	161
6.3.1	Participants	161
6.3.2	Methods	162
6.3.2.1	Study Task	163
6.3.3	Materials	164
6.3.4	Procedure	165
6.3.5	Measures	165
6.3.6	Hypotheses	166
6.3.7	Results	166
6.3.7.1	Linear Regressions	167
6.3.7.2	Non-Linear Regressions	168
6.4	Discussion	169
6.4.1	Contrast Predicts User Performance	169
6.4.2	Contrast Predicts Subjective Task Perceptions	172

6.5	Summary	172
CHAPTER 7: MODEL APPLICATIONS		174
7.1	Model Use Cases	175
7.1.1	Adaptive AR Systems	175
7.1.1.1	Establishing Model Parameters	177
7.1.1.2	Evaluating if Adaptations are Needed	178
7.1.1.3	Decide Which Adaptations to Perform	180
7.1.2	Simulating the Perceived Appearance of Imagery on Optical See-Through AR Displays	182
7.1.2.1	Methods	183
7.1.2.2	Limitations	184
7.2	Design Guidelines and Considerations	185
7.3	Summary	186
CHAPTER 8: CONCLUSION		188
8.1	Thesis Statement One - Identify Problems	189
8.2	Thesis Statement Two - Establish a Model	190
8.3	Thesis Statement Three - Apply the Model	191

APPENDIX A: IRB APPROVAL LETTERS	194
LIST OF REFERENCES	200

LIST OF FIGURES

Figure 1.1: This figure depicts how the perceived appearance of imagery on OST AR displays is a blending of the attenuated environment light and the light emitted by the AR display representing the AR imagery.	1
Figure 1.2: A figure depicting the difference between the intended and perceived appearance of text shown on OST AR display in varying colors and in varying lighting conditions.	2
Figure 1.3: This figure depicts a feedback loop where the OST AR display has information on the perceived appearance of emitted AR imagery.	4
Figure 2.1: Illustration of a contrast sensitivity function, where contrast sensitivity is the reciprocal of the lowest contrast stimulus a person is able to distinguish. . . .	15
Figure 3.1: Experimental setup and the images for the background posters used in chapter three experiment one.	31
Figure 3.2: Illustrations showing the AR text-reading task participants had to read during experiment one of chapter three.	34
Figure 3.3: Illustrations of the different experimental conditions of chapter three experiment one.	37
Figure 3.4: Results for the <i>visual acuity</i> tests of chapter 3 experiment 1.	41

Figure 3.5: Results for the <i>usability</i> estimates of chapter 3 experiment 1 using the UEQ-S questionnaire.	42
Figure 3.6: A figure demonstrating the results of the subjective preferences survey of chapter 3 experiment 1.	45
Figure 3.7: Annotated photo showing a participant in the study with the HMD seated at the table.	50
Figure 3.8: Illustrations of visual stimuli used in chapter 3 experiment 3.	51
Figure 3.9: Illustrations depicting the VR text-reading task participants had to perform during chapter 3 experiment 3.	54
Figure 3.10: Figure depicting the visual acuity tests of study two.	57
Figure 3.11: This figure depicts the visual fatigue results of study two.	59
Figure 3.12: Figure depicting the usability results for study two.	61
Figure 3.13: Figure depicting the subjective results of study two.	61
Figure 3.14: Illustration of the screen door effect for UIs of varying contrast polarity. . .	63
Figure 4.1: The figure depicts a side by side view of the visual cues used in experiment 1 and 2 of chapter 4	72
Figure 4.2: The figure depicts the color shifts used to generate our visual cues in the HSV color space.	73

Figure 4.3: This figure depicts a comparison of the false negatives from experiment one across cue type, cue level, and background.	80
Figure 4.4: This figure depicts a comparison of the false negatives from experiment one across cue type and cue level.	82
Figure 4.5: This figure depicts the overall effects of the hue of the grid of stimulus objects on the false negative rate from experiment 1 of chapter 3.	83
Figure 4.6: This figure depicts the subjective noticeability, urgency, and comfortability results from experiment 2.	88
Figure 4.7: This figure compares the false negative rates between the hue-based cues and saturation-based cues.	91
Figure 5.1: Illustration depicting the setup of the perceptual matching task used to generate imagery for study one.	103
Figure 5.2: Figure depicting the visual stimuli used in study one.	104
Figure 5.3: Figure depicting the average humanness between images used in study one. .	112
Figure 5.4: A bar chart of the average angry score with standard error bars.	114
Figure 5.5: A bar chart of the average happy score with standard error bars.	114
Figure 5.6: Representation of the 3-way effect of humanness, opacity, and race for study one.	115
Figure 5.7: The relationship between <i>happy</i> and <i>humanness</i> for avatar faces at each opacity level.	118

Figure 5.8: Pearson correlation coefficient matrix comparing humanness, competent, friendly, and dangerous.	119
Figure 5.9: Figure depicting the physical setup of experiment two.	126
Figure 5.10: Figure depicting the stimuli used in experiment two.	127
Figure 5.11: Bar charts showing effects of environment lighting on avatar skin colors and hair colors.	129
Figure 6.1: Figure depicting several types of contrast relevant to AR applications.	139
Figure 6.2: Illustration of the setup used to perform photometric measurements.	150
Figure 6.3: Figure depicting the mean illuminance values for each of the 14 tested colors at each of the 5 environment lighting conditions.	154
Figure 6.4: This figure depicts Michelson illuminance contrasts of each of the primary, secondary, and tertiary colors.	155
Figure 6.5: This figure depicts the residuals between the extended illuminance contrast values and the Michelson contrast values.	156
Figure 6.6: This figure depicts the residuals between the modeled contrast values and naive contrasts calculated based on the relative luminance of each of the colors.	157
Figure 6.7: These figures show the effects of modeled luminance contrast on user perfor- mance and subjective perception of the AR experience.	168

LIST OF TABLES

Table 1.1: The table below maps the chapters in this dissertation to thesis statements they address.	12
Table 3.1: 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.	34
Table 5.1: Final linear regressions for the humanness measure of study one.	111
Table 6.1: This table shows the RGB values for the primary, secondary, and tertiary colors measured on the HoloLens 2.	151
Table 6.2: Table depicting the illuminance values gathered from the HoloLens 2.	153
Table 6.3: This table similarly shows the <i>Michelson contrast</i> values between each color and the black level of the display on the HoloLens 2.	153

CHAPTER 1: INTRODUCTION

Augmented reality (AR) displays allow users to view 3D virtual imagery that is registered with respect to their physical environment [6]. Over the last several years, we have seen the emergence of more and more of these displays, including optical see-through (OST) displays like the Microsoft HoloLens 2 and Magic Leap 2, as well as video see-through (VST) displays such as the passthrough mode on the Meta Quest displays. Whereas at first, these displays were primarily used by researchers and in limited application areas, display manufacturers are beginning to target the population of general consumers. As this technology is adopted by the general public, these displays will be used within a much wider range of environments and contexts. In many cases, these environments will be quite different from the controlled environments where most AR research and development has so far taken place.

This wider range of environments introduces problems when it comes to the perception of imagery on AR displays, where the perceived appearance of AR imagery depends on the characteristics of the user's physical environment. While this is true for any display, this is particularly problematic

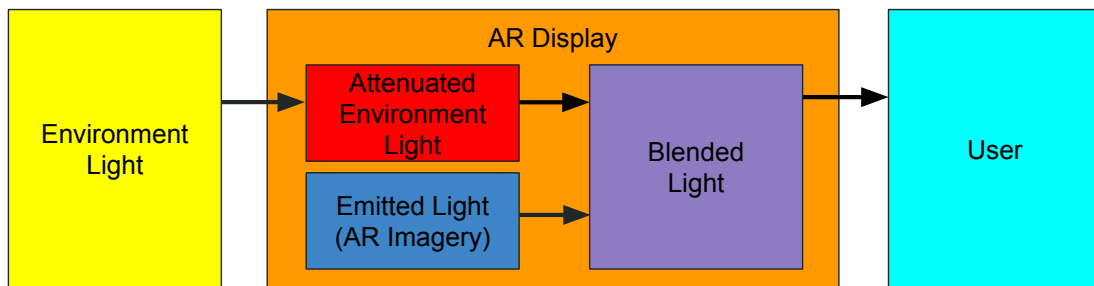


Figure 1.1: This figure depicts how the perceived appearance of imagery on OST AR displays is a blending of the attenuated environment light and the light emitted by the AR display representing the AR imagery.

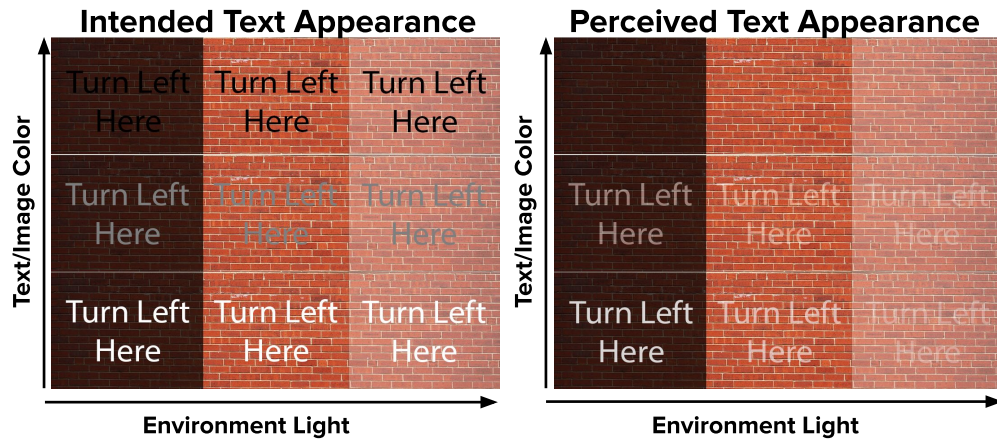


Figure 1.2: A figure depicting the difference between the intended and perceived appearance of text shown on OST AR display in varying colors and in varying lighting conditions. In this scenario, the user is facing a brick wall and is observing a message on their OST AR display that reads “Turn Left Here.” Darker text is perceived to be more transparent, where black text cannot be perceived at all and white text is least transparent. Brighter environment lighting conditions lead to more transparent and “washed out” imagery.

for OST displays, such as the HoloLens and Magic Leap, as well as spatial AR systems that use projectors to display imagery. Both OST displays and projectors are *additive displays*, in that they add light on top of the user’s view of their physical environment. Because of this, imagery on these displays takes on characteristics of the user’s environment and is perceived to be transparent or “washed out” to the user compared to how the imagery was intended to be perceived [64, 66]. This mechanism of the effect is illustrated in a flowchart in figure 1.1, where the magnitude of the effect depends on the lighting conditions in the user’s environment. Figure 1.2 shows a side by side representation of how the perceived appearance of imagery shown on OST AR displays changes depending on the user’s environment conditions, where in dark conditions, imagery is more opaque and apparent, while in brighter conditions, it may be difficult, or even impossible for the user to perceive the AR imagery [42].

With AR displays emerging on the consumer market, there will be many environments in which the

user's perception of imagery shown on the display will be quite different from how the designers of the application intended the imagery to appear. While at first, this may simply sound like an problem of aesthetics, AR displays are a source of entertainment and information, and the quality of the AR imagery can cause the user to miss important information and/or misinterpret information being conveyed, which may negatively affect the user's decisions and actions.

In general, the applications of AR displays can be split into two broad categories based on their purpose: to entertain or to inform. In either case, the user is observing virtual imagery shown on the display and is then making decisions and actions based on their perception of that imagery. If virtual imagery is perceived differently than intended, then the user is prone to make different decisions and actions as an indirect consequence of their perception. In a gaming context, the user may fail to notice a visual cue that they are meant to react to, which may cause them to lose or otherwise perform poorly in the game. In a navigation context, the user may fail to notice a cue informing them on where to turn or they may misinterpret directions shown on the display. While these previous examples are somewhat harmless, AR displays are also beginning to be used in medical and defense contexts, where misinterpreting or failing to notice the presence of vital information on the display can have more significant consequences [115, 156].

There are several potential approaches to solving this problem for AR displays. From a top-down perspective, people designing the appearance of imagery to be shown on AR displays can choose certain colors and styles that are more robust to changing environment conditions, for instance choosing colors combinations for user interfaces (UIs) that are not only high contrast but also dark mode style to avoid introducing visual elements that are prone to be perceived as transparent [48, 94]. In practice, design decisions can slightly extend the range of environments in which the perceived appearance of AR imagery will closely resemble its intended appearance. However, there is a limit to what can be done through design decisions alone, as any AR display has a fixed limit in the brightness of imagery that can be displayed, and once the user's environment lighting

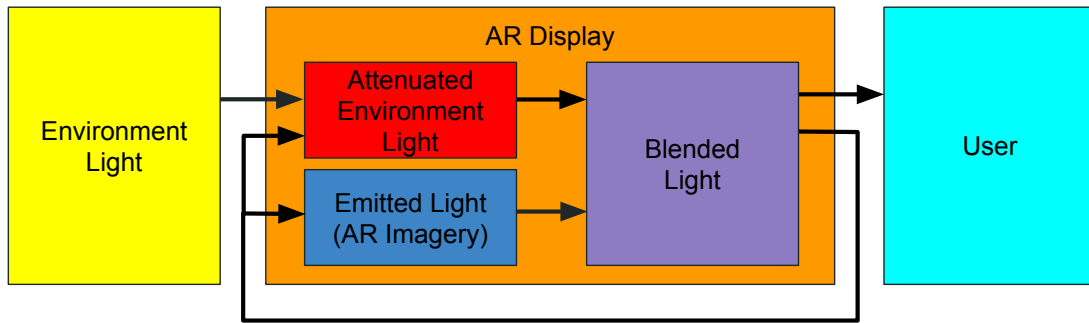


Figure 1.3: This figure depicts a feedback loop where the OST AR display has information on the perceived appearance of emitted AR imagery. In such a case, the AR system can adapt parameters within its control, such as the amount of environment light attenuated, or the representation of the virtual imagery. Future systems could even adapt characteristics of the user's physical environment. The effects of such changes on the perceived appearance of the AR imagery can be observed by the AR system so that the perceived appearance of the imagery is maintained or improved.

conditions reach a certain intensity, changes made solely to the virtual imagery are not effective at improving its perceived appearance.

Ideally, the AR system should be able to adapt on its own according to the user's current environment conditions. In this manner, the designers of AR applications do not need to worry about how imagery in their application will be perceived in the countless different environments users could be experiencing it in. Such a system could adapt the perceived appearance of virtual imagery in several different ways. It could make direct changes to the imagery, such as choosing brighter/darker colors within the imagery according to the user's viewing conditions. It could also modify aspects of the display or AR system within its control, such as changes to the brightness level of the display, or changes to the optical elements of the display. Finally, the AR system may have direct control over certain elements in the user's physical environment, such as through Internet of Things (IoT), which may allow the system to control lights, shades, curtains, or other aspects of the user's environment to control the perceived appearance of the AR imagery.

For this adaptive system to work well, it should understand how the changes it is making will affect the perceived appearance of imagery being displayed. Ideally, any changes made by the system should either maintain the perceived appearance of the virtual imagery, or bring the perceived appearance of the imagery closer to its intended appearance. Using a mathematical model, the system could simulate the effects of potential changes to ensure that the system's actions are accomplishing these goals. Figure 1.3 illustrates how such a model can be used to create a feedback loop that gives the AR display information on the perceived appearance of the AR imagery, and allows it to see how changes to parameters within its control affect the perceived appearance of the imagery.

There are several reasons why such an adaptive AR system has not yet been realized. One is that the nature of this particular problem is highly interdisciplinary and requires an understanding of the domains of computer science, human-computer interaction, visual perception, and optics. There may also be a perception amongst the community that this is a problem of aesthetics that will eventually be solved through incremental improvements to AR display technology. Additionally, the mathematical model mentioned above does not yet exist, and existing models must be adapted to consider the many factors at play in the perception of imagery shown on AR displays, including the appearance of the user's environment, the lighting conditions within the user's environment, and the characteristics of the AR display being used.

This dissertation provides several contributions in this interdisciplinary problem area. First, I demonstrate several examples of how the perceived appearance of AR imagery can be negatively affected by the user's environment and the characteristics of their AR display. I also establish a mathematical model to describe the perceived appearance of imagery on AR displays, demonstrate how this new model can be used within current and future AR displays to improve the AR experience of the user, and demonstrate why incremental improvements to AR display technology do not sufficiently address the underlying problem on their own. The following section provides additional information on the research objectives and methods described in this dissertation.

1.1 Research Methodology

My research focuses on the subset of AR displays that are *OST AR displays* [44, 86]. Such displays are widely relevant to today’s world, where new OST displays in the form of smart glasses and head-mounted displays are being released each year. These displays allow the user to visualize information that can enhance or extend the user’s perception, for example allowing the user to visualize thermal infrared information in their environment [47, 49, 52], the gaze direction of nearby people [50, 51, 124], facial expressions of people across further distances [27], and even to help user’s see in general [110]. As mentioned in the previous section, the perceived appearance of imagery shown on these displays is dependent on the characteristics of the user’s environment, which typically cannot be directly controlled by the AR display. These particular displays are typically worn by the user as opposed to being placed in a static position relative to their environment. Thus, these displays are prone to more dynamic environment conditions compared to *spatial AR displays* [13], which typically use fixed-position projectors to display imagery on static surfaces. In such a scenario, both the display system and the environment it is projecting upon are static, whereas for OST AR displays these factors are both dynamic and changes to these factors are not as easily controlled or anticipated by the AR system. For example, on OST AR displays, the simple act of the user turning their head or moving to a different location can significantly affect their perception of virtual imagery shown on the display. Additionally, these displays have less control over the user’s perception of their physical environment compared to *VST AR displays* [23], in which the user’s physical environment can be reliably occluded by superimposed AR imagery and is not prone to effects such as color blending.

While primarily focused on OST AR displays, some of the research in this dissertation potentially applies to other types of AR displays. In particular, chapter 3 presents research relevant to both OST AR displays and VST AR or VR displays. It is also possible that the effects found for OST AR

displays in this dissertation could transfer to spatial AR displays, due to the similarities between these types of displays. However, the transfer of results from OST AR to other display types is not specifically investigated in this dissertation.

In the first several chapters of this dissertation, I investigate how user perception and performance is affected by their physical environment when interacting with imagery shown on AR displays. In these chapters, I investigate three specific types of imagery that are commonly shown on AR displays: UIs in chapter 3, attention cues in chapter 4, and virtual humans in chapter 5. A review of the existing literature for visual perception, AR displays, and for these three particular domains is provided in chapter 2.

In chapter 3, I demonstrate that user performance in interacting with UIs is significantly affected by the appearance of their physical environment, the lighting conditions in their environment, and the appearance of the UIs being displayed, but that these effects are different depending on the type of AR display being used. In addition to investigating the users environment conditions, the studies in this chapter compare the contrast polarity of the UIs, comparing dark mode style UIs, consisting of white text on black backgrounds, to light mode style UIs, consisting of black text on white backgrounds. These effects observed in this chapter serve as motivation for AR systems that adapt to control the perceived appearance of AR imagery so that legibility of text, usability of UIs, and visual comfort of the user can be maintained. Portions of chapter 3 have been previously published in the following publications:

- [94] Kangsoo Kim, Austin Erickson, Alexis Lambert, Gerd Bruder, and Greg Welch. Effects of Dark Mode Visualization on Visual Fatigue and Acuity in Optical See-Through Head-Mounted Displays. In Proceedings of the 2019 ACM Symposium on Spatial User Interaction (SUI), pages 1–9, 2019.
- [41] Austin Erickson, Kangsoo Kim, Gerd Bruder, and Greg Welch. Effects of Dark Mode

Graphics on Visual Acuity and Fatigue with Virtual Reality Head-Mounted Displays. In Proceedings of IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pages 434–442, 2020.

- [45] Austin Erickson, Kangsoo Kim, Gerd Bruder, and Greg Welch. Dark/Light Mode Adaptation for Graphical User Interfaces on Near-Eye Displays. In ICAT-EGVE 2020 – International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments – Posters and Demos. The Eurographics Association, pages 1-2, 2020.
- [48] Austin Erickson, Kangsoo Kim, Alexis Lambert, Gerd Bruder, Michael Browne, and Greg Welch. An Extended Analysis on the Benefits of Dark Mode User Interfaces in Optical See-Through Head-Mounted Displays. ACM Transactions on Applied Perception, 18(3), pages 1-22, 2021.

In chapter 4, two user studies are described that investigate the effectiveness of visual attention cues in environments of different appearances and lighting levels, and for cues of several different colors and styles. This chapter compares traditional color-based visual attention cues to dichoptic variations of these attention cues, where the appearance of the cue is different between each of the user’s eyes. This chapter further reinforces that user perception of AR imagery is significantly affected by the appearance of their physical environment, the lighting conditions in their physical environment, and the perceived appearance of imagery on the AR display. The effects described in this chapter further motivate the need for AR systems that adapt to control the perceived appearance of AR imagery, so that the effectiveness of visual attention cues can be maintained across the many potential environment conditions AR users may find themselves in.

Portions of chapter 4 have been previously published in the following publication:

- [40] Austin Erickson, Gerd Bruder, and Greg Welch. Analysis of the saliency of color-based dichoptic cues in optical see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–16, 2022

In chapter 5, two user studies are described that investigate how user perception of virtual humans and avatars is affected by the lighting level within the user’s environment. It demonstrates that virtual humans are perceived to be less human like when observed in bright lighting conditions compared to dim lighting conditions, and that, if given the opportunity, users choose to adapt the appearance of their own avatars based on the lighting conditions in which they are being observed in. This chapter again reinforces that user perception of AR imagery is significantly affected by the lighting conditions within their physical environment and by the intended appearance of the virtual imagery. Additionally, it highlights a negative effect in user perception of virtual humans, where virtual humans wearing darker clothes, with darker hair, or with darker skin tones may be perceived to be less human like compared to other virtual humans depicted with otherwise brighter imagery. This introduces a potential racial bias in the presentation of virtual humans on OST and spatial AR displays, and further highlights the need for future AR systems to be able to adapt the perceived appearance of virtual imagery so that this bias can be mitigated.

Portions of chapter 5 have been previously published in the following publications:

- [131] Tabitha Peck, Jessica Good, Austin Erickson, Isaac Bynum, and Gerd Bruder. Effects of transparency on perceived humanness: Implications for rendering skin tones using optical see-through displays. *IEEE Transactions on Visualization & Computer Graphics*, 28(5), pages 2179-2189, 2022.
- [35] Doroodchi, Meelad, Priscilla Ramos, Austin Erickson, Hiroshi Furuya, Juanita Benjamin, Gerd Bruder, and Greg Welch. Effects of Optical See-Through Displays on Self-

Avatar Appearance in Augmented Reality. In 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pages 352-356, 2022.

- [38] Austin Erickson, Gerd Bruder, Greg Welch, Isaac Bynum, Tabitha Peck, Jessica Good, Perceived Humanness Bias in Additive Light Model Displays. Journal of vision. pages 1-1, 2022.

Together, these three chapters provide several examples of how the user's AR experience can be negatively affected by the appearance of their physical environment, the lighting conditions within that environment, and the intended appearance of the virtual imagery they are interacting with. These chapters provide support that future AR systems should adapt to the user's environment in order to prevent effects such as decreases in text legibility, decreases in the effectiveness of attention cues, and negative perceptions of virtual humans and avatars.

In chapter 6, I introduce a mathematical model that can relate the user, their environment, their AR display, and the imagery being observed in terms of luminance contrast or illuminance contrast. These factors were each found to significantly affect the user's perception of imagery shown on AR displays in chapters 3, 4, and 5. This model adapts Michelson's contrast equation to quantify a user's viewing conditions for their AR experience and determine the negative perceptual effects the user may experience while in these conditions. In this chapter, I demonstrate how photometric measurements can be performed on OST AR displays and on the user's environment to gather the parameters needed for the model. This chapter also details a user study that investigates how user performance in an AR search task is affected by the contrast of the virtual stimuli, as measured via the new extended contrast model. The results of this study indicate that user performance, confidence, and their perception of task difficulty can each be predicted in terms of contrast values measured via the new model. Additionally, it demonstrates that in certain environment conditions, user's are unaware that their performance is being negatively affected.

In chapter 7, I delve into the current and future applications of the new model presented in chapter 6. I demonstrate how the new contrast model can be used to simulate how potential changes made to the AR imagery, AR display, or the user's physical environment will affect the perceived appearance of the AR imagery. I also demonstrate the benefits and limitations of using the model to generate imagery that simulates the perceived appearance of AR imagery in different environment conditions, and on different AR displays. Finally, I provide a detailed description on how the new model can be used to help maintain or improve the perceived appearance of imagery across different environment conditions within AR systems that have varying levels of control over the underlying components of the AR display system.

Portions of chapter 6 and chapter 7 have been previously published and submitted for publication in the following publications:

- [46] Austin Erickson, Kangsoo Kim, Gerd Bruder, and Gregory F. Welch. 2020. Exploring the Limitations of Environment Lighting on Optical See-Through Head-Mounted Displays. In Proceedings of the 2020 ACM Symposium on Spatial User Interaction (SUI '20), pages 1–8, 2020.
- [39] Austin Erickson, Gerd Bruder, and Greg Welch. Adapting Michelson Contrast for use with Optical See-Through Displays. In 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pages 409-410, 2022.

1.2 Thesis Statements

Based on the research presented in this dissertation, I present three thesis statements that directly relate to the research objectives described in section 1.1:

- **TS1 - Identify Problems:** In increasingly common viewing conditions, AR users are susceptible to misperceiving or completely missing information shown on their AR display. Sometimes the user may not even be aware that their AR experience or their performance is being affected.
- **TS2 - Establish a Model:** The discrepancy between the intended and perceived appearance of AR imagery can be modeled in terms of luminance contrast, as a function of parameters specific to the user's viewing conditions. These parameters include the user's environment lighting conditions, the characteristics of their AR display, and the characteristics of the AR imagery.
- **TS3 - Apply the Model:** The model from TS2 can be used to evaluate the user's viewing conditions, simulate changes to their viewing conditions, and mitigate negative effects the user is prone to experience through changes made to the AR imagery, to the characteristics of the user's AR display, or to the user's environment.

Table 1.1: The table below maps the chapters in this dissertation to thesis statements they address.

	TS1 Identify Problems	TS2 Establish a Model	TS3 Model Applications
Ch. 3 - Perception of UIs	×		
Ch. 4 - Perception of Attention Cues	×		
Ch. 5 - Perception of Virtual Humans	×		
Ch. 6 - Establishing a Model	×	×	
Ch. 7 - Model Applications		×	×

1.3 Outline

The remainder of this dissertation is structured into a total of eight chapters as described below.

- **Chapter 2 - Background** This chapter provides background and explores related work relevant to this dissertation, including an overview of the human visual system, and research on AR displays, UIs, attention cues, and virtual humans.
- **Chapter 3 - Perception of User Interfaces** This chapter presents a series of two user studies that investigate the factors that influence user perception of UIs on AR displays.
- **Chapter 4 - Perception of Attention Cues** This chapter presents two user studies that investigate the factors that influence the effectiveness of visual attention cues on OST AR displays.
- **Chapter 5 - Perception of Virtual Humans** This chapter presents two user studies that investigate the factors that influence user perception of virtual avatars on AR displays.
- **Chapter 6 - Establishing a Model** This chapter presents an overview of the factors found to influence user perception of AR imagery, and establishes a model capable of describing a user's AR experience as a function of parameters describing the user's physical environment, AR display, and AR imagery being displayed.
- **Chapter 7 - Model Applications** This chapter describes how the model from chapter 6 can be used to identify and evaluate methods intended to maintain or improve the perceived appearance of AR imagery.
- **Chapter 8 - Conclusion** This chapter summarizes the above research, offers recommendations for future work in this domain, and concludes the paper.

CHAPTER 2: LITERATURE REVIEW

In this chapter, background is provided on the contrast sensitivity, contrast polarity, text legibility, visual attention cues, and their intersections with research involving AR displays. For more in depth reading on the human visual system and physiology of the eye from the ground up, we recommend referring to the online text by Stangor and Walinga [149].

2.1 Contrast

In the field of human visual perception, the ability of a person to distinguish a physical stimulus is typically based on the size and contrast of the stimulus compared to the environment. Distinguishing a visual stimulus based on its size is typically referred to as *visual acuity* and well-covered in the existing literature [54, 80, 98].

Contrast sensitivity is typically defined as a person's ability to distinguish a visual stimulus based on the differences in luminance between it and its environment [130, 133]. One way to calculate the contrast of basic visual stimuli is Michelson's contrast, which is characterized by the following equation, where I_{min} is the luminance of the stimulus (which is typically black) and I_{max} is the luminance of the background (which is typically white):

$$\textbf{Michelson Contrast: } \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \quad (2.1)$$

While this contrast can be measured using specialized light meters, there are multiple psychophysical tests to measure users' sensitivity to such contrasts. An example is the sine wave grating test, in which sine wave grating patterns are presented to the subject at varying contrast levels, spa-

tial frequencies, and direction, where subjects are tasked to identify the direction of the gratings (typically either vertical or horizontal). Contrast level is quantified as a ratio between the color of the sine pattern and the background behind it, and is varied by displaying the sine wave pattern to the user in different shades of grey, where darker shades offer more contrast than lighter shades. Spatial frequency can be thought of as size, and is quantified in cycles per degree of visual angle. It is similarly varied by increasing the amount of repeated sine wave patterns found in the set area that is presented to the subject.

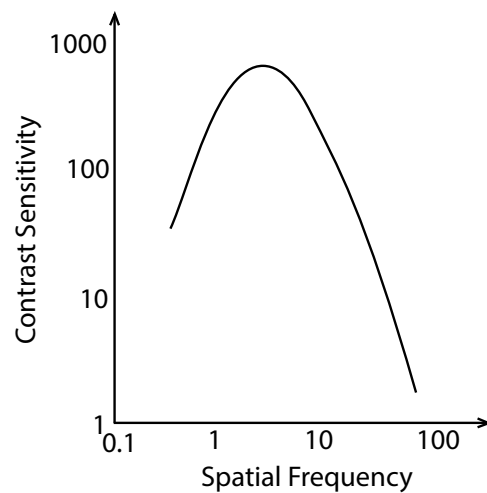


Figure 2.1: Illustration of a contrast sensitivity function, where contrast sensitivity is the reciprocal of the lowest contrast stimulus a person is able to distinguish.

When a person's contrast sensitivity levels and spatial frequency levels are plotted on the axes of a figure in logarithmic scale, then the resulting graph of the person's contrast sensitivity function typically takes on the approximate form of an inverted parabola (see figure 2.1). This shape is expected, as the stimulus is more difficult to identify when contrast is low, as there is little difference in color between the sine grating and the background, or when the spatial frequency is high, as there is little separation between repeated wave patterns on the stimulus.

Visual stimulus that falls outside a person’s contrast sensitivity function can be dealt with in two different manners, either by increasing the spatial frequency (e.g., its size), or by increasing the contrast. However on OST-HMDs, screen space is valuable due to the limited field of view on most devices. Because of this, increasing spatial frequency in reduced contrast situations only works so long as the image fits within the field of view of the device. With this upper bound on spatial frequency, any additional change to make imagery distinguishable to the user must come from adjusting either the virtual image itself or the conditions within the user’s physical environment.

In general, higher contrast is considered better when it comes to making something more easily distinguishable or readable. However, it has been suggested that there exists an optimal luminance contrast value for which greater contrast results in decreased readability of text [107], although the reasons for why this occurs are not completely understood.

2.1.1 Contrast of Additive Imagery

With OST AR imagery, there are several ways in which luminance contrast can be compared. It can be compared between two points in the user’s physical environment seen through the display, between two points in the virtual image itself, or between a point in the user’s environment and a point in the virtual image. For the work presented here, we focus primarily on contrast between points in the virtual image and contrast between the virtual image and the user’s physical environment.

When it comes to contrast involving a virtual image on an OST-HMD, there are several factors that can cause the image to be perceived with poor luminance contrast, which may make it difficult for the user to identify features within the image. These factors include environmental factors such as environment lighting conditions and the colors that comprise the user’s physical environment, as well as display factors such as the luminance capability of the display and the colors that comprise

the virtual image. Tinted visors attached to OST displays are often used as a practical solution for improving luminance contrast in response to these factors, where reducing environment lighting as it combines with the light emitted from the display has the effect of improving contrast within the virtual image itself, and between the image and the user's physical environment. However, this tint reduces the contrast within the user's physical environment, especially in dim lighting conditions, and changes the manner in which users' perceive color [113, 114].

To address perception issues in scenarios where virtual imagery has poor luminance contrast, researchers conducted studies that investigate the perception thresholds and design guidelines for augmented content in OST displays. For example, Harding et al. [74] presented an HMD simulation model that could simulate different see-through background images with overlaid white-color symbology, and showed that the perceptual quality of the symbology was greatly influenced by the visual complexity of the physical backgrounds, which was characterized by the standard deviation of small patches of luminance in the images. Beyond the white-color symbology, Harding et al. [75] further studied luminance and color contrast requirements while discussing color choices for effective symbology in OST-HMDs.

Gabbard et al. [62] also pointed out the challenges associated with low contrast imagery on OST AR displays, specifically in presenting text-based information in outdoor environments, where lighting conditions and environment appearance tend to be uncontrollable. They found that text legibility is highly affected by the text drawing style, the visual structure of the physical background, and the interaction between the two. Merenda et al. [121] considered in-car HUD interfaces and conducted a study that investigates the user's color identification performance in different color-blending circumstances where virtual imagery shown on the HUD blends with the appearance of the physical environment behind it. They found that participants generally chose brighter high contrast colors as compared to the original source color of the content, but certain colors, e.g., blue, green, and yellow, were more accurately identified.

There is also a continuous effort on developing display techniques and devices for improved luminance contrast and color reproduction. Itoh et al. [84, 85] presented a series of display methods to improve the visual quality of virtual content in OST-HMDs. For example, they proposed a color calibration method that pre-processes the input image to match the user-perceived color on the display to the original image color [84] and introduced a new light attenuation approach that forms virtual images by spatially subtracting colors instead of the traditional additive color approach [85]. Hincapié-Ramos et al. [78] addressed a similar problem by investigating methods of altering the appearance of text to preserve the intended color, and introduced methods of preserving the intended color while also increasing contrast. Donval et al. [34] proposed an adaptive smart filter for OST-HMDs that could change the filter transmission according to the background illumination conditions. Leykin and Tuceryan [108] examined predictions of the legibility of text by using machine learning based classifiers, however their work did not take into account the how text annotations appear more transparent as environment lighting conditions increase.

Several other methods have been established for improving the contrast of virtual text specifically, and these largely fall into methods that either alter the appearance of the text, the appearance of the area surrounding the text, or methods that reposition the text. Among these, Gabbard et al. [64] and Kim et al. [94] found that for OST-HMDs, the user experience is typically improved by utilizing bright colored text over dark colored backgrounds, which is the opposite of what users tend to prefer when using other display mediums such as virtual reality HMDs [41], where an interaction effect between the text appearance and virtual lighting occurs such that users perform better with light colored font in dark virtual environments and better with dark colored font in bright virtual environments.

When users were tasked with manually placing annotations, Jia et al. [88] found that users preferred text labels over uniform surfaces, and tended to place annotations over the sky. They further established a method which employs image analysis to identify such preferred locations automatically

for annotations. Orlosky et al. [128] developed a similar automatic annotation system, however their procedure involves prioritizing locations which are darker and uniform, so that annotations are less affected by environment illuminance.

In 2021, Zhang and Murdoch demonstrated that the contrast between a virtual image on an OST display and the physical environment behind it can be used to predict how transparent users will perceive the imagery to be [166]. They present a model that accurately predicts perceived transparency as a function of contrast, which is important to consider when designing applications for OST displays since perceived transparency has been shown to have other non-intuitive effects, such as affecting perceived humanness of virtual humans, and affecting perception of ground contact [1, 131].

2.1.2 Contrast Polarity

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 [122]) 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 [8, 18, 25, 135, 137, 161]. 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 [151]. This was also later confirmed to be the case by Piepenbrock et al. [136]. 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 [19], 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 [119]. 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 [21]. 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 [77].

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) [15]. 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 [22]. 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 Mojave¹.

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 [41, 99, 102, 104, 168]. Even in the domain of AR research, researchers investigated the effects of real background patterns and focal distance on visual fatigue and acuity [61, 121]. We see parallels between in-car heads-up displays and current-state OST-HMDs in the use of additive display designs, and the overall desire not only to ensure

¹Apple, “How to Use Dark Mode on your Mac” (<https://support.apple.com/en-us/HT208976>).

legibility of the displayed text but also to retain natural viewing of the physical environment behind the display without inducing severe visual fatigue.

2.2 Visual Attention Cues

Visual cues are commonly used to either *direct* or *attract* the attention of an observer to something specific in their environment. The effectiveness of such cues is referred to as *saliency*, the ability of a visual cue to stand out from its surroundings. Saliency is affected by many different factors, which can be separated out into two main groups: features of the cue itself, and features in the environment near or surrounding the cue. Kamkar et al. provides an insightful overview of how these factors affect the saliency of visual cues [92].

Within these two groups are a multitude of different features which affect the saliency of visual cues in different manners. In general, a cue can be made more salient by changing its appearance to differ from the environment's appearance in one or multiple ways [126]. Color, form, motion, and positioning changes have been shown in the past to be particularly effective at increasing cue saliency [162]. Increasing the number features that differ between a cue and its surroundings has also been shown to further increase the saliency of the cue, for example combining a luminance-based cue with size-based cue, or a color-based cue with a motion-based cue [82, 127].

The saliency of a visual cue can be measured in several different manners, such as through performance analysis of search tasks performed by participants, or through subjective measures. For search tasks, there are two main varieties: feature search and conjunction search. In feature search, sometimes referred to as a pop-out task, the participant is tasked with identifying an object within an arrangement of distractor objects that differs from the others in one or more certain distinct features [153]. Participant performance in such tasks has been shown to be unaffected by increas-

ing the number of distractor objects if the cue is successful at “popping-out” to the user [120]. In conjunction search, participants are similarly tasked with identifying an object with certain features, however in this task the distractor objects share a subset of the features with the object the participant is searching for [145, 153]. Participant performance in this type of task has been shown to be significantly reliant on the number of distractor objects presented, implying that participants must perform a serial search of all objects to find the one with the specified features [153].

2.2.1 Dichoptic Visual Cues

Dichoptic visual cues are cues that appear differently between each of the observer’s eyes, creating a phenomenon known as binocular rivalry [14]. This difference between eyes can be in the form of hue, lightness, size, or positioning. Wolfe and Franzel studied these types of cues in a series of experiments in 1988 [163]. They specifically investigated form rivalry, color rivalry, and binocular luster, where binocular luster is achieved by having a disparity in the perceived lightness of the visual cue such that one eye sees a darker cue while the other sees a lighter cue. Of these techniques, only binocular luster was found to successfully pop-out to study participants, however they concluded that binocular rivalry may be an effective manner of guiding the attention of an observer.

Dichoptic color rivalry, sometimes referred to as “forbidden colors,” were investigated by several other researchers since 1983 [12, 28, 81]. Such colors were thought to form when opposing colors, such as red and green, or blue and yellow, blended together on one or both of the observer’s eyes, resulting in a color that appears as both of the input colors rather than a blending of the two [12, 28]. However in 2006, Hsieh and Tse [81] suggested that the term “forbidden” is misleading and that the combined color is actually a blending of the two colors. Whether it is a blending or not, such color combinations between the observer’s eyes are not commonly seen, and thus may be an

effective method of drawing the attention of an observer, even if they are not necessarily processed preattentively, as shown by Wolfe and Franzel [163].

More general dichoptic cues were revisited by Zou et al. in 2017, where they confirmed that such cues may be effective at guiding the attention of an observer, but they achieved results that suggested that the luster effect did not pop out as strongly as originally thought in the 1988 experiments by Wolfe and Franzel [169].

More recently, Krekhov and Krüger investigated the potential of using monoscopic variations of visual cues in order to attract the attention of the observer in a technique they called DeadEye [101]. Their technique involved removing the image of the visual cue in one of the observer's eyes to cause binocular rivalry, and when tested in a pop-out task with varying numbers of distractor objects, they demonstrated that this type of cue successfully pops-out to observers in both feature search tasks with homogeneous or heterogeneous distractors, and thus can be processed preattentively by the observer. Their initial work was performed on a stereoscopic 3D flat-panel display, but was later replicated on an immersive VR display, where similar results were achieved [100].

It is also important to consider how the effectiveness of dichoptic cues varies in response to factors besides its appearance, such as the eye dominance of the observer and the appearance of the background behind or surrounding the visual cues.

Eye dominance was investigated in 2006 by Shneur et al. where they concluded that the dominant eye has priority in visual processing tasks, which they demonstrated via increased task performance in a pop-out style search task [146]. They later showed that this increase in performance for the dominant eye extends to conjunction searches as well [147]. More recently, the impact of eye dominance on user performance in using monocular displays was investigated by Bayle et al. where they found that only four of their 18 participants exhibited better tracking performance of a monocular visual cue with one eye compared to the other [9]. However, in the investigation of

DeadEye cues by Krekhov and Krüger, they did not find any significant effect of eye dominance on user performance [101]. In a study by Browne et al. involving user performance in using a monocular display in a flight simulator environment, eye dominance was again shown to not have a significant effect on user performance [17]. It appears as though there has yet to be a consensus on whether or not eye dominance is an important factor in determining user performance in applications involving monocular cues, which are similar to the conditions involved when identifying dichoptic visual cues.

It is also possible that the appearance of the background behind and surrounding the dichoptic cues could influence the observer's ability to distinguish the cue. Although little work has been done to investigate this specifically for dichoptic cues, there has been several studies that have investigated similar effects that the background appearance can have on user performance when using monocular displays. This was investigated by Grudin in 2002, where he found that dynamic moving backgrounds were detrimental to user performance in using a monocular display to perform a look-up task [72]. He also showed that a visually complex static background had similar, albeit lesser, effects on user performance, and concludes that monocular displays may not be well suited for dynamic or complex backgrounds. However, this was also evaluated in 2021 by Bayle et al. where it was found that high spatial frequency physical backgrounds led to increased user performance in a tracking task involving a monocular AR display compared to conditions with a low spatial frequency background [9].

2.2.2 *Virtual Humans*

AR displays enable users to observe and interact with environments consisting of real physical objects within the user's local environment and virtual imagery that appears to be collocated within the same environment. When virtual imagery is used to represent a real person or a virtual agent, it

is referred to as a *virtual human*. Typically the appearance and behavior of the virtual human can be controlled by a person or a virtual agent in order to accomplish tasks such as navigating virtual environments, communicating with other users, and interacting with physical or virtual objects.

One common use of virtual humans in AR is to allow remote communication and collaboration between people when it is not possible or is otherwise inconvenient to meet in person. While this domain has been more extensively explored with combinations of AR and virtual reality (VR) displays [26, 71, 96, 138], several recent works have explored this context solely using AR displays [155, 159, 164]. For instance, the AR telepresence project Holoportation reconstructed a user's appearance in real time and presented it to other remote AR users [129].

Virtual avatars are also sometimes used to change a user's self perception of their own body or identity through virtual embodiment illusions [70,93]. A recent survey by Genay et al. investigated existing work in this domain, paying particular attention to how the user's sense of embodiment changes when applying various levels of avatarization to oneself in AR [68]. In their work, they describe a continuum of self-body avatarization with AR displays, ranging from using the person's real body (in collocated settings), to accessorization of the user's body with virtual imagery, to partial avatarization with additional virtual modifications, to full avatarization where the user appears as a purely virtual entity.

When not representing a real human, virtual humans are sometimes used to visually represent a virtual agent. A systematic review of this particular domain was conducted by Norouzi et al. [125], which indicated that the recent literature in AR focused on four main application areas of virtual agents: assistive/collaborative tasks, entertainment/media, healthcare, and training. A common theme in this domain is that virtual agents are often used as a stand in for a real person or a human-controlled avatar in contexts where it is expensive, dangerous, or otherwise difficult to have a real person.

CHAPTER 3: PERCEPTION OF USER INTERFACES

In this chapter, a series of two user studies are described that investigate user perception of user interfaces (UIs) shown on optical see-through augmented reality (OST AR) displays and video see-through (VST) AR displays. Portions of chapter 3 have been previously published in the following publications:

- [94] Kangsoo Kim, Austin Erickson, Alexis Lambert, Gerd Bruder, and Greg Welch. Effects of Dark Mode Visualization on Visual Fatigue and Acuity in Optical See-Through Head-Mounted Displays. In Proceedings of the 2019 ACM Symposium on Spatial User Interaction (SUI), pages 1–9, 2019.
- [41] Austin Erickson, Kangsoo Kim, Gerd Bruder, and Greg Welch. Effects of Dark Mode Graphics on Visual Acuity and Fatigue with Virtual Reality Head-Mounted Displays. In Proceedings of IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pages 434–442, 2020.
- [45] Austin Erickson, Kangsoo Kim, Gerd Bruder, and Greg Welch. Dark/Light Mode Adaptation for Graphical User Interfaces on Near-Eye Displays. In ICAT-EGVE 2020 – International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments – Posters and Demos. The Eurographics Association, pages 1-2, 2020.
- [48] Austin Erickson, Kangsoo Kim, Alexis Lambert, Gerd Bruder, Michael Browne, and Greg Welch. An Extended Analysis on the Benefits of Dark Mode User Interfaces in Optical See-Through Head-Mounted Displays. ACM Transactions on Applied Perception, 18(3), pages 1-22, 2021.

The research presented in this chapter is used to partially support thesis statement 1:

- **TS1 - Identify Problems:** In increasingly common viewing conditions, AR users are susceptible to misperceiving or completely missing information shown on their AR display. Sometimes the user may not even be aware that their AR experience or their performance is being affected.

Specifically, this chapter demonstrates that text is less legible and UIs are less usable when the user's lighting conditions introduce a discrepancy between the perceived appearance of the UI and the intended appearance of the UI. This discrepancy is demonstrated through comparison of similar studies run on an OST AR display and subsequently on a VR or VST AR display. To support this thesis statement, the terms *perceived appearance* and *intended appearance* both need to be defined. These terms will share the same definitions across chapters 3 through 5.

Here, I use the term *intended appearance* to refer to the appearance of the AR imagery as chosen by its creator/designer. I use the term *perceived appearance* to refer to the appearance of AR imagery as perceived by the user of an OST AR display. In this manner, the *intended appearance* of the imagery is the “correct” appearance, while the *perceived appearance* of the imagery is different to an extent determined by the user's viewing conditions. As discussed in chapter 2, imagery on OST AR displays is prone to effects such as color blending and reduced contrast, where the extent of these effects is dependent on the characteristics of the user's physical environment, the characteristics of their display, and the characteristics of the particular imagery being displayed.

3.1 Overview

The studies presented in this chapter investigate several factors that may affect the user's perception of imagery shown on AR displays, including the appearance of the user's environment, the lighting

conditions in the user’s environment, and the contrast polarity of the UI. I demonstrate that a subset of these factors have significant effects on the user’s perception of AR imagery. Additionally, I demonstrate that for OST AR displays, negative polarity UIs (dark mode style UIs) are more robust to changing environment conditions compared to positive polarity UIs (light mode style UIs), resulting in increased text legibility, reduced visual fatigue, and increased usability of the UI in all tested conditions. For VST AR displays, I demonstrate that the lighting conditions in the user’s environment determine which UI style will provide the most benefits to the user, where in bright lighting conditions positive polarity (light mode style) UIs offer increased text legibility and increased usability of the UI, while in dim lighting conditions these benefits are found when using negative polarity (dark mode style) UIs. Additionally, I demonstrate that negative polarity (dark mode style) UIs offer reduced visual fatigue compared to light mode style UIs in all tested conditions.

3.2 Experiment I

In this section we describe the first of two user studies presented in the chapter. In this study, participants were asked to read text displayed on an OST AR display, the Microsoft HoloLens. This text was shown in two different *vision modes*, which differed in contrast polarity. Positive polarity UIs consisted of black text shown on a white virtual background, while negative polarity UIs consisted of white text on a black virtual background. Throughout the remainder of this section, these vision modes will be referred to as light mode when positive polarity UIs are used and dark mode when negative polarity UIs are used. During the study participants were asked to complete visual acuity tests, to rate their subjective experience, and to rate their preference of these different UIs.

3.2.1 *Participants I*

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) [142], 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.

3.2.2 *Materials I*

3.2.2.1 *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 3.1 and figure 3.2). 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

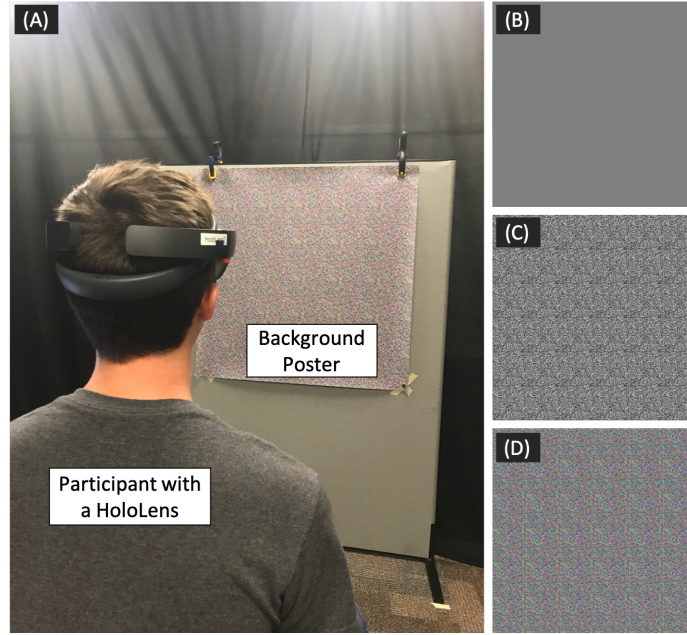


Figure 3.1: Experimental setup and the images for the background posters used in chapter three experiment one.

1268 \times 720 pixels per eye. The HoloLens 1 leverages SLAM-based tracking [20] 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. 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 [106], 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 Questions¹ in a 2×2 grid text board in a size of 72×72 centimeters (figure 3.2). We further developed an AR version of a common visual acuity test chart similar to a Golovin–Sivtsev Table with Landolt C characters [57, 58, 160], which are characterized by circles with a missing piece on either of four sides (figure 3.3). 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 0.9×0.9 meters. 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 [80]. 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 1268×720 pixels per eye, and a reported

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

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 UIs 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 [42]. 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 3.1.

Table 3.1: 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.

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

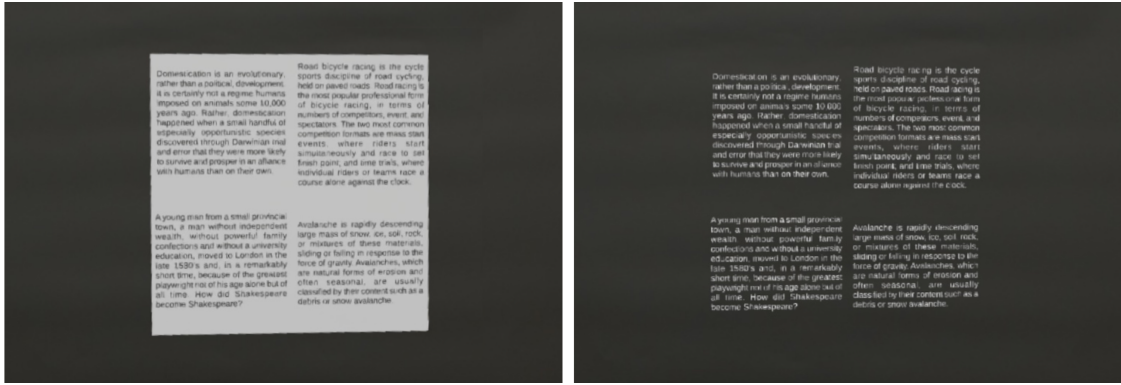


Figure 3.2: Illustrations showing the AR text-reading task participants had to read during experiment one.

3.2.2.2 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 3.1 A). We created different backgrounds for the experiment by mounting large-scale printed posters on a partition wall in front of the participants (figure 3.1 B–D). The posters were made of 0.9×0.9 meter Premium Archival Matte papers. The three considered background conditions were as follows (figures 3.1 and 3.3):

- **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.
- **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.

3.2.2.3 *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 3.3):

- **High Light:** The environment was well-lit due to diffuse indirect ceiling lighting in the room (with 200–270 lux²).
- **Low Light:** The environment was reasonably dark due to dimmed lighting (with 10–12 lux).

3.2.3 *Methods I*

We used a full-factorial within-subjects design in this experiment. As described in Section 3.2.2, the independent variables were as follows:

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

²Measured by a URCERI Light Meter Digital Illuminance Meter

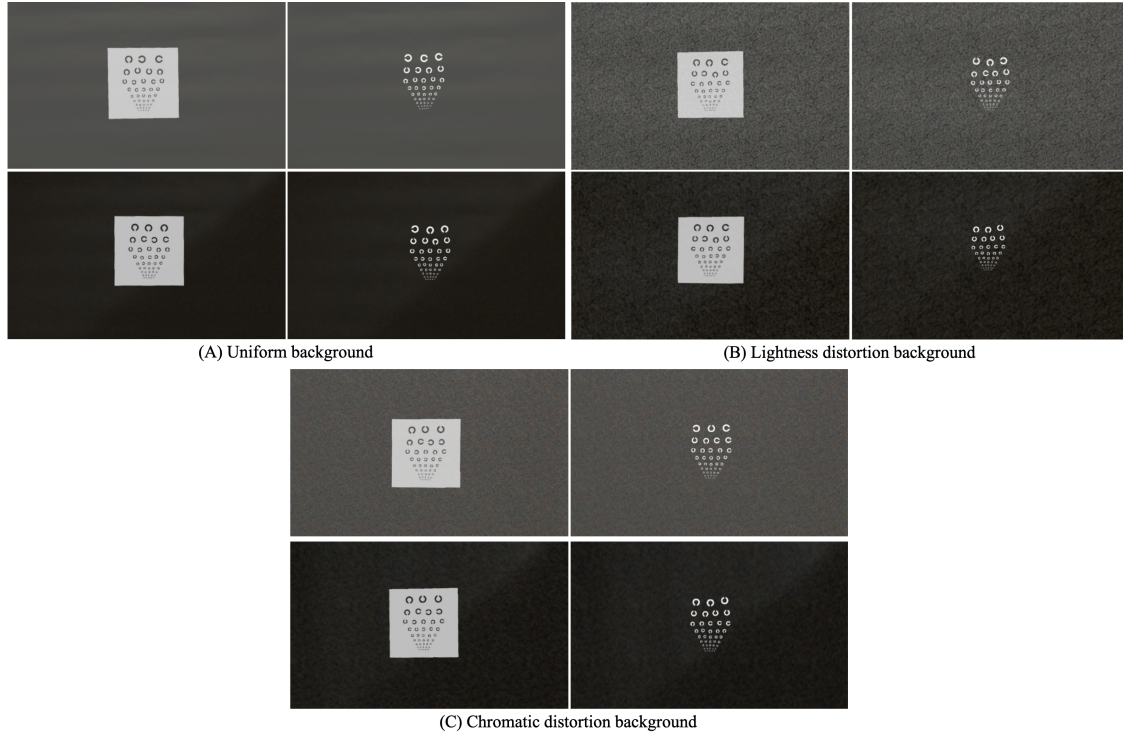


Figure 3.3: Illustrations of the different experimental conditions of experiment one.

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.

3.2.3.1 Procedure I

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 3.1). 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 3.2). 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 3.3). 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.

3.2.3.2 Measures I

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 3.2.2.1, we used a visual acuity test based on a Golovin–Sivtsev Table with Landolt C characters [57, 58, 160]. 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 [63, 67], 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) [143]. 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.

3.2.4 Hypotheses I

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.

3.3 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 [97, 103]. 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.

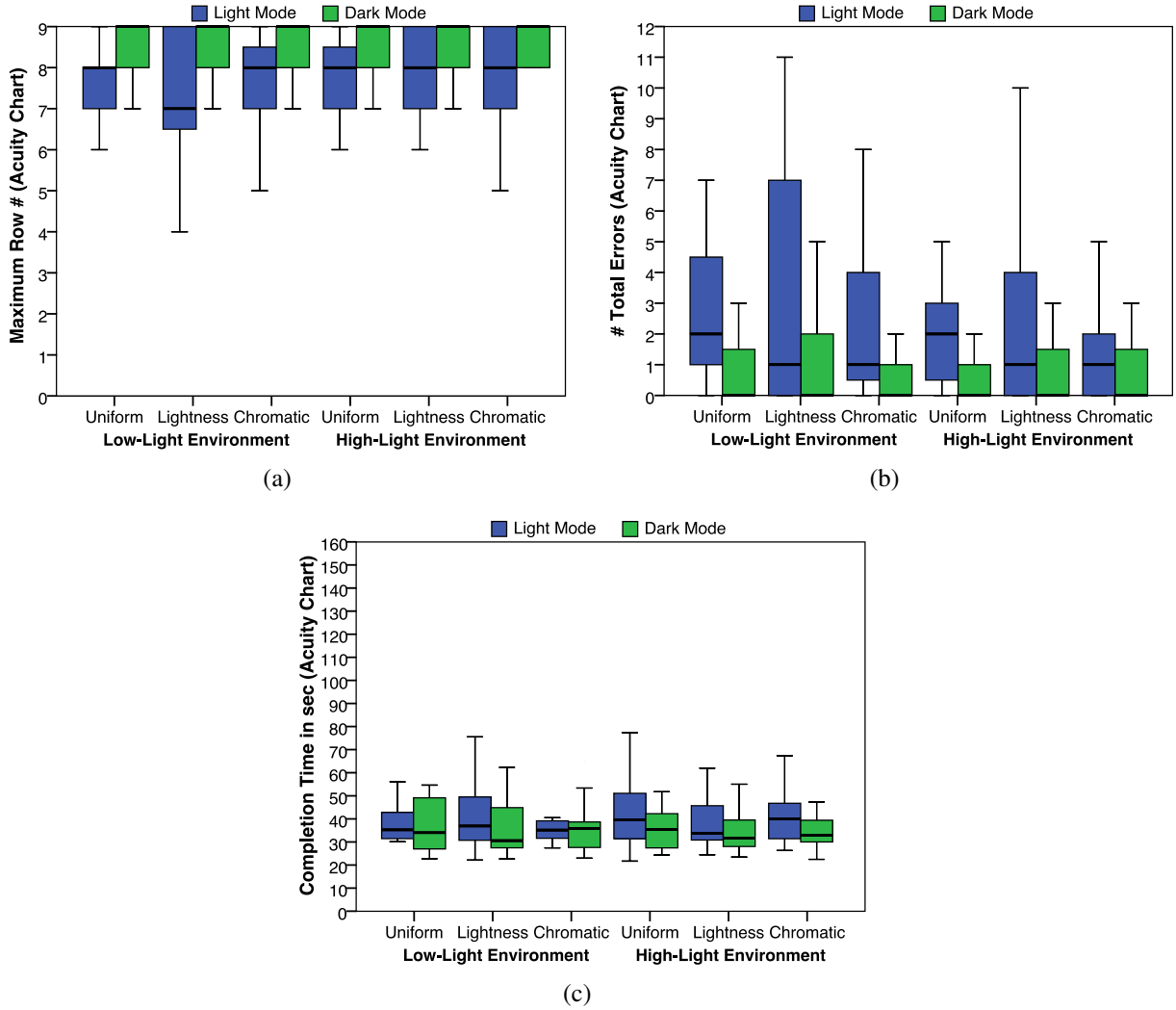


Figure 3.4: Results for the *visual acuity* tests of chapter 3 experiment 1.

3.3.1 Visual Acuity

The results for the visual acuity test are shown in figure 3.4.

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, $F(1, 18) = 9.20$, $p = \mathbf{0.007}$, $\eta_p^2 = 0.338$, indicating that participants

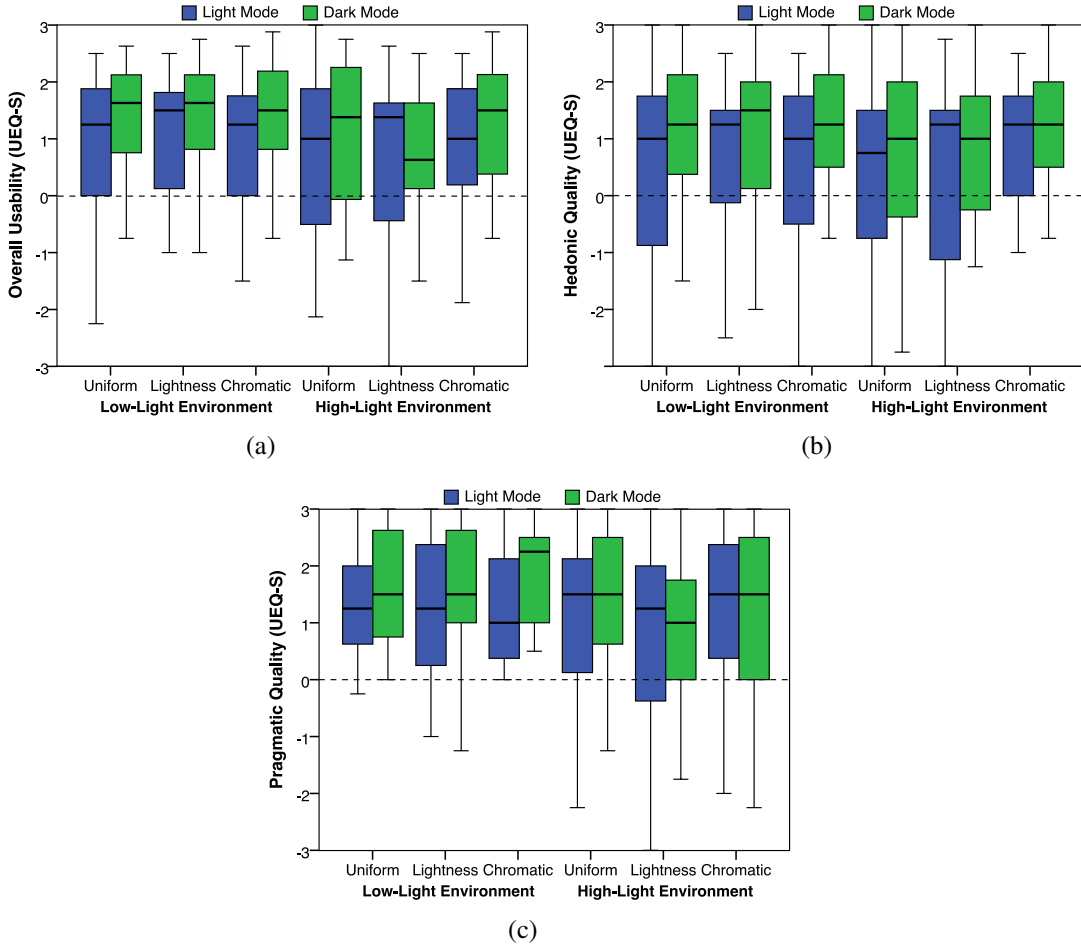


Figure 3.5: Results for the *usability* estimates of chapter 3 experiment 1 using the UEQ-S questionnaire.

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, $F(1, 18) = 12.65$, $p = 0.002$, $\eta_p^2 = 0.413$, 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 acu-

ity 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.

3.3.2 Usability

The results for usability (UEQ-S) are shown in figure 3.5. 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, $p = \mathbf{0.041}$, $\eta_p^2 = 0.223$, and on *hedonic quality*, $F(1, 17) = 6.97$, $p = \mathbf{0.017}$, $\eta_p^2 = 0.291$. 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 $p = 0.082$.

3.3.3 Subjective Preferences

The subjective preferences of our participants are shown in figure 3.6.

We found significant main effects for **physical lighting** on *overall preference*, $F(1, 18) = 6.09$, $p = \mathbf{0.024}$, $\eta_p^2 = 0.253$, on *visual comfort*, $F(1, 18) = 7.44$, $p = \mathbf{0.014}$, $\eta_p^2 = 0.292$, on *easy to read*, $F(1, 18) = 9.21$, $p = \mathbf{0.007}$, $\eta_p^2 = 0.338$, and on *perceived performance*, $F(1, 18) = 19.36$, $p < \mathbf{0.001}$, $\eta_p^2 = 0.518$. 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*, $F(1.51, 27.17) = 5.78$, $p = \mathbf{0.013}$, $\eta_p^2 = 0.243$. 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*.

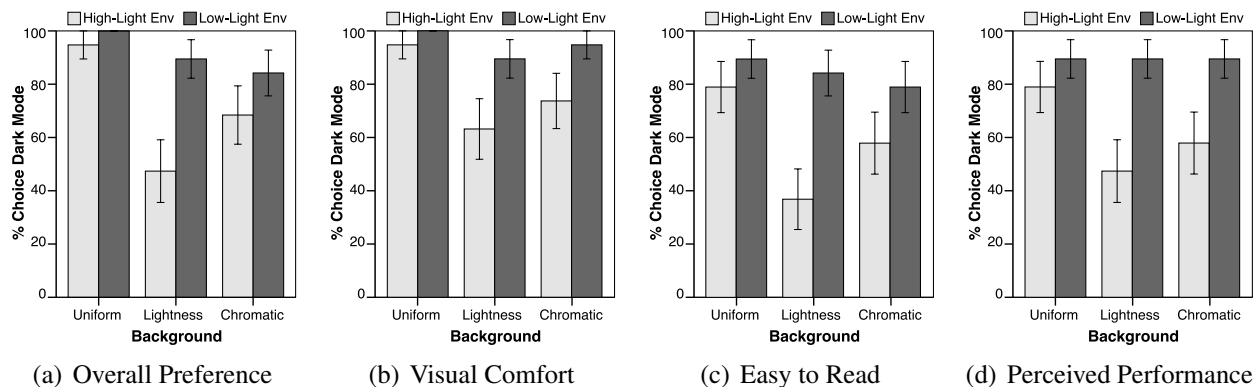


Figure 3.6: A figure demonstrating the results of the subjective preferences survey of chapter 3 experiment 1.

3.4 Discussion I

In this section, we summarize the main findings of experiment one and discuss implications for the use of UIs with OST AR displays.

3.4.1 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 [148]. As our stimuli consisted solely of black and white colored UIs, 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 relationship between environment luminance and pupil diameter.

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 [43]. 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 [30, 167]. However, there were some works that directly contradict those that were obtained here [55]. 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.

3.4.2 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 3.6, 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.

3.4.3 *Limitations*

It is possible 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 3.2.1, 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 UIs, 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.

3.5 Experiment II

In this section we present a user study in which we evaluate similar factors from the previous two studies differ when the user is tasked with interacting with study stimuli on a commercial virtual

reality (VR) display as opposed to the OST AR display used in the previous two experiments. Many current VR displays, such as the Meta Quest line of displays, have features that allow the user to use the display as a VST AR display. When using the display in this manner, the user sees their physical surroundings through imagery streamed from outward-facing cameras mounted on the front of the display. Hence, the results of this study can potentially apply to both similar VR displays as well as similar VST AR displays. This particular study investigates an additional independent variable that intends to manipulate the user's perceptual state to increase legibility of text by illuminating a ring of light around the outer perimeter of the display. In addition to this new variable, contrast polarity, and virtual lighting were manipulated to investigate how they influence the legibility of text on the user interface, visual fatigue, and the subjective preferences of the user.

3.5.1 Participants II

After initial pilot tests, we estimated the effect size of the expected strong effects, and based on a power analysis, we made the decision to recruit 18 participants, which proved sufficient to show significant effects in our experiment. We recruited a total of 15 male and 3 female participants (ages between 19 and 35, $M = 24.5$, $SD = 4.8$). Eligible for participation in the experiment were only healthy people who did not have any cognitive or motor impairments. All of our participants had normal or corrected-to-normal vision. Seven wore glasses and four wore contact lenses during the experiment. None of the participants reported known visual or vestibular disorders, such as color or night blindness, dyschromatopsia, or a displacement of balance. The participants were student or non-student members of the local university community, who responded to open calls for participation, and received monetary compensation for their participation. All participants had used a VR HMD before.

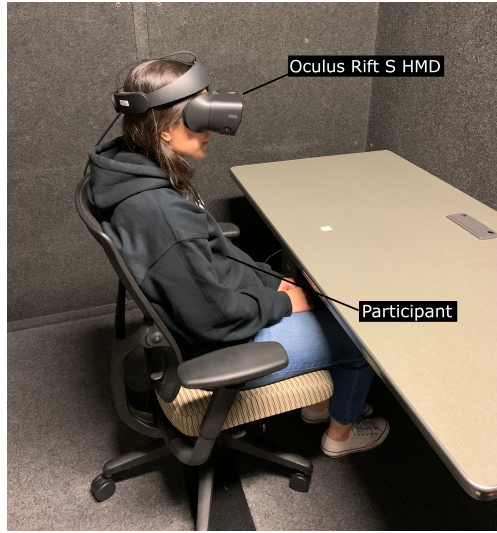


Figure 3.7: Annotated photo showing a participant in the study with the HMD seated at the table.

3.5.2 *Material II*

In this section, we describe the material used for our experiment.

3.5.2.1 *Physical Setup*

Figure 3.7 shows a photo of a participant in the study. Participants were seated in an office chair and were instructed to wear an Oculus Rift S VR HMD. The HMD was tracked in position and orientation using a the built in inside-out tracking, where position and orientation updates are handled internally by the HMD through the use of cameras placed on the device. The HMD has a resolution of 1280×1440 pixels per eye for a total resolution of 2560×1440 and a refresh rate of 80 Hz. The virtual environment was rendered in Unity 2018.2.21f1 on a host PC tethered to the HMD (Intel Core i7-8700k @ 3.70 GHz, 32Gb Ram, NVIDIA GTX 1070Ti graphics card, Windows 10 Pro).

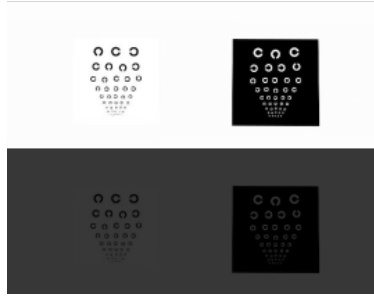


Figure 3.8: Illustrations of visual stimuli used in chapter 3 experiment 3.

3.5.2.2 Virtual Environment

The visual stimuli consisted of a virtual room in which the participant was placed near a wall facing a floating panel that contained text relevant to the conditions being displayed to them. Figure 3.8 shows the virtual content that we used for the tasks in the study.

The *floating panel* was designed to match realistic lighting conditions such that a diffuse white background would appear white in bright virtual lighting, but would appear gray in dim virtual lighting. It also meant that a diffuse black background would remain black independent of the amount of virtual light. The virtual lighting in the experiment could be adjusted to *bright* or *dim* by varying the intensity of a virtual point light located above the participant's head in the Unity scene. Bright in this case means that white pixels on the display were drawn at an RGB value of (1, 1, 1), and dim means that environment lighting was reduced by 90% so white pixels on the display appeared at an RGB value of (0.1, 0.1, 0.1).

It is important to note that the RGB values described above only describe the colors specified to the unity engine and are not indicative of the actual amount of light that was displayed from the Oculus S HMD. The apparent contrast between black and white pixels on a display will vary across different HMDs and displays depending on the contrast ratio of the device as well as parameters

associated with the display hardware.

Separate from the amount of ambient light in the virtual environment, we also implemented a *perimeter lighting mode* in which bright light originated from a ring of the out-most pixels at the perimeter of the HMD screen. This lighting mode was inspired by previous work by Jones et al. [90], who found that light reaching a VR user's eyes from the far periphery can affect (improve) spatial perception. The ring was scaled to take up 348 pixels or 13.6 percent of the total width of the display and 338 pixels or 23.4 percent of the total height of the display. These values were chosen based on pilot testing which suggested that lesser values were not noticeable to some users.

3.5.2.3 Task Stimuli

We implemented a *visual acuity test* that incorporated tumbling Landolt C characters, which could be oriented normally or at 90 degrees incremental rotations, so that the opening on the 'C' character could face up, down, left, or right [31]. These characters were randomly generated each time a chart was displayed, and by using this tumbling 'C' format, all acuity tests were of comparable difficulty to one another. The Landolt C characters were chosen in favor of a traditional Snellen variety acuity chart in order to avoid the possibility of having differing degrees of difficulty between same sized letters between participants. While a Snellen variety chart would have been possible, it would have required careful design to ensure that the charts used in each condition were similar in difficulty to each other to avoid potentially biasing a subset of conditions. The acuity chart was positioned at a distance of 1.52 meters (5 feet) away from the user, and at the eye level of the user. This distance was chosen because the letter sizes on our custom acuity chart were modeled after an acuity chart that was designed to be read specifically from this distance.

We further implemented a *reading task*, which consisted of four paragraphs from the Pearson Test of English Read Aloud Practice Questions, which were presented in the Liberation Sans font. This

task was chosen to allow a standard time for the user to be exposed to the condition lighting and potentially induce visual fatigue prior to reading the acuity chart. These paragraphs were displayed to the participant during each condition at the same depth as the acuity chart (1.52 meters) and at a consistent field of view between all conditions as a means of evaluating the amount of eye strain induced and the readability of text in each condition.

3.5.3 *Methods II*

The study used a $2 \times 2 \times 2$ full-factorial within subjects design in which each participant experienced all eight of the different conditions, and the conditions were counterbalanced among participants through the use of a Latin square. The evaluated independent variables were:

- **Color Mode:** *light mode* graphics consisting of black text on a white background or *dark mode* graphics consisting of white text on a black background.
- **Virtual Lighting:** *bright* or *dim* ambient lighting in the virtual environment.
- **Perimeter Lighting:** *enabled* or *disabled* perimeter lighting.

3.5.3.1 *Procedure II*

To begin the study, participants were led into the laboratory and were asked to read over a consent form describing what would take place during the experiment. After giving consent, the participants were asked to complete two questionnaires: one which gathered demographic information, and an Ocular Surface Disease Index (OSDI) survey that gathered information about the current level of comfort of the participant's eyes [142].



Figure 3.9: Illustrations depicting the VR text-reading task participants had to perform during chapter 3 experiment 3.

For each of the eight conditions, participants were then asked to don the HMD and observe a virtual panel consisting of four short paragraphs of text which they would read non-verbally to themselves (see figure 3.9). After one minute of observation, the paragraphs would disappear and be replaced with a visual acuity chart consisting of Landolt C characters rotated at 90 degree increments [31] (see figure 3.8). The participants would then read through the chart until they reached the bottom or the characters were too difficult for them to distinguish. We measured the number of errors

participants made when reading the letters, the maximum row that participants could read without errors, as well as the time it took them to complete the task. Following the completion of the acuity test, the participant would take off the HMD and complete two short questionnaires: a Short User Experience Questionnaire (UEQ-S) that gathered information on the usability of the graphics interface under the testing conditions [143], and a Convergence Insufficiency Symptom Survey (CISS) questionnaire that gathered information on any eye strain noted by the participant during the condition [16]. Immediately after completing the questionnaires, participants were instructed to re-don the HMD and continue onto the next condition with no additional time to rest.

After completion of all eight conditions, participants were asked to don the HMD one final time to measure their subjective preference of the different conditions as well as which conditions they found to be easiest to read.

Specifically, we asked them to indicate their preference of *color mode* (dark mode or light mode) on four questions:

1. **Preference:** Which condition do you prefer?
2. **Comfort:** Which condition was more comfortable?
3. **Easy to read:** Which condition was easier to read?
4. **Performance:** Which condition do you think you performed better, e.g., fast and accurate reading?

We further asked them to indicate their preference of *perimeter lighting* being enabled or disabled.

3.5.4 Hypotheses II

Inspired by the body of literature on vision modes, most notably recent work by Kim et al. [95], we defined the following hypotheses.

- **H1a (Virtual Lighting Affects Visual Acuity):** Participants will make fewer errors on their visual acuity test with bright virtual lighting than dim virtual lighting.
- **H1b (Virtual Lighting Affects Eye Strain):** Participants will experience less eye strain in dim virtual lighting conditions than they will in bright virtual lighting conditions.
- **H2a (Color Mode Affects Visual Acuity):** Participants will make fewer errors on their visual acuity test in the dark mode condition than in the light mode condition.
- **H2b (Color Mode Affects Eye Strain):** Participants will experience less eye strain when experiencing the dark mode than the light mode.
- **H3a (Perimeter Lighting Affects Visual Acuity):** Participants will make fewer errors on their visual acuity test while experiencing the perimeter lighting than they will without perimeter lighting.
- **H3b (Perimeter Lighting Affects Eye Strain):** Participants will experience more eye strain in conditions with perimeter lighting than they will in conditions without it.
- **H4 (Subjective Preference of Color Mode):** Users will prefer the dark mode over the light mode.

3.6 Results II

In this section we present the analysis and results of our experiment. We used parametric statistical tests to analyze the responses in line with the ongoing discussion in the field of psychology indi-

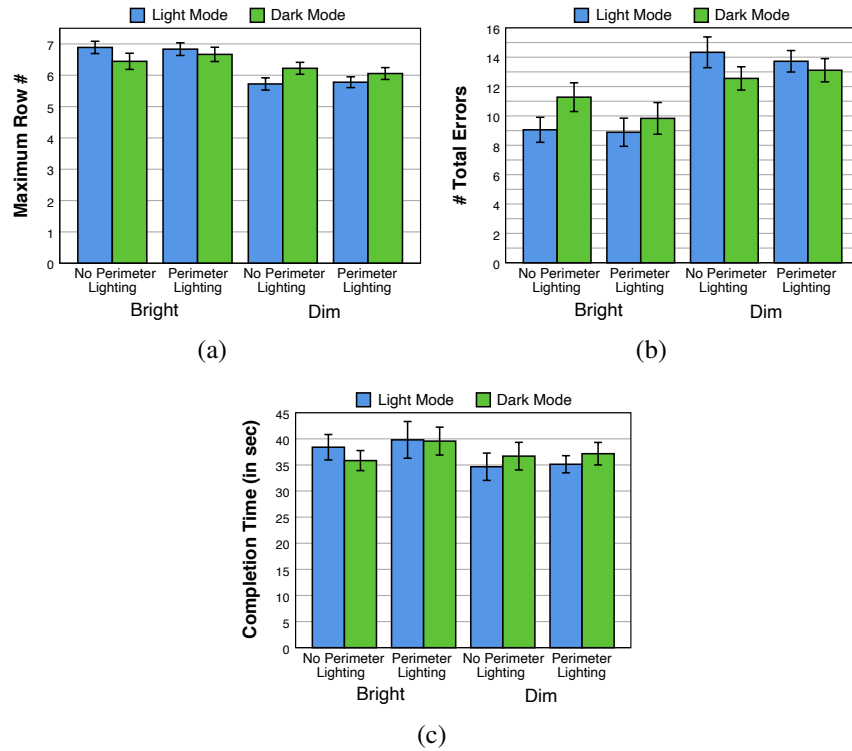


Figure 3.10: Results for the *visual acuity* tests: (a) maximum row on the acuity chart that could be completed without errors (between 0 and 9; higher is better), (b) total number of errors on acuity chart (lower is better), and (c) completion time for the acuity chart (lower is better).

cating that parametric statistics can be a valid and informative method for the analysis of combined experimental questionnaire scales with individual ordinal data points measured by questionnaires or coded behaviors [97, 103]. 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. We had to remove two questionnaire data sets from the analysis due to incomplete responses by our participants. We only report the significant effects.

3.6.1 Visual Acuity

Figure 3.10 shows the results for the visual acuity tests.

In line with hypothesis **H1a**, we found that users made fewer errors and could complete more rows without errors on the visual acuity charts if **virtual lighting** was bright instead of dim, while it took participants longer to complete the tasks. We found a significant main effect of virtual lighting on the number of errors made by participants on the visual acuity chart, $F(1, 17) = 72.33, p < 0.001$, $\eta_p^2 = 0.81$, indicating that bright lighting resulted in fewer errors than dim lighting. We also found a significant main effect of virtual lighting on the maximum row without errors on the visual acuity chart, $F(1, 17) = 32.12, p < 0.001$, $\eta_p^2 = 0.65$, which indicates that more rows could be completed with bright lighting than dim lighting. Further, we found a significant main effect of virtual lighting on the time spent on the visual acuity chart, $F(1, 17) = 4.54, p = 0.048$, $\eta_p^2 = 0.21$, indicating that it took participants longer to complete the charts under bright lighting than under dim lighting.

Contrary to hypothesis **H3a**, our results did not show any significant effects of the **perimeter lighting** on the participants' errors, maximum rows, or time when completing the visual acuity charts. As we are not seeing significant benefits of perimeter lighting, we are focusing on the results for color modes without perimeter lighting in the following. We further found a significant interaction effect between virtual lighting and color mode on the number of errors, $F(1, 17) = 13.99, p = 0.002$, $\eta_p^2 = 0.45$, and on the maximum row without errors, $F(1, 17) = 9.43, p = 0.007$, $\eta_p^2 = 0.36$, so we present the corresponding significant effects in the following. We found no significant effects on time.

In line with hypothesis **H2a**, without perimeter lighting, we found significant effects of the **color mode** on the results. However, interestingly, the results show the opposite effect depending on the virtual lighting:

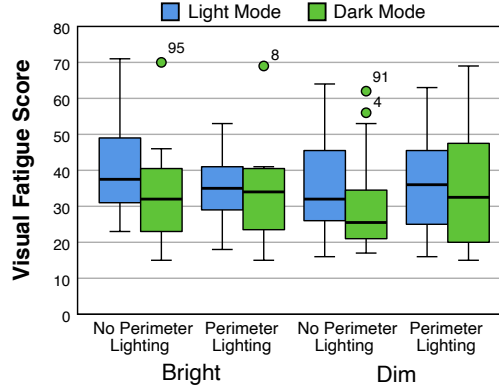


Figure 3.11: Results for the *visual fatigue* questionnaire (CISS): overall fatigue scores (lower is better).

For *bright* environments, we found a significant effect of the color mode on the number of errors participants made when reading the visual acuity chart, $F(1, 17) = 9.15$, $p = \mathbf{0.008}$, $\eta_p^2 = 0.35$, indicating fewer errors for the light mode over the dark mode. We also found a non-significant trend between the color mode and the maximum row that a participant could reach without making any errors, $F(1, 17) = 3.28$, $p = \mathbf{0.088}$, $\eta_p^2 = 0.16$, suggesting that more rows may be completed with the light mode than the dark mode.

For *dim* environments, we found a significant effect of the color mode on the number of errors participants made when reading the visual acuity chart, $F(1, 17) = 4.91$, $p = \mathbf{0.041}$, $\eta_p^2 = 0.22$, indicating fewer errors for the dark mode over the light mode. We also found a significant effect of the color mode on the maximum row that a participant could reach without making any errors, $F(1, 17) = 5.28$, $p = \mathbf{0.035}$, $\eta_p^2 = 0.24$, indicating that more rows could be completed with the dark mode than the light mode.

3.6.2 Visual Fatigue

Figure 3.11 shows the visual fatigue results for the CISS questionnaire.

In line with hypothesis **H2b**, we found a significant main effect of **color mode** on the visual fatigue scores, $F(1, 15) = 8.10$, $p = 0.014$, $\eta_p^2 = 0.34$, indicating lower visual fatigue for the dark mode compared to the light mode. Interestingly, this result is independent of the virtual lighting and applies to both bright and dim environments (see results for visual acuity).

Contrary to hypothesis **H3b**, our results did not show any significant effects of the **perimeter lighting** on the visual fatigue scores. As we are not seeing significant benefits of perimeter lighting, we are focusing on the results without perimeter lighting in the following.

In line with hypothesis **H1b**, without perimeter lighting, we found a significant effect of **virtual lighting** on the visual fatigue scores, $F(1, 15) = 5.17$, $p = 0.038$, $\eta_p^2 = 0.26$, indicating higher visual fatigue for the bright environment compared to the dim environment.

3.6.3 Usability

Figure 3.12 shows the usability results for the UEQ-S questionnaire, in which users rated various aspects of the condition using a seven point scale [143]. We found no significant effects of virtual lighting, perimeter lighting, or color mode on the usability results.

3.6.4 Subjective Preferences

Figure 3.13 shows the participants' preferences of the dark mode for all conditions in the experiment.

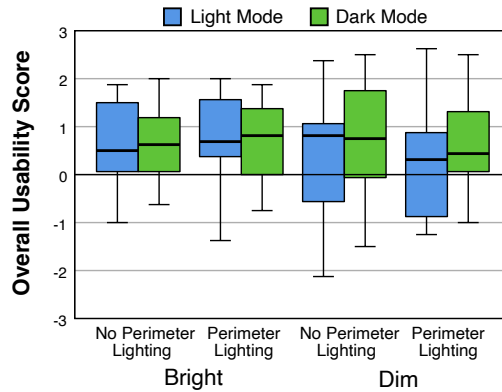


Figure 3.12: Results for the *usability* questionnaire (UEQ-S): overall usability scores (higher is better).

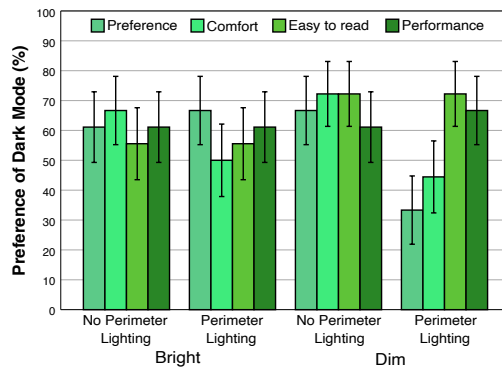


Figure 3.13: Subjective results in percent of participants who preferred the dark mode in the different experimental conditions.

We performed a two-tailed binomial test analysis on the subjective preference data with a test value of 0.5 and a confidence interval of 95%, where users responded to questions about their preference of color mode between either dark mode or light mode under each of the study conditions. Users were specifically asked which color mode they preferred, which was more comfortable, and which was easier to read. We found a non-significant trend in the number of participants who preferred the dark mode when trying to read in a dim virtual environment with perimeter lighting turned on ($p = 0.096$, $Proportion = 0.722$) and turned off ($p = 0.096$, $Proportion = 0.722$). We also found

a non-significant trend in the number of participants who preferred the dark mode as more visually comfortable in dim lighting conditions with perimeter lighting turned off ($p = 0.096$, $Proportion = 0.722$). We also found that a significant number of participants preferred having perimeter lighting turned off as opposed to turned on ($p = 0.008$, $Proportion = 0.833$).

3.7 Discussion II

In this section, we discuss the main findings and their implications for VR HMDs.

3.7.1 *Dark Mode Improves Visual Acuity Only in Dim Lighting Conditions*

Our results shown in Section 3.6.1 indicate that the *dark mode* improves the visual acuity of the user in dim lighting conditions on VR HMDs, effectively making it easier for users to identify Landolt C characters or make out small visual details. Conversely, the *light mode* improves the visual acuity of the user in bright virtual environments.

This result stands in partial contrast to the results of prior work by Kim et al., who investigated dark mode UIs in AR OST HMDs and found that the dark mode yielded better visual acuity regardless of lighting conditions [95]. It stands to reason that the difference in the display’s light model, in particular the *additive light model* [64] of current OST displays as well as the *screen door effect* that is prevalent in current immersive HMDs have a strong effect on the results.

In VR HMDs, the screen door effect is very prominent when pixels on the display are illuminated with bright light, and is more obscured from view when pixels are darker. However, even in the presence of an increased screen door effect in bright virtual environments, we found a significant main effect that the user’s visual acuity under bright lighting is significantly higher than under dim

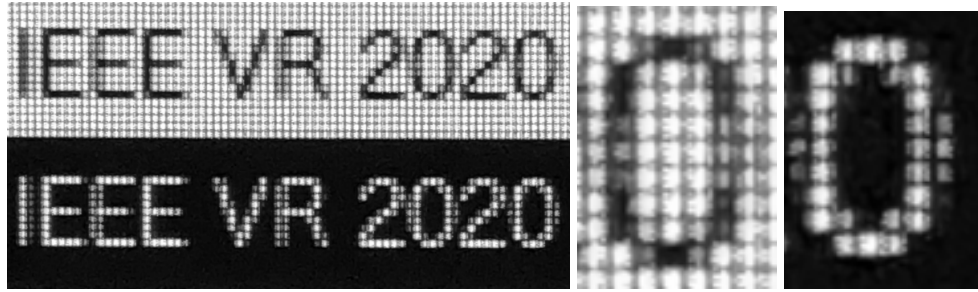


Figure 3.14: Illustration of the *light mode* (top) and *dark mode* (bottom) color schemes on virtual reality head-mounted displays. The dark grid indicates the *screen door effect* in VR HMDs, which affects the signal-to-noise ratio of the foreground light/dark text on a dark/light background.

lighting. This result matches previous work, which indicates that increasing the amount of light reaching the user’s eyes will reduce their *pupil diameter*, which in turn is known to increase the quality of the retinal image with greater depth of field and less spherical aberration [151]. It is interesting to see that the positive effects of the increased light out-weighted the negative effects of the increased screen door effect in VR HMDs.

As shown in figure 3.14, with the *dark mode*, the screen door effect is primarily hidden from view in the dark background but does appear directly over the foreground letters. In contrast, in the *light mode*, the screen door effect is clearly noticeable in the light background but is more obscured in the foreground letters. One’s first intuition may suggest that having the screen door effect over the letters and not the background would make them more difficult to see, but our results suggest that this only occurs in bright lighting conditions and that the opposite occurs in dim lighting conditions.

It is further interesting to note that the aforementioned study by Kim et al. incorporated a *lighting-independent text mode* (as used in AR heads-up displays) for the visual acuity charts that were displayed to the users, meaning that the light/dark RGB color values on the chart were constant and were not affected by virtual lighting in the AR environment [95]. In contrast, our study took

place in VR as opposed to AR, where the RGB color values of text in the virtual environment were affected by changes in the amount of virtual light, denoted as a *lighting-dependent text mode*. Because of this, when the virtual lighting is bright, white text appears as RGB value (1, 1, 1) and black text as (0, 0, 0). However, as the virtual lighting dims, the white text darkens to a value between (1, 1, 1) and (0, 0, 0), while black text remains black (0, 0, 0) and unaffected by the amount of virtual light. A decrease in virtual light thus reduces the visual contrast between the light colors and dark colors and the signal-to-noise ratio between the foreground text and its background, which is known to reduce the visual acuity [89].

Our results indicate that it is advantageous to use the *light mode* under bright virtual lighting, but when the contrast between letters and their background is reduced due to dim virtual lighting, then it is advantageous to switch to the *dark mode*. We believe that this is partially due to a *color bleeding effect* that occurs when a light colored letter is presented on a dark background, where the light from the letter partially illuminates neighboring background pixels and results in a letter that appears slightly larger [60]. It stands to reason that the magnitude of this effect is affected by virtual lighting paired with the nature of the letter identification task. In our study, participants were asked to identify Landolt C characters on the visual acuity chart, and if the magnitude of the color bleeding effect was too significant (in the case of bright virtual lighting) then it is possible that while the letters did appear slightly larger, the opening on the ‘C’ is reduced to appear more as an ‘O,’ and thus the direction of the opening is more difficult to distinguish. In the case of dim lighting conditions, the characters still appear slightly larger, but the magnitude of the color bleeding effect is not as strong as in the bright lighting condition, resulting in an opening on the ‘C’ that is easier to distinguish than for the light mode.

If this color bleeding effect is responsible for the results obtained here, then it is possible that different results may be obtained from a similar future study where the pixel density of the VR HMD is increased. This increased pixel density may result in less of a color bleeding effect around

the perimeter of the letters, which means that letters will appear slightly smaller on the high-density display and thus be more difficult to read. However, there is a trade-off to this reduced color bleeding effect as the openings in letters will be easier to identify than when this effect is more apparent, which should make letters with similar features such as ‘C,’ ‘E,’ and ‘O’ easier to distinguish from one another.

3.7.2 *Dark Mode Decreases Visual Fatigue*

As shown in Section 3.6.2, our results show that the *dark mode* resulted in significantly lower visual fatigue (CISS) scores than the *light mode*, which suggests that the dark mode causes less eye strain than the light mode. This result was also observed by Kim et al. for AR OST HMDs [95]. Further, in line with related work in the field, we also found that increasing the amount of (virtual) lighting caused more eye fatigue than our tested dim lighting condition [10]. For the least amount of visual fatigue in VR HMDs, our guideline is to dim the amount of virtual lighting and make use of the dark mode when presenting text or other visual details.

3.7.3 *Preference of Dark Mode over Light Mode*

As shown in Section 3.6.4, the majority of participants responded with a preference of the dark mode over the light mode, although this was only a non-significant statistical trend and further research would be required to come to a more general conclusion. Both color modes offer benefits to the users’ visual acuity depending on the virtual lighting of the scene. A slight shift in preference for the dark mode might stem from perceived benefits due to reduced visual fatigue. It is possible that the preferences would have become clearer in favor of the dark mode after a longer VR exposure.

3.7.4 Perimeter Lighting Showed no Significant Effects

As shown in Section 3.6.4, our results indicate that the majority of participants preferred the *perimeter lighting* to be turned *off*. We also found no significant effects of perimeter lighting on visual acuity, visual fatigue, or usability. We were surprised to not see clear benefits of the perimeter lighting on the results as the relevant literature suggested that a decrease in pupil size due to added light should improve the retinal image due to greater depth of field and reduced spherical aberration [151]. In theory, it should not matter whether the light that is affecting the user's pupil size originates in the center or the periphery/perimeter of the display.

Some of our participants commented on the perimeter lighting, e.g., stating that turning the perimeter light on *felt* like the rest of the virtual environment was getting darker. Another mentioned that they felt as though a dark gradient was placed over the center of the screen when the perimeter lighting was on.

For future work in this direction, we suggest looking into far-periphery lighting (instead of perimeter lighting) as used by Jones et al. [90] or Lubos et al. [117], who added an LED strip around the screen in the periphery of a VR HMD. We expect that an increased amount of peripheral light might result in benefits for visual acuity in VR, although we also see potential drawbacks due to increased visual fatigue.

3.8 Summary

The studies presented in this chapter investigated several factors that have significant effects on the user's perception of imagery shown on AR displays, including the lighting conditions in the user's environment, and the contrast polarity of the UI. When viewing imagery on any display, there is always a discrepancy between the intended and perceived appearance of the virtual imagery, as

the user's viewing conditions affect their perception of the imagery. For OST AR displays, this discrepancy is more apparent, as virtual imagery blends and takes on characteristics of the user's physical environment.

The results of study one support thesis statement one by demonstrating that the legibility of text and the usability of the UIs is significantly affected by the lighting conditions in the user's environment. The lighting conditions tested in this study are relatively dim (10 and 250 lux) compared to the range of environment lighting conditions an AR user could potentially experience, since for instance direct sunlight could yield lighting conditions greater than 100,000 lux. Yet despite the limited range tested, significant effects are still observed. Thus, there is likely a range or several ranges of environment lighting conditions in which the user may experience difficulty reading text shown on the display, and once lighting levels exceed a particular threshold the user will not be able to see the imagery on the display at all, rendering the display useless until conditions improve.

In study two, the lighting conditions of the study were manipulated in two different manners. When perimeter lighting was applied on the VST AR display, the amount of light reaching the user's eyes was increased in a manner similar to the increase in environment lighting between conditions in study one. However, we found that increasing lighting in this manner did not have significant effects on the user's perception of the UIs in study two, while it did for study one. Additionally, the lighting conditions in the virtual environment were manipulated in a manner that affected the perceived color of the UIs in the study, where in darker conditions white portions of the UI would instead appear to be grey. This particular lighting manipulation had significant effects on the legibility of text, where brighter virtual lighting conditions resulted in UIs that were more legible. In chapter 6, I demonstrate that both of these effects can be explained in terms of luminance contrast. I also demonstrate how the user's environment lighting conditions and the color of the UI being displayed can be considered to calculate contrast values used to predict the difficulty the user will have when observing and interacting with UIs on their AR display.

In the next chapter, I investigate how similar factors affect the user's ability to recognize the presence of visual attention cues shown on OST AR displays.

CHAPTER 4: PERCEPTION OF ATTENTION CUES

In chapter 3, I demonstrated how the lighting conditions in the user's environment affect their perception of UIs shown on OST AR displays. However, there are other factors that similarly affect the user's perception of AR imagery, such as the appearance of the user's environment and the color of the AR imagery being displayed. In this chapter, two user studies are described that investigate the effectiveness of visual attention cues presented on optical see-through (OST) augmented reality (AR) displays in many different colors and in environments with varying appearances.

Portions of this chapter were previously published in the following publication:

- [40] Austin Erickson, Gerd Bruder, and Greg Welch. Analysis of the saliency of color-based dichoptic cues in optical see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–16, 2022

The research presented in this chapter is used to partially support thesis statement one:

- **TS1 - Identify Problems:** In increasingly common viewing conditions, AR users are susceptible to misperceiving or completely missing information shown on their AR display. Sometimes the user may not even be aware that their AR experience or their performance is being affected.

In particular, this chapter demonstrates that visual attention cues are less effective when there is a discrepancy between the perceived appearance of the visual attention cue and the intended appearance of such a cue.

4.1 Overview

AR visual attention cues can be used to draw or direct the attention of the user to the specific regions of their physical environment or to specific virtual elements occupying the space around them. However, unless the AR system analyses the appearance of the user's physical environment, it may be difficult to choose or generate an attention cue that effectively stands out in the user's environment. For instance, an attention cue may not be very effective if the user is in an environment consisting of colors similar to the attention cue. For this reason, this chapter investigates the differences in effectiveness between several different dichoptic visual attention cues, some of which have been previously demonstrated to be effective at drawing the attention the user.

Three main types of dichoptic visual cues are investigated, which involve differences in the color of the cue shown between each of the observer's eyes. For example, one of the user's eyes may see a green cue while the other eye sees a red cue. As mentioned in section 2.2, certain dichoptic cues have been previously shown to be processed preattentively by the observer when displayed on a VR or VST AR display, such as the DeadEye style visual cue, where the cue is shown monoscopically to the user such that one eye does not see the visual cue while the other does [100, 101]. This particular cue is also investigated in the following experiments, which allows us to compare the results of our study to the previous work by Krekhov et al. and determine how factors specific to the additive nature of the OST display impact the effectiveness of visual cues.

4.2 Experiment 1

In this section, we describe an experiment that compares the saliency of dichoptic cues on an OST AR display.

4.2.1 Participants I

We recruited 20 participants ages 18–56 (mean 26.3, SD 9.2) 14 male, 6 female from the population of our university. The participants were screened for exclusion criteria, including pregnancy, history of seizures/epilepsy, neurological and motor impairments, color blindness, strong eye dominance, night blindness, and visual conditions that otherwise impair their visual acuity. Eight participants wore glasses during the experiment and all participants reported having normal visual acuity (with correction if needed).

Participants were asked to rate their level of experience with using stereoscopic displays, such as watching 3D movies or using AR/VR head-mounted displays, using a seven point scale with 1 meaning “least experienced” and 7 meaning “most experienced.” Participants reported a mean level of experience of 5.5 with a standard deviation of 1.7. Participants were also asked to rate their level of experience with using monocular displays using the same seven point scale, which resulted in a mean of 2.8 and standard deviation of 1.5.

4.2.2 Materials I

In order to present dichoptic imagery to the participant that differed in manners other than parallax effects, we set up a scene in the Unity engine (version 2019.4.26) in which the main camera rig consisted of separate cameras for each of the participant’s eyes. Objects in the unity scene were duplicated in place and set to a layer mask such that one object would appear solely to the participant’s left eye and one would appear solely to the participant’s right eye. In this manner, different materials could be set for each, otherwise identical, object in order to generate binocular rivalry.

We investigated three main types of visual cues involving binocular rivalry in this study, each

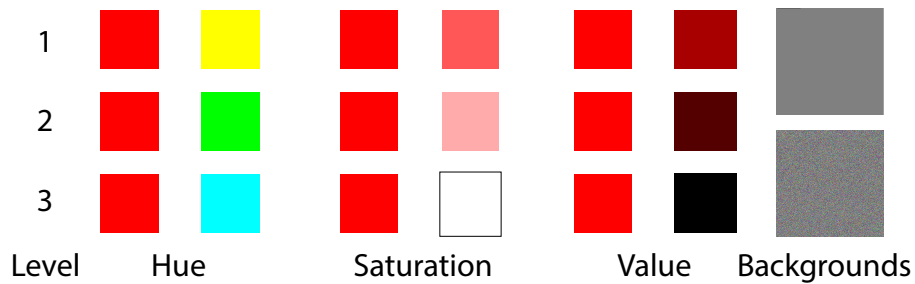


Figure 4.1: The figure depicts a side by side view of the visual cues used in experiment 1 and 2 separated into what each of the participant's eyes would see, as well as illustrations of the grey and randomized pixel patterns used to create the background posters. For this illustration, the original hue chosen was 0 (red) for the left eye, and is shifted to create a dichoptic cue by changing the color on the right eye. Note that the starting hue was randomized for each trial in the study procedure. Also note that, due to the additive light model employed by the AR display, the color depicted in the bottom right corner (for the Value cue at level 3) is black, and thus appears as transparent when shown on the OST AR display. Additionally, the appearance of these cues, as observed by the participant, will be shifted in color space as the light from the background poster blends with the light emitted from the AR display.

with three different levels of intensity. As the level of intensity increases, the imagery between the participant's eyes differs to a greater degree. As an initial exploration into the effectiveness of these types of visual cues, we investigated the three dimensions of the hue, saturation, value (HSV) color space. This color space is commonly visualized as a cylinder, where hue represents the radial angle, saturation represents the radial length, and value represents the vertical height.

- **Hue:** These cues appeared with a different hue between the participant's eyes (see figure 4.1 left.) The intensity level was varied by increasing the amount of degrees of rotation away from the randomly chosen initial hue of the object for one of the participant's eyes (see figure 4.2 left.) The three levels of intensity for this type of visual cue involved 60, 120, and 180 degrees of rotation respectively.
- **Saturation:** These cues appeared with a different saturation level between the participant's

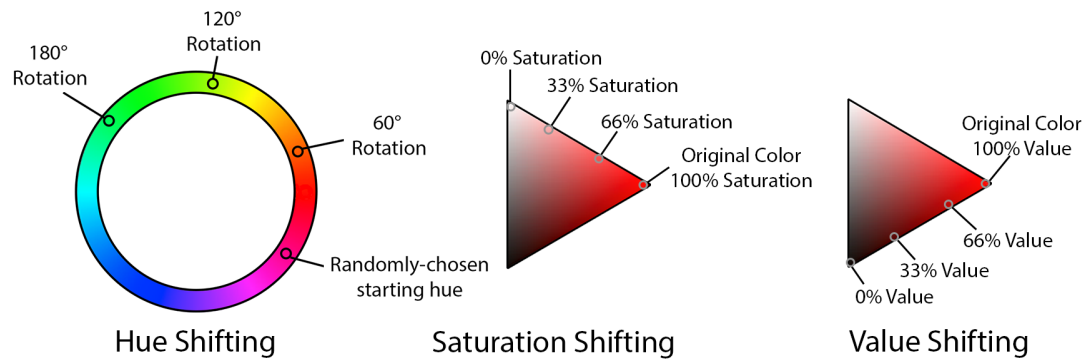


Figure 4.2: The figure depicts the color shifts used to generate our visual cues in the HSV color space.

eyes (see figure 4.1 middle), effectively making the image whiter in one eye compared to the other. The intensity level was varied by specifying a percentage of saturation at levels of 0%, 33%, and 66% respectively (see figure 4.2 middle).

- **Value:** These cues appeared with a different value level between the participant's eyes (see figure 4.1 right.) Since an additive light model OST HMD was used to display these cues, where darker colors appear more transparent than lighter colors, reducing the value level has a similar effect to reducing the transparency level of affected object. The intensity level was varied by specifying a percentage of value at levels of 0%, 33%, and 66% respectively (see figure 4.2 right.) At 0% value levels, the object is completely transparent on the display, making this particular level similar to the DeadEye cue described in section 4.1.

For each of these types of visual cues, the participant is presented with a different color for each eye, and the relative difference between these two colors can be measured as a distance between two points in the HSV color space. It is possible that this distance correlates with the saliency of the dichoptic cue, and is quantified in terms of the radius, r , and height, h , of the HSV color space. We designed the dichoptic cues in this study to investigate the whole range of the color space available

to the AR display. Therefore, the hue-based cues involve HSV color space distances of r , $r\sqrt{3}$, and $2r$ for intensity levels 1, 2, and 3 respectively. The saturation-based cues involve HSV color space distances of $\frac{r}{3}$, $\frac{2r}{3}$, and r for intensity levels 1, 2, and 3 respectively. Finally the value-based cues involve HSV color space distances of $\frac{h}{3}$, $\frac{2h}{3}$, and h for intensity levels 1, 2, and 3 respectively.

It should also be mentioned that due to the additive light model employed by the AR display, the colors of the visual cues observed by participants blend with the color of the physical environment behind it [64]. Since the two colors used in the dichoptic cue are blending towards a single background color, the distance between the two colors of the dichoptic cue will decrease after color blending. The only time this is not the case is if the background is pure black, in which case the distance between the colors will remain unchanged.

Following the example set by Krekhov et al. we arranged our visual cues within a four by four grid of distractor objects [100]. These distractor objects took the form of cubes that were sized to be 1.8 degrees of visual angle with horizontal and vertical spacing between the cubes of 1.2 degrees of visual angle, resulting in a grid that measured 10.8 by 10.8 degrees of visual angle. These values were somewhat reduced from those used by Krekhov et al. due to the limited vertical field of view of the HoloLens 2. For each trial during the experiment, the hue of the cubes in the grid was set to a randomly chosen number between 0 and 359, while saturation and value were set to 100%. Half of the trials experienced by participants involved a dichoptic cue being applied onto a randomly-chosen cube from the grid, shifting its hue, saturation, or value away from the initial color in one eye, while its appearance remains unchanged in the other eye. The eye chosen to present the shifted color to is determined by the **Presented Eye** independent variable described in section 4.2.3.1. The remaining half of trials (those without a dichoptic cue) are presented as a homogeneous grid of cubes.

We chose to use the Microsoft HoloLens 2 as the OST AR display for this study. The HoloLens 2

has a resolution of 2048×1080 per eye, and a field of view of 43×29 degrees¹. The brightness of the HoloLens was verified to be set to maximum for each participant, and the device was remotely connected to a PC controlling the sequencing of study events via the Unity engine and Holographic Remoting. As with other waveguide based OST AR displays, there are brightness non-uniformities when observing virtual content through the HoloLens 2 [106]. These can be observed monoscopically as variations in the appearance of virtual content within different parts of the field of view of the HoloLens 2, and they can also be observed stereoscopically since these uniformities appear to be in different regions of the field of view for each eye/display. Because of this, and similar to previous works [100], we ensured that the position of the visual cue within the arrangement of distractor objects was randomized for each trial the participant experienced. This randomization of cue position on the display should reduce the overall impact that these non-uniformities have on the results of the study.

We used two background posters during this study, a solid grey background that would introduce a uniform color blending between the AR imagery and the participants' view of their physical environment, and a "chromatic aberration" background that introduced a randomized color blending between the AR imagery and the participants' view of their physical environment. The posters were printed on premium archival matte paper and measured 0.9x0.9 meters. The solid grey background poster was generated using a uniform grey pixel intensity of 128 on a scale of 0 to 255. The chromatic aberration background poster was generated by randomly assigning pixel values across the range of available color space (randomizing between 0 and 255 for the red, green, and blue color channels independently). The pixels on the chromatic distortion background were sized at 1.58 millimeters, which corresponds to a visual angle of 0.0455 degrees (2.725 arc minutes) at two meters distance. When viewed under the LED lighting from the participant's perspective, the illuminance of the background posters were measured using an Urceri MT-912 light meter prior

¹<https://uploadvr.com/hololens-2-field-of-view/>

to starting participant sessions, where the average illuminance was found to be 115 Lux and the standard deviation was 7.65 Lux.

The testing environment was set up so that the participant was seated in a chair two meters in front of a wall, upon which the background posters would be hung. Two LED light sources were positioned just to the left and right sides of the participant, also at a distance of two meters away from the wall, and at a height of 1.83 meters (measured from the top of the light).

4.2.3 *Methods I*

4.2.3.1 *Study Design*

The experiment consisted of four independent variables:

- **Cue Type: (3)** The type of visual cue presented to the participant, which could consist of color differences in the form of 1. *Hue*, 2. *Saturation*, or 3. *Value*.
- **Cue Intensity: (3)** Each cue had three varying levels of intensity that spanned the range of HSV color space on the HoloLens 2, as described in section 4.2.2.
- **Physical Background: (2)** There were two physical background posters that were displayed behind the virtual imagery shown on the HoloLens 2, which were 1. *Solid Grey* and 2. *Chromatic Aberration*, as described in section 4.2.2.
- **Presented Eye: (2)** The color of the visual cue would be changed away from that of the distractor objects on either the participant's 1. *dominant eye* or 2. *non-dominant eye*. For example, if a visual cue was shown within the grid of distractor objects to the participant's dominant eye, then their non-dominant eye observed a grid of homogeneous objects.

We used a $3 \times 3 \times 2 \times 2$ within subjects study design where the order of each condition was randomized for each factor except the physical background, which was counterbalanced to avoid wear and tear on the paper posters. This led to the experiment taking place in two segments, the first of which consisted of all conditions for one particular background poster, after which the poster was exchanged and the remaining conditions were presented. Each condition was presented to the participant in a block of ten trials, which were randomized so that half of them had a visual cue within the grid of distractor objects and the other half did not have a visual cue and consisted of homogeneous distractor objects. These ten repeated trials were all performed in succession, which allowed for calculation of measures such as false negatives and false positives for each condition.

4.2.3.2 *Measures I*

The main objective measure of the experiment involves the accuracy of the participants' responses, and can be broken down into several combinations based on whether or not a visual cue was present for a particular trial and whether the participant observed the presence of the visual cue:

1. Correct Response: The participant correctly indicated the presence or absence of the visual cue.
2. False Positive: The visual cue was not presented within the grid and the participant responded that they observed the presence of the stimulus.
3. False Negative: The visual cue was presented within the grid and the participant responded that they did not observe the presence of the stimulus.

Based on these possible combinations, we examined error rate, which was the total number of false positives and false negatives divided by the total number of trials (10) for each condition. We also individually examined the false negative rate and false positive rate for each condition.

4.2.3.3 Procedure I

As participants arrived to the testing environment, they were asked to read an informed consent document and provide their verbal consent to participate in the study. While the participant read the consent form, the HoloLens 2 was sanitized via a UV box along with all other equipment the participant would come into direct contact with. Participants were then asked to perform a dominant eye test by performing the hole-in-card test at a distance of two meters to a target on a wall directly in front of them. Participants were then asked to sit in a chair positioned two meters in front of the physical background poster that would be used for their first section of study conditions. The experimenter then explained how to properly don/doff the HMD, and asked the participant to then put on the HMD and hold a keyboard on their lap. Following this, the experimenter opened the virtual scene on the HMD and explained the instructions and sequencing of the study procedure.

For each condition, participants were shown a message and two virtual cursors within the HMD. The message directed the participant to align the two AR virtual cursors while keeping their gaze fixated on the cursors, one of which was world-fixed to the center of the physical background poster positioned two meters in front of the participant, and the other was head-fixed to the participant, also at a distance of two meters. Once participants had aligned the two cursors, they pressed the space button to begin the trial. After pressing space, the message disappeared and only the two cursors were visible for a duration of 2.5 seconds. Following this, the cursors disappeared and the 4×4 grid of cubes appeared for a duration of 250 milliseconds and then disappeared. The participant was then asked whether or not they noticed that one of the cubes appeared differently than the others. The participant pressed ‘y’ or ‘n’ on the keyboard for “yes” or “no”, and were then returned to the initial screen with the two cursors to begin the next trial.

After ten repeated trials of one condition (five with a cue present, five without), the next condition was randomly chosen from a list of remaining conditions. This process repeated until all conditions

for the first background poster were completed. Following this, the experimenter exchanged the background posters, and the process repeated again for all conditions involving that poster.

Upon completing all conditions, the participant was instructed to remove the HMD and place it and the keyboard on a table next to them. Finally, the participant completed a demographics questionnaire and was compensated with fifteen dollars for their time. This procedure took between 45–55 minutes per participant depending on how quickly the participants moved through the conditions.

4.2.3.4 *Hypotheses I*

Based on the previous literature, we formulated the following hypotheses:

- **H1** Hue-based visual cues will be more effective than saturation and value-based cues.
- **H2** Visual cues with higher intensity levels will be more effective (lower error rates) compared to cues with low intensity levels.
- **H3** Visual cues will have similar effectiveness (similar error rates) when the shifted color of the dichoptic cue is presented to the participant's dominant eye compared to when it is presented to their non-dominant eye.
- **H4** Visual cues will be more effective (lower error rates) when displayed over uniform backgrounds compared to visually complex backgrounds.

4.3 Results 1

This section describes the results of the first experiment gathered through analysis of the participants' error rate, number of false negatives, and number of false positives. All analysis took place

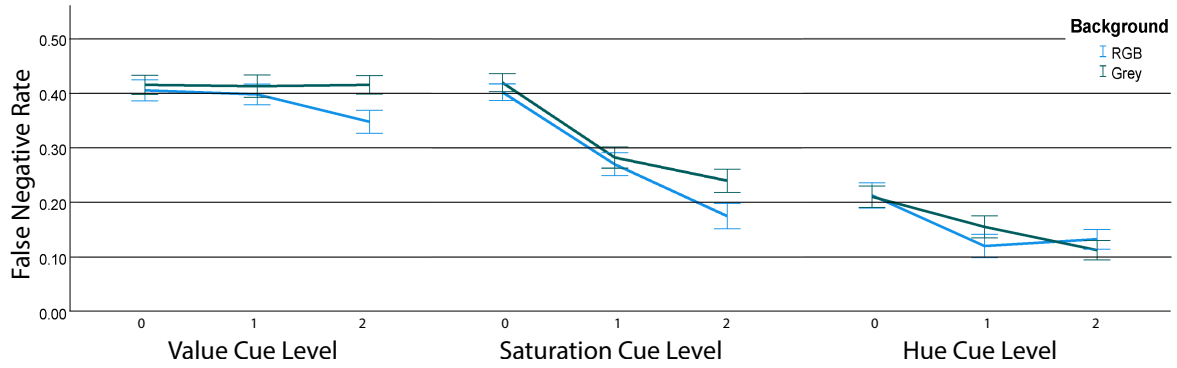


Figure 4.3: This figure depicts a comparison of the false negatives from experiment one across cue type, cue level, and background. Error bars represent standard error with a multiplier of 1.

using SPSS version 28.0.0.0.

Since each condition was repeated in ten trials, prior to analysis these ten trials were aggregated via SPSS to generate mean error rate, mean false negatives, and mean false positives for each condition. Following this, the aggregated data was analyzed with a repeated-measures ANOVA with four factors (cue type, cue intensity, physical background, and presented eye) with levels of 3, 3, 2, and 2 respectively. From this, Tukey multiple comparisons tests were performed with Bonferroni correction at the 5% significance level. We confirmed the normality of the results using Shapiro-Wilk tests set to 5% level and QQ plots.

Our results indicated several significant main effects ($p < 0.05$). Cue type was found to have a significant effect on mean error rate, $F(2,38)=106.91$, $p < 0.001$, $\eta_p^2 = 0.85$, and mean false negatives, $F(2,38)=118.38$, $p < 0.001$, $\eta_p^2 = 0.86$. No significant effect was found on the number of false positives, $F(2,38)=1.02$, $p=0.37$, $\eta_p^2 = 0.05$. Pairwise comparisons revealed significant differences both for error rate and false negatives for each comparison of cue types ($p < 0.001$). These results indicated that value-based cues had the highest error rate ($m=0.462$, $SE=0.011$) and false negatives ($m=0.399$, $SE=0.018$). Hue-based cues had the lowest error rate ($m=0.209$, $SE=0.020$) and false

negatives ($m=0.157$, $SE=0.017$). Finally, saturation-based cues fell between them in both error rate ($m=0.361$, $SE=0.018$) and false negatives ($m=0.298$, $SE=0.015$). Figure 4.4 shows a comparison of the false negative rates broken down by the three cue and the three cue levels.

Cue intensity level was found to have a significant effect on mean error rate, $F(2,38)=42.93$, $p<0.001$, $\eta_p^2 = 0.69$, and false negatives, $F(2,38)=46.77$, $p<0.001$, $\eta_p^2 = 0.71$. No significant effect was found for false positives, $F(2,38)=0.82$, $p=0.45$, $\eta_p^2 = 0.04$. Pairwise comparison revealed significant differences for both error rate and false negatives for each comparison of cue types (all $p<0.01$). These results indicate that overall error rate and false negative error rate reduce with increased cue intensity level, and therefore with increased amounts of color distance between colors shown between the participants' eyes. Cue level 0, which involved the least amount of color distance between the participants' eyes, yielded the highest error rate ($m=0.406$, $SE=0.011$) and the highest false negatives ($m=0.344$, $SE=0.014$). Cue level 1 yielded lower error rate ($m=0.335$, $SE=0.017$) and false negatives ($m=0.273$, $SE=0.016$). Finally, cue level 2, which involved the greatest color distance between the participants' eyes, yielded the lowest error rates ($m=0.292$, $SE=0.018$) and false negatives ($m=0.237$, $SE=0.017$). Again, this effect can be seen in figure 4.4, as well as figure 4.3, which breaks down the results further by including cue type, cue level and background.

Physical background was found to have a significant effect on mean error rate, $F(1,19)=10.79$, $p=0.004$, $\eta_p^2 = 0.36$, and mean false negatives, $F(1,19)=5.47$, $p=0.03$, $\eta_p^2 = 0.22$. No significant effect was found on the number of false positives, $F(1,19)=0.29$, $p=0.59$, $\eta_p^2 = 0.02$. This effect on false negative rates can be seen in figure 4.3.

Finally, presented eye was not found to have any significant effects on error rate, $F(1,19)=0.33$, $p=0.58$, $\eta_p^2 = 0.02$, false negatives, $F(1,19)=0.44$, $p=0.52$, $\eta_p^2 = 0.02$, or on false positives, $F(1,19)=0.01$, $p=0.93$, $\eta_p^2 < 0.01$.

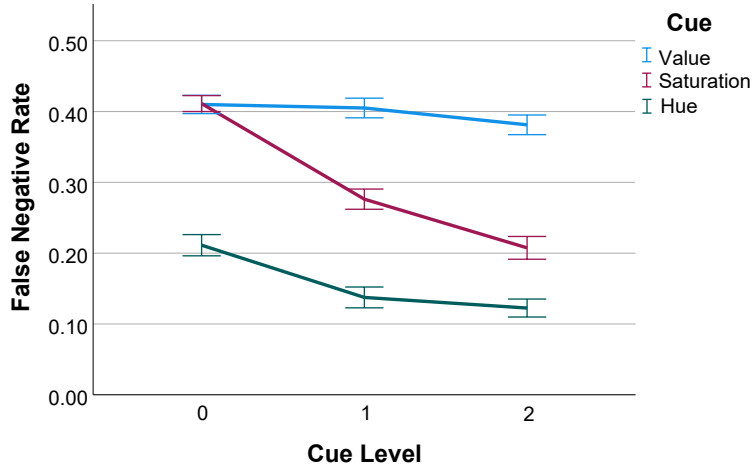


Figure 4.4: This figure depicts a comparison of the false negatives from experiment one across cue type and cue level. Error bars represent standard error with a multiplier of 1.

When examining the interaction effects, several significant results were found. Here we report only the significant interaction effects. There was a significant interaction effect between cue type and cue level on error rate, $F(4,76)=12.68$, $p<0.001$, $\eta_p^2 = 0.40$, and on false negatives, $F(4,76)=15.38$, $p<0.001$, $\eta_p^2 = 0.45$. There was a significant interaction effect between physical background, cue type, and cue level on false negatives, $F(4,76)=2.79$, $p=0.03$, $\eta_p^2 = 0.13$. There was a significant interaction effect between physical background, cue level, and the presented eye on false negatives, $F(2,38)=6.42$, $p=0.004$, $\eta_p^2 = 0.25$. Finally, there was a significant interaction effect between all four independent variables on both error rate, $F(4,76)=3.28$, $p=0.016$, $\eta_p^2 = 0.15$, and on false negatives, $F(4,76)=3.18$, $p=0.018$, $\eta_p^2 = 0.14$.

4.3.1 Hue

We performed an additional analysis on the overall effects of the hue of the objects within the grid of stimuli on the number of false negatives experienced by the participants. Since hue was randomized for each individual trial of each condition, our results showed a relatively uniform

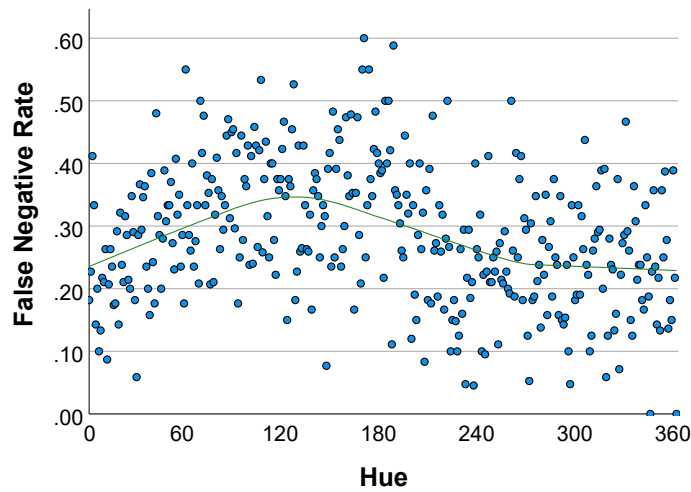


Figure 4.5: This figure depicts the overall effects of the hue of the grid of stimulus objects on the false negative rate from experiment 1 of chapter 3.

distribution of trials across all hues in the range of (0–359) for the HSV color space. Assuming that the hue of the visual cue had no impacts on the rate of false negatives, then we should expect to see a uniform distribution in the plots of hue versus false negative rate, with a horizontal linear fit line. However, as shown in figure 4.5, we see that this is not the case, and the false negative rate increases for hue values near 120, which correspond to green-tinted hues. For reference, red corresponds to both 0 degrees and 360 degrees of hue rotation, green corresponds to 120 degrees, and blue corresponds to 240 degrees.

We examined this in SPSS via use of the locally estimated scatterplot smoothing (Loess) fit line (shown in green in the figure). From this fit line, we observed a distribution with its mean centered at 125.8 hue on the x-axis with a standard deviation of approximately 60. The amplitude was measured to be approximately 0.14, and the vertical shift was measured to be 0.23.

4.4 Experiment II

In this section, we describe an experiment that investigates the subjective qualities of the visual cues investigated in the previous experiment. All participants from the first experiment participated in this experiment immediately after completing the first, therefore all participant data is the same and can be seen in section 4.2.1.

4.4.1 Materials II

All materials used in this experiment are the same as in experiment I (see section 4.2.2).

4.4.2 Methods II

4.4.2.1 Study Design

The experiment consisted of two independent variables:

- **Cue Type: (3)** The type of visual cue presented to the participant, which could consist of color differences in the form of 1. *Hue*, 2. *Saturation*, or 3. *Value*.
- **Cue Intensity: (3)** Each cue had three varying levels of intensity that spanned the range of color space on the HoloLens 2. These intensities were set to the same values used in experiment one.

We used a 3×3 within subjects study design where the order of each condition was randomized for each factor. For this study, all conditions were presented solely in front of the solid grey background poster. The “chromatic aberration” poster was not used because we wanted the participants

to be able to observe the visual cues with a uniform color blending between the virtual imagery and the physical environment. Each condition was presented to the participant along with a user interface depicting three different questions. Participants would respond to these three questions using the number keys on the keyboard, and then the next condition would be displayed. In this section, no distractor objects were used, and participants were free to observe the conditions without time constraints while responding to the questions.

4.4.2.2 *Measures II*

The measures for experiment two consisted of three subjective prompts for the participants to respond to.

1. **Noticeability:** Participants were asked to rate how noticeable the visual cue was compared to the other cues they observed. They rated noticeability using a 9-point scale with 1 meaning least noticeable and 9 meaning most noticeable.
2. **Urgency:** Participants were asked to rate the sense of implied urgency of the visual cue compared to the other cues they observed. “Urgency” was described using an anecdote in which a visual cue was being used as a system notification, and could be a low-urgency notification such as a new text message, or a high urgency notification such as an amber/disaster alert. Participants responded to the prompt using a 9-point scale where 1 meant least urgent and 9 meant most urgent.
3. **Comfortability:** Participants were asked to rate the level of general visual comfort experienced while observing the visual cue while responding to the three subjective prompts. Participants used a 9-point scale where 1 meant least comfortable and 9 meant most comfortable.

4.4.2.3 Procedure II

Upon finishing experiment one (described in section 4.2), the participants immediately started experiment two. Participants were instructed to look to the center of the solid grey background poster presented in front of them at a distance of two meters. The experimenter then started the Unity scene for experiment two on a remote PC via Holographic Remoting.

Upon loading, the participant would see a single randomly chosen visual cue from the set of conditions for experiment two. Above the cue was a user interface that initially asked the participant to rate the level of noticeability of the cue using a scale with a range of one to nine. The participant indicated their response by pressing the number on the keyboard, after which the user interface would update to ask the participant to rate the level of implied urgency of the visual cue. The participant would indicate their response in the same manner, then the user interface would update to ask the participant to rate the level of comfortability of the visual cue. Once the participant responded to this third prompt, a new randomly-chosen cue was picked from the list of remaining visual cues, and the process repeated. Once the participant had responded to all nine visual cues, they were instructed to remove the HoloLens 2.

4.4.2.4 Hypotheses II

Based on the results of the experiment one and those found in the literature, we formulated the following hypotheses:

- **H5** Hue-based visual cues will be rated as more noticeable, more urgent, and less comfortable compared to saturation-based cues and value-based cues.
- **H6** Dichoptic visual cues displayed at higher intensity levels will be rated as more noticeable,

more urgent, and less comfortable compared to lower intensity levels.

4.5 Results II

This section describes the results of experiment two. All analysis took place using SPSS version 28.0.0.0. Similar to experiment one, the data was analyzed with a repeated-measures ANOVA with two factors (cue type, cue intensity) with levels of three and three respectively. From this, Tukey multiple comparisons tests were performed with Bonferroni correction at the 5% significance level. We confirmed the normality of the results using Shapiro-Wilk tests set to 5% level and QQ plots.

Our main results can be seen in figure 4.6. Our results indicated several significant main effects ($p < 0.05$). Cue type had a significant main effect on the sense of implied urgency of the cue, $F(2,38)=3.85$, $p=0.03$, $\eta_p^2 = 0.17$, and on the comfortability of the cue, $F(2,38)=11.82$, $p<0.001$, $\eta_p^2 = 0.38$. No significant main effect was found for cue type on the noticeability of the cue, $F(2,38)=2.17$, $p=0.13$, $\eta_p^2 = 0.10$. Pairwise comparisons revealed a significant difference ($p=0.023$) in the sense of implied urgency between the value-based cues ($m=4.667$, $SE=0.252$) and the hue-based cue ($m=5.592$, $SE=0.278$), indicating significantly higher sense of urgency for the hue-based cues. They also revealed a significant difference ($p<0.001$) in comfortability between the hue-based cues ($m=4.450$, $SE=0.386$) and the value-based cues ($m=5.942$, $SE=0.254$), indicating that value-based cues were more comfortable than the hue-based cues. A second significant difference in comfortability ($p=0.029$) was found between the hue-based cues and the saturation-based cues ($m=5.308$, $SE=0.312$), indicating significantly higher comfort in the saturation-based cues compared to the hue-based cues.

Cue level was found to have a significant main effect on the comfortability of the cue, $F(2,38)=9.51$, $p<0.001$, $\eta_p^2 = 0.33$. No significant effects were found for cue level on noticeability, $F(2,38)=1.52$,

$p=0.23$, $\eta_p^2 = 0.07$, or on the sense of implied urgency, $F(2,38)=1.48$, $p=0.24$, $\eta_p^2 = 0.07$.

Pairwise comparisons revealed a significant difference in comfort ($p=0.008$) between cue intensity level 0 ($m=5.975$, $SE=0.239$) and level 1 ($m=5.108$, $SE=0.313$), as well as a significant difference ($p=0.006$) between cue intensity level 0 and level 2 ($m=4.617$, $SE=0.401$). No significant difference was found between cue intensity level 1 and level 2.

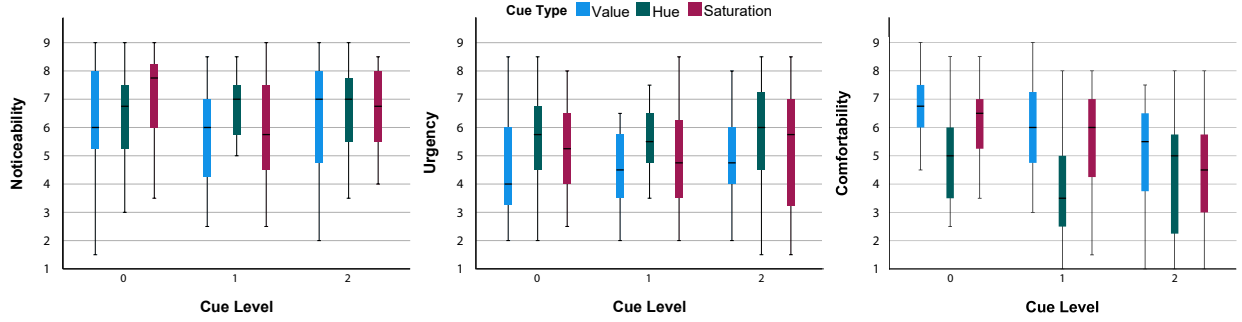


Figure 4.6: This figure depicts the subjective noticeability, urgency, and comfortability results from experiment 2.

4.6 General Discussion

As presented in the 2018 survey by Kamkar et al. the saliency of a visual cue is determined by many different factors, including the cue's appearance in relation to that of distractor objects and the background, as well as factors such as binocular rivalry [92]. In the following sections, we show that the salience of some of the investigated cues can be modeled as a function of HSV color space distances between the two colors that comprise dichoptic cues. These color space distances were previously established in section 4.2.2 and will be discussed in detail in the proceeding sections.

4.6.1 *Saliency of Color-Based Dichoptic Cues*

Through experiment one, hue-based visual cues were found to have the lowest error rates in the pop-out task, followed by saturation-based cues and then finally value-based cues. Further, hue-based visual cues were rated as appearing significantly more urgent compared to the value-based cues. From this, we accept hypothesis **H1** and partially accept hypothesis **H5**. Intensity level was also shown in experiment one to have significant impacts on the error rates of participants, where increased intensity level led to reduced error rates, which leads us to accept hypothesis **H2**. However, in experiment two, intensity level was only shown to have effects on the perceived comfort of the dichoptic cue and no effects on noticeability or urgency were found, therefore we can only partially accept hypothesis **H6**.

As mentioned above, it is possible that the HSV color space distances introduced in section 4.2.2 can be used to predict the saliency of the dichoptic cues. In comparing between the hue-based cues and the saturation-based cues, they both involve HSV color space distances that rely on the radius of the cylindrical color space. Further, they both involve an intensity level with the same distance, where the hue-based cue at intensity level 0 and the saturation-based cue at intensity level 2 both have a distance equivalent to r , the radius of the cylindrical color space. If the HSV color space distance between the two colors used in the dichoptic cue is an effective way to compare saliency levels of the dichoptic cues, then we would expect to see overlapping values for these two specific conditions. Looking at figure 4.4, we can see that this is the case, and that the false negative values for these two conditions do indeed overlap and fall within the range of a standard error (multiplier 1) of each other. However, because we did not originally plan to compare the study conditions in this manner, we only have one point of overlap in the data for these two cue types, and we do not know whether or not hue-based cues would follow the same pattern as the saturation-based cues when measured at the color space distances of $\frac{r}{3}$, and $\frac{2r}{3}$.

For the purpose of this discussion, we ran four additional participants through the study procedure of experiment one, testing only hue-based cues at HSV color space distances of $\frac{r}{3}$ and $\frac{2r}{3}$, so that these conditions could be compared with the saturation-based conditions at the same HSV distances. The data from these participants was plotted along with the initial data from experiment one, and can be seen in figure 4.7. It appears as though participants perform similarly with hue-based cues and saturation-based cues, and that the differences measured previously were a factor of color space distance and not necessary cue type. This implies that HSV color space distance is an effective way to compare the saliency level of color-based dichoptic cues.

We performed a curve estimation regression on the data from figure 4.7 via SPSS that revealed that a power model best fits the data. This curve can be seen as a red line in the figure, and the equation and R^2 values for this model are reported in the caption of the figure. While we acknowledge that four participants is not necessarily an adequate sample size for confirming this, we include the data here as a method of explaining the results of the above studies, and emphasize that this should be evaluated further in future work.

With this model, we can see that hue-based dichoptic cues achieve lower error rates compared to saturation-based cues because they are able to achieve higher HSV color space distances, and thus higher levels of saliency, compared to the saturation-based cues. Saturation-based cues are limited in that the maximum color space distance they can achieve (without combination with other cues) is equal to the radius of the cylindrical HSV color space, since 0 saturation refers to a point on the central vertical axis of the cylinder and 100% saturation refers to a point on the outer edge. Hue-based dichoptic cues have a maximum distance of two times the radius of the color space, which is achieved by performing a 180 degree hue rotation at 100% saturation within the color space.

Unfortunately, it is more difficult to compare value-based cues in a similar manner because value-

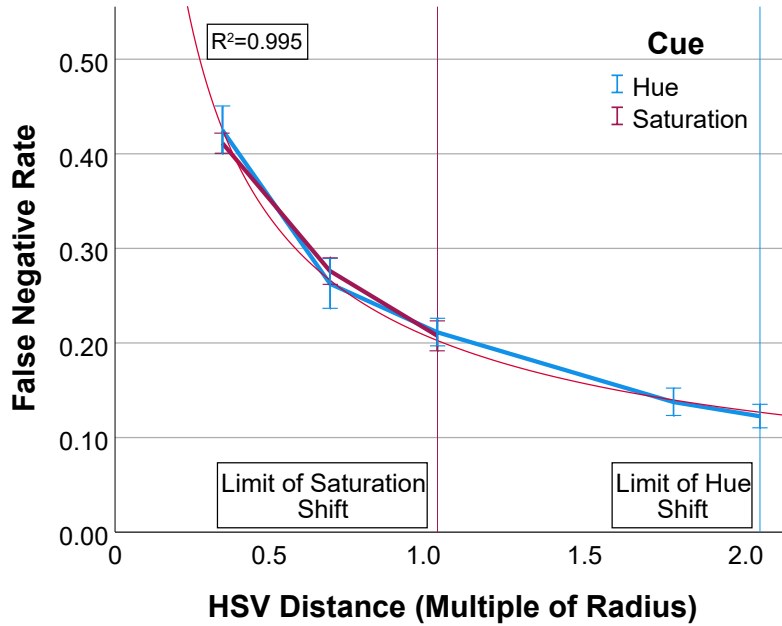


Figure 4.7: This figure compares the false negative rates between the hue-based cues and saturation-based cues in terms of HSV color space distance measured in multiples of the radius of the color space.

based cues have a maximum distance determined by the height of the color space, and we have not established a relationship between the radius and height of the color space. However, the intensity levels of the value-based cues had HSV color space distances with the same multiples of height that the saturation-based cues had as multiples of the radius. Therefore, it is likely that the height of the color space must be less than the radius, which would mean that value-based cues would be less salient than the saturation-based cues, which is supported by our data.

While this is certainly a strong possibility, it is not something that we can determine from this work alone. In order to further evaluate the relationship between value-based dichoptic cues and the other cue types, future work should investigate value-based dichoptic cues in combination with hue or saturation-based cues. For example, comparing user performance in conditions with a variable intensity hue-based cue alone to conditions consisting of both a variable hue and variable value shift. From this, a potential relationship between the radius and height of the color space can

be determined and then value-based cues could be directly compared with hue or saturation-based cues in the saliency model established above.

Another potential reason for the increased false negative rates in the value-based conditions is due to the brightness non-uniformities of the HoloLens 2. Several of our participants mentioned that the appearance of the imagery shown on the display changed based on where it was positioned within the field of view of the device. While the cursor alignment portion of the procedure ensured that the grid of objects would appear in the same portion of the device's field of view for each trial, participants were able to notice this effect on the text-based UIs when orienting their heads downward to look at the keyboard and then back forward towards the background poster. This is also something that we noticed when implementing the project and observing the grid of distractor objects for durations longer than the 250 milliseconds it was displayed for during the study. It is possible that this effect in the grid of objects would make some of the objects appear similar to some of the value-based cues, particularly those displayed at the lower intensity levels. Since this effect should be equally apparent for each trial of the study, it makes sense that participants would treat this effect as a part of the regular appearance of the virtual imagery, and therefore make false negative errors. Future work could determine if this is the case by seeing if user performance improves in a similar experiment using a different OST AR display with less brightness non-uniformities, such as a beam-splitter based OST AR display. This could also potentially be measured on the HoloLens 2 by performing a task in which users observe cues in randomized positions within the device's FOV and are tasked with indicating whether or not the cue is a value-based dichoptic cue or a normal stereoscopic cue.

4.6.2 *Effect of Eye Dominance*

The literature suggested that eye dominance was not an important factor when displaying dichoptic visual cues to an observer [17, 101]. Therefore, based on the results obtained by Logothetis et al. as well as Krekhov et al. we hypothesized in **H3** that there would be a similar effectiveness (similar error rates) between conditions in which the shifted color of the dichoptic cue was presented to the participant's dominant eye versus their non-dominant eye [101, 116]. We observed similar magnitudes of error rates in these conditions in line with this hypothesis.

This result is convenient for the potential applications of dichoptic cues in AR displays, as it implies it is not necessary for the designer of the AR imagery to know the dominant eye of the participant in order to achieve an effective dichoptic cue. Instead, the designer can produce a color shift away from the virtual cues' original color in either eye, while keeping the image the same on the other eye. While all of the conditions investigated in this paper used this technique of creating a dichoptic cue, it is possible that different techniques could be used to gain a similar, or perhaps even better, effect. For example, a slightly different approach could be taken where dichoptic cues are generated via color shifts that are performed for both of the participant's eyes in opposite directions across the color space. In this manner, the cue would likely stand out more from homogeneous distractor objects of the original color in a pop-out task, since neither eye is observing a homogeneous grid, however it is less intuitive to imagine how cues generated by these two techniques would compare in a more practical task, such as a timed search task where participants respond to dichoptic cues positioned within their physical-virtual environment.

4.6.3 *Effects of Background Appearance*

An interesting result of experiment one is that participants made significantly fewer errors when viewing the cues over the chromatic aberration pixelated background compared to the solid grey background. We hypothesized in **H4** that this would be the other way around, therefore we cannot accept this hypothesis.

In forming hypothesis **H4**, we thought that cues presented in front of the uniform background would be easier to identify since there is a uniform color shifting of the cue when the imagery from the AR display and the environment blends together. However, due to color blending, this background solely introduces a luminance shift when the virtual image of the cue blends with that of the physical background. On the other hand, the chromatic aberration background is comprised of pixels with different hues and luminances. So in this case, when the colors of the physical background and virtual dichoptic cue mix, the resulting blended color has both a luminance shift and a hue shift away from its original color. This background also has a higher spatial frequency compared to the uniform background.

The literature on saliency of visual cues has previously established that saliency is increased when multiple features differ between the cue and the distractor objects [24, 92, 140], therefore its possible that the presence of hue-based features and/or spatial frequency-based features in addition to the luminance-based feature contributed to the higher saliency of visual cues displayed over the chromatic distortion background poster.

Turatto and Galfano investigated hue and luminance independently in 2000, where they found that either of these features can be an effective way of capturing the attention of the observer [154]. In 2009, Engmann et al. similarly found that both luminance and hue based gradients applied over a scene significantly bias points of fixation made by their participants, and that when both types of

gradients were applied simultaneously their effects combined linearly to produce a stronger bias in participant fixations [37]. Therefore, it is possible that the hue and luminance features introduced through color blending on the chromatic aberration background conditions combined in a similar manner to produce improved saliency in the AR dichoptic cue.

This result is promising when considering its implications for AR dichoptic cues “in the wild,” since many of the environments that users of pervasive AR systems will find themselves in will contain a multitude of different hues and levels of visual complexity in the physical environment. This implies that such colorful or “noisy” scenes would actually be beneficial in that AR dichoptic cues presented in them are actually easier to detect compared to the more uniform and controlled laboratory settings where these studies typically take place. In such environments, it is also possible that the AR dichoptic cues could combine with the relative motion of the user or of elements in their environment to produce an additional feature that further increases the saliency of the AR cues. Such effects should be investigated in future work.

4.6.4 Subjective Qualities of Dichoptic AR Visual Cues

In experiment two, we hypothesized in **H5** that hue-based dichoptic cues would be rated as more noticeable, more urgent, and less comfortable to observe compared to the saturation-based and value-based cues. Part of this hypothesis was motivated by the results of experiment one, where hue-based cues significantly out-performed the other types of cues, so we expected this would carry over in terms of subjective noticeability and urgency. The comfort portion of this hypothesis was motivated by the observation that the hue-based cues involved larger color space distances between the colors shown to each of the participant’s eyes. We did not find any significant effects in terms of noticeability of the cues in experiment two, and found that hue-based cues were only more urgent compared to value-based cues and not saturation-based cues. However, hue-based

cues were less comfortable compared to both of the other types of cues, therefore we partially accept **H5**.

We also hypothesized in **H6** that as the intensity level of the dichoptic cues increases, their noticeability and sense of implied urgency would increase while the sense of comfort would decrease. The results of experiment two revealed that the intensity level of the cues only had significant effects on the comfortability of the cues, and that as intensity level increased, comfortability decreased, therefore we also partially accept **H6**.

These effects involving the comfortability of the dichoptic cues present an interesting challenge in using the cues effectively “in the wild,” as the cues that were most effective at being noticed by participants were also rated as the most uncomfortable to observe. We should note that in experiment two, participants were allowed to examine the dichoptic cues for as long as they wanted to while indicating their responses on the scales, whereas in experiment one, participants were only exposed to the presence of the cues for durations of 250 milliseconds at a time. When integrating these types of cues into the applications that will use them, exposure time should therefore be carefully considered. Applications could make use of eye-tracking in order to determine that the user has shifted their attention to the cue, after which the cue’s appearance could be shifted back to its original (non-dichoptic besides parallax effects) form. In this manner, potentially negative comfortability effects could at least be reduced and managed.

In terms of the sense of implied urgency of the cues, the only significant difference observed was between the hue-based cues and the value-based cues, which are respectively the most and least salient cues according to the results of experiment one. It makes sense that the most salient cues are perceived as implying a more urgent message or notification, while the least salient cues are perceived as implying one that is less urgent or more casual. If these types of visual cues are used in future applications, then these perceptions should be considered when pairing a cue with a

message or notification, and perhaps hue-based dichoptic cues should be reserved for cases where the message or notification is of high urgency or importance.

It is interesting that no effects were found in experiment two in terms of the subjective noticeability of the different dichoptic cues. We believe that this may be due to the way in which the cues were presented to participants in this experiment. In experiment two, no grid of distractor objects were employed, and the cues were presented one by one. Therefore, it may have been difficult for participants to accurately rate how noticeable each of the cues were, since there was nothing else in the virtual scene to compare it with. This could potentially be revisited in future work, by presenting pairs of cues and comparing subjective factors such as noticeability or implied urgency between them.

4.6.5 Using Dichoptic Cues on OST AR Displays

From the results of our two experiments, we recommend using hue-based dichoptic cues over saturation-based or value-based dichoptic cues. These cues have the widest available range of color space to work with compared to the other two cue types, and thus have the most versatility when it comes to creating cue types at different intensity levels. While we did not compare hue-based dichoptic cues to any more-traditional non-dichoptic cues in this work, it is possible that this type of cue may be a more robust way to draw user attention, since it does not necessarily rely on understanding the appearance and structure of the user's physical-virtual environment in order to successfully stand out. Future work should examine how hue-based dichoptic cues compare with other non-dichoptic cues in at least a subset of many varied environments that future pervasive AR users will find themselves in.

Experiment one revealed the interesting result that green-hued dichoptic cues performed worse compared to cues that were more red or more blue (see figure 4.5.) This finding is similar to what

was shown by Etchebehere and Fedorovskaya in 2017, where they noted a similar decrease in noticeability of visual cues for this color region [53]. As a result of this, we recommend avoiding green-hued dichoptic cues in applications where high saliency of the visual cue is important.

4.7 Summary

In comparing the results of study one to the results obtained by Krekhov et al., the DeadEye attention cues are less effective on OST AR displays compared to flat panel displays, and compared to VR or VST AR displays [100, 101]. While this serves as only a single point of comparison, it is possible that other types of visual attention cues shown on OST AR displays are less effective compared to similar cues shown on more traditional display types. Although environment lighting level was not one of the independent variables tested in the studies of this chapter, chapter 3 demonstrated that the lighting conditions in the user's physical environment affect the perceived appearance of the AR imagery, where dimming the lighting increased the legibility and usability of UIs, and brightening the lighting reduced the legibility and usability of UIs. Therefore, it is possible that environment lighting conditions would have a similar effect on the effectiveness of the visual cues, where their effectiveness may have been more comparable to the results of Krekhov et al. when tested in dimmer lighting conditions, and their effectiveness would be further reduced in brighter conditions. While this effect seems likely, it cannot be concluded from the results of this chapter alone, and future work should investigate the effectiveness of visual cues on OST AR displays in varying lighting conditions to check if a result similar to that from chapter 3 is found.

In this chapter, the visual appearance of the user's environment was shown to have a significant effect on the effectiveness of the visual cues. The study demonstrated that the visual attention cues are more effective when displayed in front of a non-uniform physical background compared to a uniform one. When the attention cues are presented in front of the uniform grey background, they

are perceived to be more opaque and closer to their intended appearance. Switching to the non-uniform background will cause the appearance of the cue to be more transparent as the color of the cue blends with the physical background, and yet this results in increased effectiveness of the attention cue. This particular effect supports thesis statement one by demonstrating that in environments with uniform visual appearances, visual cues may be less effective than when displayed in other environments with non-uniform appearances. This implies that the user may be more likely to fail to notice the presence of certain cues or notifications when in uniform environments. Despite this increase in effectiveness when in non-uniform environment, the cues still do not compare in effectiveness to the similar cues investigated in the studies by Krekhov et al. [100, 101], which implies that while the physical appearance of the user's environment does have a significant effect on the user's perception of virtual imagery shown on the OST AR display, the effect is not as strong as the effect observed of the lighting level in the user's environment from chapter 3.

In the next chapter I investigate how similar factors affect the user's perception of virtual humans shown on OST AR displays.

CHAPTER 5: PERCEPTION OF VIRTUAL HUMANS

The previous two chapters investigated the specific types of AR imagery related to user interfaces (UIs) and visual attention cues, which are types of imagery that are commonly used in many different applications on AR displays. In this chapter, I investigate one additional specific type of AR imagery that is commonly used in applications on AR displays, virtual humans. These studies presented in this chapter were previously published in the following publications:

- [131] Tabitha Peck, Jessica Good, Austin Erickson, Isaac Bynum, and Gerd Bruder. Effects of transparency on perceived humanness: Implications for rendering skin tones using optical see-through displays. *IEEE Transactions on Visualization & Computer Graphics*, 28(5), pages 2179-2189, 2022.
- [35] Doroodchi, Meelad, Priscilla Ramos, Austin Erickson, Hiroshi Furuya, Juanita Benjamin, Gerd Bruder, and Greg Welch. Effects of Optical See-Through Displays on Self-Avatar Appearance in Augmented Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 352-356, 2022.
- [38] Austin Erickson, Gerd Bruder, Greg Welch, Isaac Bynum, Tabitha Peck, Jessica Good, Perceived Humanness Bias in Additive Light Model Displays. *Journal of vision*. pages 1-1, 2022.

The research presented in this chapter is used to partially support thesis statement 1:

- **TS1 - Identify Problems:** In increasingly common viewing conditions, AR users are susceptible to misperceiving or completely missing information shown on their AR display.

Sometimes the user may not even be aware that their AR experience or their performance is being affected.

In particular, this chapter demonstrates that users misperceive virtual humans on OST AR displays depending on the lighting conditions in their physical environment, where users perceive virtual humans to be less human like when environment lighting causes the virtual humans to appear more transparent. This chapter additionally demonstrates that, for the specific case of users creating their own avatars, they choose to adapt the appearance of their avatar based on the particular set of viewing conditions they find themselves in, supporting the idea that future AR systems should adapt to maintain the intended appearance of AR imagery.

5.1 Overview

As mentioned above, the research in this chapter is focused on the particular type of AR imagery that is virtual humans. Imagery of virtual humans can be used to depict a user's avatar, as well as to depict intelligent agents and characters that may be assisting or otherwise interacting with the user. This type of AR imagery is different compared to UIs and visual attention cues in that depictions of virtual humans typically consist of many different colors, rather than a select few colors related to the theme of the UI or application. Because of this, it is possible that user perception is affected differently by the discrepancy between the perceived and intended appearance of the virtual humans on the OST AR display. In particular, this chapter investigates users' subjective perception of virtual avatars and images of real people shown on an OST AR display.

In study one, we create an aggregate measure of humanness through analysis of users' subjective responses to images of real people and virtual avatars as they may appear on OST AR displays. Using this measure, we demonstrate that user perception of the humanness of the people in the

images is directly related to the perceived opacity of the person in the image, where reduced opacity (increased transparency) causes the users to view the people in the images as less human like. We also demonstrate that, because of this effect, virtual humans consisting of darker colors (wearing darker clothes, having darker hair, or having darker skin tones) may be perceived to be less human like compared to other virtual humans.

In study two, we demonstrate that users choose to adapt the appearance of their own avatars when the lighting conditions in their environment are changed, changing aspects of their avatar such as their hair color and skin color in response to the change in environment lighting. This study shows that users are aware of the discrepancy between the perceived and intended appearance of virtual imagery, at least for the case of their own avatars, and if given the chance they would adapt the appearance of the imagery to better match its intended appearance. This study provide support for future AR systems capable of adapting to maintain the intended appearance of virtual imagery in response to changes in the user's viewing conditions.

5.2 Subjective Qualities of Transparent AR Virtual Humans

In this section, we present a user study that investigates users' subjective perceptions of virtual humans and avatars as they would be perceived on an OST AR display under varying lighting levels.

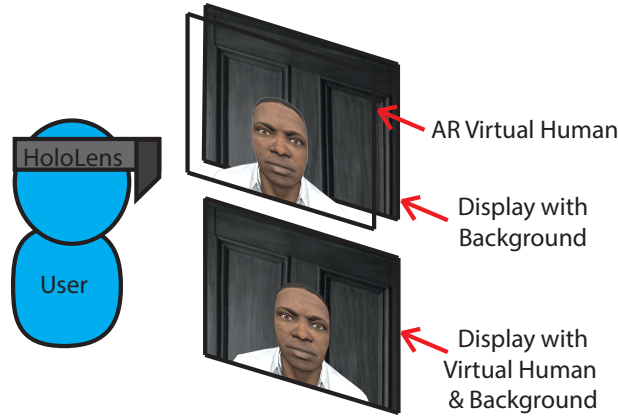


Figure 5.1: Illustration depicting the setup of the perceptual matching task, in which a participant compares the appearance of an AR virtual human (top) to an identical virtual human displayed on a flat-panel display (bottom). Note that the AR imagery was displayed at the same depth as the flat panel display and appears in front of the background in this illustration only to emphasize that the top virtual human consists of a combination of AR imagery and a flat-panel display.

5.2.1 Methods

5.2.1.1 Image Generation

In order to generate images that would be representative of what the user would observe on an OST-HMD we needed to understand how to model the transparency and color blending effects that are inherent in OST AR imagery. While this effect has been formally modeled by Gabbard et al. [64] for OST displays in terms of the blending of light from the user’s environment and light from the display, to our knowledge there does not exist a formulaic method of generating 2D images that capture such effects. However, in 2020 Gabbard et al. pointed out that color accuracy and robustness can be appropriately measured through the use of *perceptual matching* tasks where users are tasked with matching imagery displayed on an optical-see through display to imagery shown on a controlled display for which color parameters have already been established [66]. Thus to capture the effects of color blending and transparency for our experimental images we

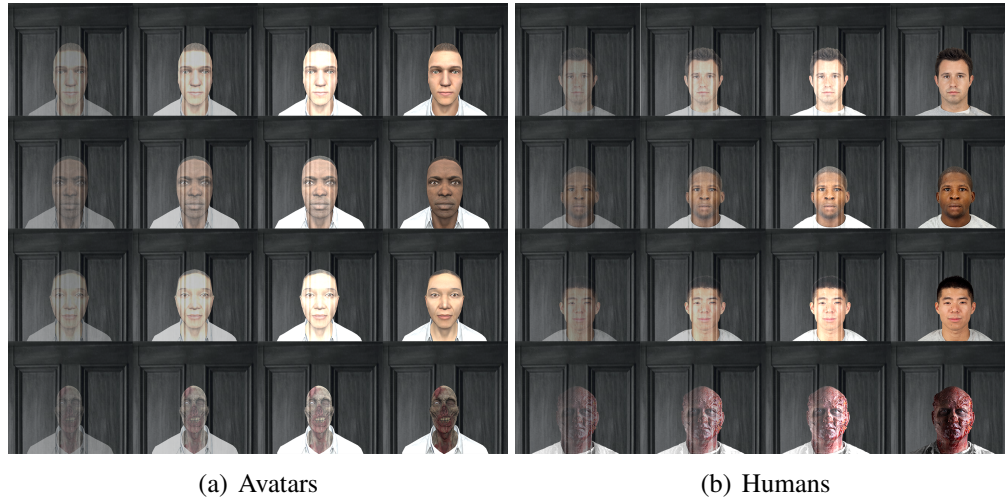


Figure 5.2: Visual stimuli used in the experiment: (a) avatars representing (from top to bottom) White, Black, East Asian, and Zombie categories, (b) humans representing (from top to bottom) White, Black, East Asian, and Zombie categories. For both (a) and (b), columns represent (from left to right) 35% opacity, 68% opacity, 100% opacity, and 100% opacity with color correction.

chose to perform a perceptual matching task.

The perceptual matching task was performed by one participant, who wore an AR HMD, the Microsoft HoloLens 1. They compared the appearance of an AR virtual human, displayed on the device in front of a flat panel display, to the appearance of the same virtual human displayed on a second flat panel display at the same distance as the AR virtual human (see figure 5.1.) The AR avatar was positioned to be in front of a display depicting the virtual background of a dark door image (see figure 5.2). A second display was directly beneath the flat panel display and was connected to a PC running a Photoshop 2021 project containing on separate layers 1) images of the same avatar segmented from the background and 2) the background image.

The participant adjusted the layer blending options and layer opacity levels of the virtual human in Photoshop to best match the appearance of the AR avatar on the HoloLens. Due to the potential impact of the tinted visor on the front of the HoloLens, the user would alternate between looking

through the HMD at the AR avatar and lifting the HMD to see the avatar from the Photoshop project. The results of this task suggested that using the linear dodge (additive) layer blending option on the avatar's Photoshop layer, along with a layer opacity of 35 percent yielded images that closely matched each other.

It should be noted that there are several limitations with generating images in this manner. Most notable is that the resulting Photoshop parameters are dependent on both the physical lighting conditions present in the testing environment as well as the luminance capabilities of the OST-HMD. Decreasing the intensity of the physical lighting in the testing environment or increasing the luminance capability of the HMD would yield higher layer opacity parameters than what was obtained here, and vice versa. While this would be problematic if we wanted to compare avatars in different physical environments or different physical lighting conditions, these factors remain consistent across all avatar images used in this study. Thus this limitation should not impact our results. Images generated in this manner also do not take into account the effects of luminance non-uniformity on the OST- HMD [29, 106], which may or may not have effects on the user's perception of the avatars. However, this effect is difficult to quantify, even with the proper optical equipment, and so we leave investigation of this factor to future work.

We used the above-mentioned parameters to generate the first set of virtual human images that can be seen in the left-most columns of figure 5.2 a and b. These images provide a baseline for how virtual humans would appear on a current consumer OST display, specifically the HoloLens 1, under indoor lighting conditions. To compare against these images, we generated three additional sets of images that consider how the appearance of the avatars would change as the display technology advances. The next two sets of images (second and third from the left column in figure 5.2 a and b) depicts how the avatars would appear should the luminance capability of the display increase to result in opacity parameters in a similar perceptual matching task of 68 percent and 100 percent, respectively. The value of 68 percent was chosen as a midway point between the current opacity

values found for the HoloLens 1 and the maximum values possible at 100 percent opacity. For the final set of images (rightmost column in figure 5.2 a and b), we reverted the layer blending setting on the avatar from linear dodge (additive) to normal mode, which represents what users may see on a future OST display in which the issues of transparency and color blending are resolved. These image sets make up the four different opacity levels referenced in the remainder of the paper:

- **35% Opacity:** Current Display Opacity
- **68% Opacity:** Improved Display Opacity
- **100% Opacity:** Maximized Display Opacity (without color correction)
- **100% Opacity CC:** Maximized Display Opacity with Color Correction

Since these images would be viewed by participants on their own personal displays, it is possible that there were slight variations between how the images appeared to each participant. Such variations could potentially include: differences in brightness, contrast, and color tone of the image. However, since we are interested in comparing participants' collective perceptions between the different images rather than perceptual differences between participants, our results should not be negatively impacted by this variation. Further, such display variations should only reduce the likelihood of finding significant effects, as extreme display settings would make it more difficult to identify differences between images, effectively dampening significant effects with cases in which no effect was observed.

All 3D avatars used in our study were collected from the Microsoft Rocketbox Avatar Library ¹ except for the zombie model, which was obtained from Adobe Mixamo ². The avatars were all

¹<https://github.com/microsoft/Microsoft-Rocketbox>

²<https://www.mixamo.com/>

retextured to appear as though they were wearing the same white collared shirt. For the real human conditions, images were gathered from the Chicago Face Database ³, except for the zombie image which was obtained from a stock image library. Images from the Chicago Face Database are standardized photographs of human faces that are intended for controlled scientific research. The faces selected for this experiment came from the main data set with neutral, non emotive faces. The door image used as a background in all conditions was similarly obtained from a stock image library. This exploratory study limited images to only male-appearing faces since racial stereotyping and dehumanization literature supports these effects on men. However, future work should investigate the effects on humanness caused by opacity and race for female and gender-neutral appearing faces as well.

5.2.1.2 Experimental Procedure

The procedure and all materials were approved by the Davidson College Institutional Review Board prior to beginning data collection. Participants initially accessed the study by clicking on the study URL from within the CloudResearch recruitment listing. The study was conducted online (hosted via the Qualtrics survey platform). Participants first read and completed informed consent and an eligibility checklist, which verified that they were at least 18 years old, had normal or corrected to normal vision, and were proficient in written and spoken English (since all study materials were presented in English).

Eligible, consenting participants were then randomly assigned to view and evaluate images from 2 of the 4 racial group conditions for a total of 16 images (4 opacity level \times 2 appearance \times 2 race). As recommended by Little and Jubin [111] we used a planned missing data design and chose to capture participants' evaluations of only 2 of the 4 racial groups in order to reduce participant

³<https://www.chicagofaces.org/>

fatigue and poor quality responding. Planned statistical imputation of missing data allowed for estimation of the other two racial groups for each participant. For each image, participants rated the figure using a series of 14 questions (see section 5.2.1.3). Participants were presented with the 16 images (and subsequent questions) in random order. Embedded within the image questions were two attention check questions which were used to screen out inattentive participants. After evaluating all images, participants completed demographic questions. Participants were compensated according to agreed upon rates with CloudResearch; participation lasted on average 19.35 minutes (min = 8.33, median = 15.30).

5.2.1.3 Measures

Participants evaluated the extent to which each figure seemed human, animal-like, robotic, competent, friendly, dangerous, angry, happy, creepy, and unearthly. Participants also evaluated the extent to which each figure had emotions, felt physical pain, had complex thoughts, and had control over their actions. All responses were given on a 100-point sliding scale ranging from “not at all” to “very much” with the initial selection snapped at the midpoint. Participants were required to move each slider before continuing to the next image.

Our primary interest was in assessing *humanness*, though we also included two items assessing emotion in order to control for possible baseline differences in facial expression across the images. Research has demonstrated that White participants tend to perceive faces of Black Americans as angrier than faces of White Americans [73, 83] even when facial musculature/expression is held constant. We also included items assessing stereotypical perceptions of racial groups in order to verify that our manipulation of the racial groups was successful. Psychological research on the Stereotype Content Model indicates that social groups are stereotyped along two primary axes, warmth and competence [56]. Contemptuous prejudice is directed at groups that are stereotyped

as low in both competence and warmth (e.g., Black Americans), while envious prejudice is directed at groups that are stereotyped as high in competence but low in warmth (e.g., Asian Americans). Groups that are stereotyped as high in both warmth and competence generally receive social admiration (e.g., White Americans).

Therefore, participants rated how *happy* and how *angry* each figure was, and we included these items in the overall experimental model in order to account for any variation in facial expression. We conducted separate analyses with the *friendly*, *competent*, and *dangerous* items in order to verify expected racial group differences in warmth and competence as well as the common stereotype that Black men are dangerous [32,33].

For the remaining items assessing humanness, we conducted an exploratory principal components factor analysis with varimax rotation to determine how many factors best fit the data. Two factors with eigenvalues greater than 1 emerged. We retained items that loaded above .5 on only one of the two factors. The first factor (eigenvalue = 4.36, 48.41% variance explained) included 6 items (human, robot, unearthly, emotions, physical pain, and complex thoughts) and showed high scale reliability (Cronbach's alpha = .904). the second factor (eigenvalue = 2.01, 22.30% variance explained) included 3 items (animal, creepy, and control over actions) and demonstrated acceptable scale reliability (Cronbach's alpha = .762), but low conceptual meaning. The three items on the second factor seemed to be measures of humanness that did not fit the current study context well given that the targets in the present study were virtual (not animalistic) and static (did not have action). Therefore, we computed a single composite measure of *humanness* as the mean of the following items: human, 100-robot, 100-unearthly, emotions, physical pain, and complex thoughts.

5.2.1.4 Participants

160 participants, recruited and compensated through the CloudResearch platform, passed the attention checks and were included in the analysis. Participants self-verified that they met all inclusion criteria. A general population of participants were recruited to match the United States Census to support the generalizability of results. 49% of participants identified as women, 49% as men, and 2% as non-binary. The racial make-up of participants was 76% White, 12% Black of African American, 4% Hispanic/Latinx, 4% East Asian, 2% American Indian or Alaska Native, and the remaining 2% as other races. Education level ranged from 3% with some high school, 38% with a high school degree, 6% with a 2-year college degree, 33% with a 4-year college degree, 14% with a masters degree, and 6% with a doctoral degree. House-hold income demographics included 22% earning under \$25,000, 27% earning between \$25,000–\$50,000, 19% earning between \$50,000–\$75,000, 10% earning between \$75,000–\$100,000, 5% earning between \$100,000–\$125,000, 6% earning between \$125,000–\$150,000, 3% earning between \$150,000–\$175,000, and 8% over \$175,000. Finally, 36% were non-video-game players, 44% casual video-game players, 12% core-video game players, and 8% hard-core video game players.

5.2.2 Results

Analysis was performed in R 4.0.0. Planned missing data were imputed using predictive-mean matching [112] and implemented with the mice 3.11.0 library. After imputation, submeasures were calculated as described in section 5.2.1.3.

Table 5.1: Final linear regressions for the *humanness* measure considering opacity, race, angry, and happy. Human faces are in the left column and avatars are in the right column. Significance codes: *** < 0.001, ** < 0.01, * < 0.05, · < 0.1.

	Humans $R^2 = .34$					Avatars $R^2 = .39$				
	β	Std. Error	t-value	p		β	Std. Error	t-value	p	
(Intercept)	82.97	2.72	30.51	<0.001	***	75.35	3.76	20.03	<0.001	***
gender	–	–	–	–		-5.30	2.04	-2.60	0.01	*
happy	-0.02	0.06	-0.38	0.70		0.21	0.08	2.56	0.01	*
angry	-0.67	0.16	-4.29	<0.001	***	-0.74	0.15	-4.92	<0.001	***
Opacity	9.54	3.31	2.88	<0.01	**	12.13	4.35	2.79	<0.01	**
RaceBlack	-2.15	3.74	-0.58	0.56		8.70	4.98	1.74	0.08	·
RaceWhite	-5.58	3.69	-1.51	0.13		4.91	4.86	1.01	0.31	
happy:angry	0.00	0.00	0.64	0.53		0.00	0.00	0.10	0.92	
happy:Opacity	-0.04	0.07	-0.65	0.52		-0.43	0.11	-3.95	<0.001	***
angry:Opacity	-0.17	0.23	-0.76	0.45		-0.18	0.20	-0.86	0.39	
happy:RaceBlack	0.14	0.09	1.60	0.11		-0.05	0.13	-0.43	0.67	
happy:RaceWhite	0.10	0.08	1.17	0.24		-0.12	0.12	-1.04	0.30	
angry:RaceBlack	0.12	0.17	0.68	0.50		0.27	0.18	1.55	0.12	
angry:RaceWhite	-0.03	0.19	-0.18	0.86		-0.12	0.19	-0.65	0.52	
Opacity:RaceBlack	1.19	4.69	0.25	0.80		-5.23	6.23	-0.84	0.40	
Opacity:RaceWhite	3.60	4.66	0.77	0.44		-6.63	6.10	-1.09	0.28	
happy:angry:Opacity	0.00	0.00	0.98	0.33		0.00	0.00	1.79	0.07	·
happy:angry:RaceBlack	0.00	0.00	-0.55	0.59		-0.01	0.00	-1.51	0.13	
happy:angry:RaceWhite	0.00	0.00	0.37	0.71		0.00	0.00	0.54	0.59	
happy:Opacity:RaceBlack	-0.18	0.12	-1.58	0.12		0.17	0.18	0.96	0.34	
happy:Opacity:RaceWhite	-0.07	0.11	-0.60	0.55		0.11	0.17	0.66	0.51	
angry:Opacity:RaceBlack	0.44	0.24	1.80	0.07	·	0.03	0.24	0.13	0.90	
angry:Opacity:RaceWhite	0.55	0.26	2.11	0.04	*	0.21	0.25	0.82	0.41	
happy:angry:Opacity:RaceBlack	0.00	0.00	-0.91	0.36		0.00	0.01	-0.28	0.78	
happy:angry:Opacity:RaceWhite	-0.01	0.01	-1.81	0.07	·	0.00	0.01	-0.76	0.45	

5.2.2.1 Validation Analyses

To validate the *humanness* measure we performed a 2 (*zombie*: zombie, people) \times 2 (*appearance*: avatar, human) within-participants analysis of variance (ANOVA). See figure 5.3. Zombies ($M = 69.02$, $SD = 30.21$) had significantly lower *humanness* compared to people ($M = 74.86$, $SD = 25.88$), $F(1, 159) = 4.17$, $\mathbf{p} = .04$, $\eta^2 = .02$. Further, humans ($M = 77.94$, $SD = 23.24$) had significantly higher *humanness* compared to avatars ($M = 68.96$, $SD = 29.87$), $F(1, 159) = 180.76$, $\mathbf{p} < .0001$, $\eta^2 = .03$. Finally, there was a significant interaction, $F(1, 159) = 16.74$, $\mathbf{p} < .0001$, $\eta^2 = .003$. When considering human faces, people had higher *humanness* compared

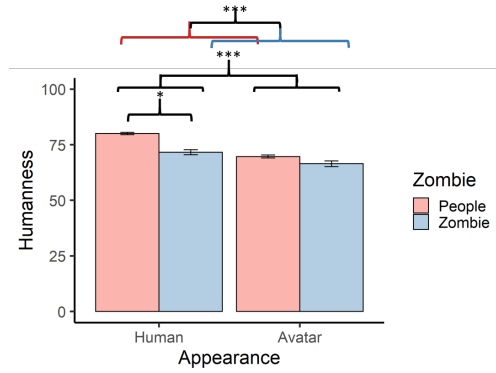


Figure 5.3: A bar chart of average humanness with standard error bars. Zombie faces (blue) had significantly lower humanness compared to people (red). Further, avatar faces had lower humanness compared to human faces.

to zombies, $t(175) = 2.88$, $p = .004$, $d = .37$. However, no significant difference was found between zombies and people in the avatar group, $t(175) = 1.10$, $p = .27$, $d = .11$.

These results support that the *humanness* metric did measure humanness since people had higher humanness compared to zombies and humans had higher humanness compared to avatars. The zombie faces were removed from the remainder of this analysis since they were included only to determine the construct validity of the humanness measure.

Because only one human and one avatar face was used to represent each racial category, we tested for expected differences in stereotyping based on racial category as a way of validating the manipulation of race. We tested for an effect of race on each of the *competent*, *friendly*, and *dangerous* measures using a 2 (*appearance*: avatar, human) \times 3 (*race*: Asian, Black, White) ANOVA. Post-hoc analysis was performed using estimated marginal means with Tukey method adjustments for repeated tests by comparing race pairwise within each separate appearance face group. The highest-order significant effect is reported. Human faces ($M = 70.81$, $SD = .72$) were perceived to be significantly more *competent* compared to avatar faces ($M = 63.46$, $SD = .78$),

$F(1, 159) = 83.04$, $\mathbf{p} < .0001$, $\eta^2 = .02$. A significant race \times appearance interaction was identified for the *friendly* measure, $F(1, 318) = 4.98$, $\mathbf{p} = .007$, $\eta^2 = .002$. The Asian human face ($M = 34.59$, $SD = 1.49$) was rated significantly friendlier than the Black human face ($M = 24.77$, $SD = 1.21$), $t(362) = 2.76$, $\mathbf{p} = .02$, $d = .86$. No differences were identified within the avatar faces. Finally, a significant race \times appearance interaction was identified for the *dangerous* measure, $F(1, 318) = 3.15$, $\mathbf{p} = .04$, $\eta^2 = .001$. Consistent with expectations based on stereotypes [33], the Black human face ($M = 19.43$, $SD = 1.10$) was rated significantly more dangerous than the Asian human face ($M = 11.97$, $SD = .85$), $t(377) = 2.72$, $\mathbf{p} = .02$, $d = .62$. No differences were identified within the avatar faces.

Having verified that we successfully manipulated racial category membership, we next sought to ensure that race was the only variable manipulated and that facial expression was controlled across the various targets. We therefore tested for an effect of race on the angry and happy measures using a 2 (appearance) \times 3 (race) ANOVA. See figures 5.4 and 5.5. For the *angry* measure, significant main effects of both appearance, $F(1, 159) = 10.98$, $\mathbf{p} = .001$, $\eta^2 = .002$, and race, $F(1, 318) = 7.86$, $\mathbf{p} = .0004$, $\eta^2 = .03$, were found. Avatar faces ($M = 17.75$, $SD = .62$) were rated as angrier than human faces ($M = 15.81$, $SD = .58$), $t(159) = -3.31$, $\mathbf{p} = .001$, $d = .07$. Further, Black faces ($M = 22.25$, $SD = .85$) were rated as angrier than both Asian faces ($M = 12.26$, $SD = .60$) and White faces ($M = 15.83$, $SD = .70$), $t(318) = -3.91$, $\mathbf{p} = .0003$, $d = .08$ and $t(318) = 2.51$, $\mathbf{p} = .03$, $d = .05$ respectively. This finding is consistent with past research showing that anger is more likely to be associated with Black faces compared to White faces [73, 83].

When considering the *happy* measure, significant main effects of both appearance, $F(1, 159) = 49.95$, $\mathbf{p} < .0001$, $\eta^2 = .01$, and race, $F(1, 318) = 8.99$, $\mathbf{p} < .0001$, $\eta^2 = .03$, were found. Additionally, there was a significant appearance \times race interaction $F(1, 318) = 15.01$, $\mathbf{p} < .0001$, $\eta^2 = .007$. The Asian human face ($M = 38.27$, $SD = 1.32$) was perceived to be happier than

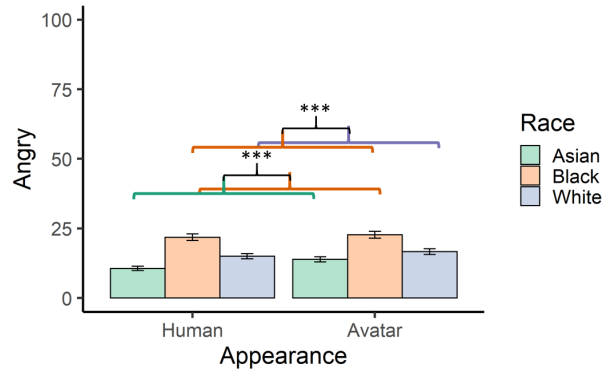


Figure 5.4: A bar chart of the average angry score with standard error bars. Black faces (orange, middle) were rated angrier than both Asian (green, left) and White (purple, right) faces.

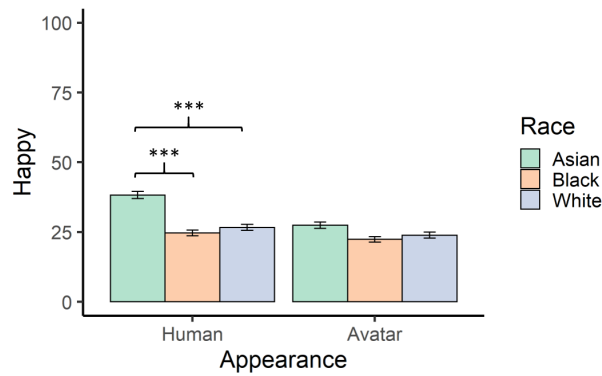


Figure 5.5: A bar chart of the average happy score with standard error bars. Within the human faces (left), the Asian (green) face were rated happier than both the Black (orange) and White (purple) faces.

both the Black human face ($M = 24.70$, $SD = 1.03$) and the White human face ($M = 26.63$, $SD = 1.08$), $t(405) = 5.43$, $p < .0001$, $d = .63$ and $t(405) = 4.66$, $p < .0001$, $d = .58$ respectively. No significant differences were identified between the avatar faces.

Because perceptions of emotion in facial expression differed across the various faces, we sought to use angry and happy as covariates in our primary analyses of the effect of opacity and race on

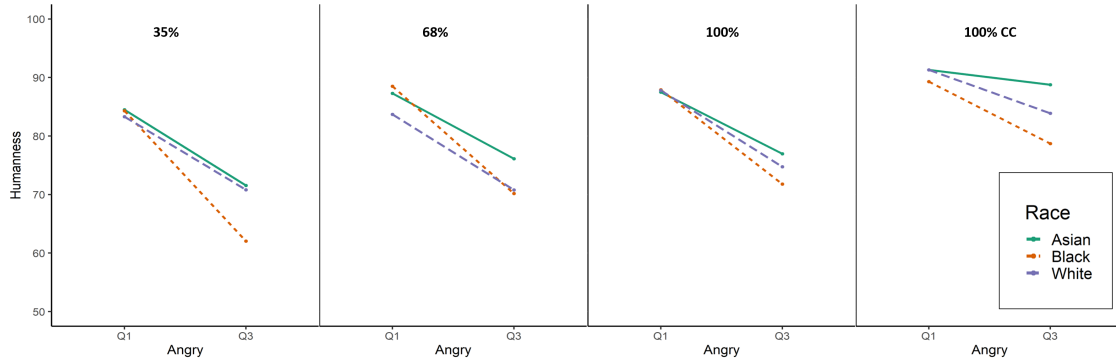


Figure 5.6: A visual representation of the 3-way *humanness* \times *opacity* \times *race* interaction for human faces. Each graph displays data for one opacity level increasing from left (35% opacity) to right (100% with color correction) for each race (Asian: green, solid, Black: orange, small-dash, White: purple, long-dash). The humanness level (y-axis) based on each race's perceived angry level (x-axis) at quartile 1 (left) and quartile 3 (right). As opacity increases, humanness increases. As perceived angryness increases, humanness decreases.

humanness. However, the angry and happy measures both failed the assumption of homogeneity of regression slopes, a necessary assumption for analysis of covariance. Therefore, the two items were entered as factors rather than covariates in our subsequent analyses. Finally, avatar and human faces were analyzed separately since *humanness* was significantly different between the two groups.

5.2.2.2 Human Faces

Multiple regression was used to test experimental effects on *humanness* for human faces. Hierarchical models were built, sequentially adding opacity, angry, race, and happy measures, in that order. Starting with opacity, adding angry significantly improved the model fit, $\chi^2(2) = 758.53$, $\mathbf{p} < .0001$. Adding race, again significantly improved model fit, $\chi^2(8) = 34.42$, $\mathbf{p} < .0001$, and finally adding happy improved model fit, $\chi^2(12) = 40.91$, $\mathbf{p} < .0001$. Adding participant gender as a covariate did not significantly improve the model and it was therefore removed. The final

regression can be seen in table 5.1.

Significant main effects of angry, $t(1775) = -4.29$, $\mathbf{p} < .0001$, $d = .20$, and opacity, $t(1753) = 2.88$, $\mathbf{p} = .004$, $d = .14$, were found, such that lower perceptions of anger and greater opacity were associated with higher ratings of humanness. These main effects were modified by a higher-order angry \times opacity \times race 3-way interaction, $t(1775) = 2.11$, $\mathbf{p} = .03$, $d = .10$. The significant 3-way interaction was evaluated for each race by comparing slopes, pairwise, with Tukey adjustments. See figure 5.6. No significant differences in slope were identified for the Asian human face. However, for both the Black and White human faces significant differences were found between 100% opacity with color correction and all other levels. For the Black face, the slope at 100% opacity with color correction ($\beta = -.29$) was less steep compared to 100% opacity ($\beta = -.35$), $t(481) = -3.50$, $\mathbf{p} = .003$, 68% opacity ($\beta = -.41$), $t(481) = -4.39$, $\mathbf{p} = .0001$, and 35% opacity ($\beta = -.48$), $t(484) = -5.46$, $\mathbf{p} < .0001$. A similar pattern was seen for the White face, the slope at 100% opacity with color correction ($\beta = -.31$) was less steep compared to 100% opacity ($\beta = -.57$), $t(484) = -4.31$, $\mathbf{p} = .0001$, 68% opacity ($\beta = -.45$), $t(484) = -3.73$, $\mathbf{p} = .001$, and 35% opacity ($\beta = -.56$), $t(481) = -4.62$, $\mathbf{p} < .0001$. In other words, anger was less predictive of humanness for the White and Black fully opaque faces with color correction compared to the White and Black faces at all other opacity levels. Whereas the effect of anger on humanness did not differ based on opacity for Asian faces. For all faces, lowering opacity reduced humanness and faces that were perceived as angrier were perceived as less human. Further, the Black face was perceived to be significantly angrier than both the Asian and White faces (see figure 5.4) thus having a greater effect on perceived humanness at lower opacity levels (see figure 5.6, left).

5.2.2.3 Avatar Faces

The same hierarchical regression analysis was performed to test experimental effects on humanness for avatars. Starting with opacity, adding race significantly improved the model fit, $\chi^2(4) = 11.31$, $\mathbf{p} = .02$. Adding angry, again significantly improved model fit, $\chi^2(6) = 947.68$, $\mathbf{p} < .0001$, adding happy again improved the model fit $\chi^2(12) = 67.40$, $\mathbf{p} < .0001$, and finally adding participant gender as a covariate improved the model $\chi^2(2) = 7.12$, $\mathbf{p} = .03$. The final regression can be seen in table 5.1.

A significant main effect of gender (male, female, non-binary) was found, $F(2, 157) = 3.80$, $\mathbf{p} = .03$, $\eta^2 = .05$. Women ($M = 66.66$, $SD = 1.00$) gave significantly lower humanness ratings to avatars compared to men ($M = 72.62$, $SD = .89$), $t(157) = 2.74$, $\mathbf{p} = .02$, $d = .15$. There was also a significant main effect of angry, $t(1794) = -4.92$, $\mathbf{p} < .0001$, $d = .23$. As found with human faces, avatars that were perceived to be angrier were perceived to be less human, ($\beta = -.62$), $\mathbf{p} < .0001$.

Significant main effects of happy, $t(1793) = 2.57$, $\mathbf{p} = .01$, $d = .12$, and opacity, $t(1759) = 2.80$, $\mathbf{p} = .005$, $d = .13$, were qualified by the higher-order significant happy \times opacity interaction, $t(1770) = -3.94$, $\mathbf{p} < .0001$, $d = .19$. The significant 2-way interaction was evaluated by comparing slopes, pairwise, with Tukey adjustments. Significant differences in slope were identified between 35% opacity ($\beta = .16$) and 100% opacity with color correction ($\beta = -.27$), $t(1752) = 4.47$, $\mathbf{p} < .0001$, $d = .21$. Trends were found between 35% opacity and both the 100% opacity without color correction ($\beta = -.19$) and 68% opacity ($\beta = -.07$), $t(1746) = -2.43$, $p = .07$, $d = .12$ and $t(1740) = 2.44$, $p = .07$, $d = .12$ respectively. See figure 5.7. In essence, for the most transparent avatars (35%), perceived happiness predicted greater humanness, whereas for the most opaque avatars (100% with color correction), happiness predicted lower humanness.

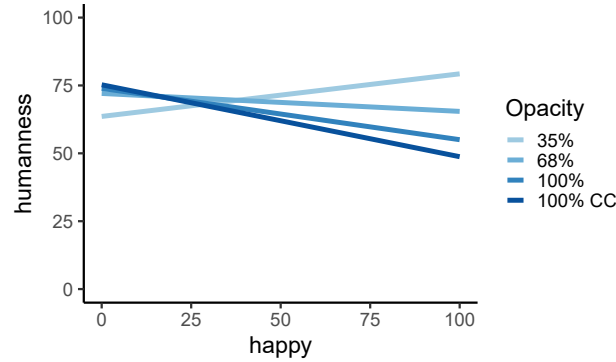


Figure 5.7: The relationship between *happy* and *humanness* for avatar faces at each opacity level.

5.2.2.4 Humanness Implications

Although the present study was exploratory, with the main goal being to test the effect of opacity on *humanness*, our final analyses explored possible implications of humanness. In other words, given that reduced opacity predicts lower ratings of humanness, what are the possible downstream effects of reduced humanness? To explore this, we computed bivariate correlation coefficients between humanness and the measures of stereotyping (competent, friendly, dangerous). The results are shown in figure 5.8. Humanness significantly correlated with friendly, competent, and dangerous for both avatars and humans with all $p < .0001$. Of note, greater humanness was associated with less danger (Human: $r = -.6$, Avatar: $r = -.66$), greater competence (Human: $r = .56$, Avatar: $r = .63$), and less friendliness (Human: $r = -.22$, Avatar: $r = -.31$).

5.2.3 Discussion

The primary goal of the present study was to explore whether a known inconsistency in the rendering of certain colors with OST-HMDs injected unintentional racial bias into AR applications. Because the color black appears transparent, this means that dark-skinned (i.e., Black) virtual hu-

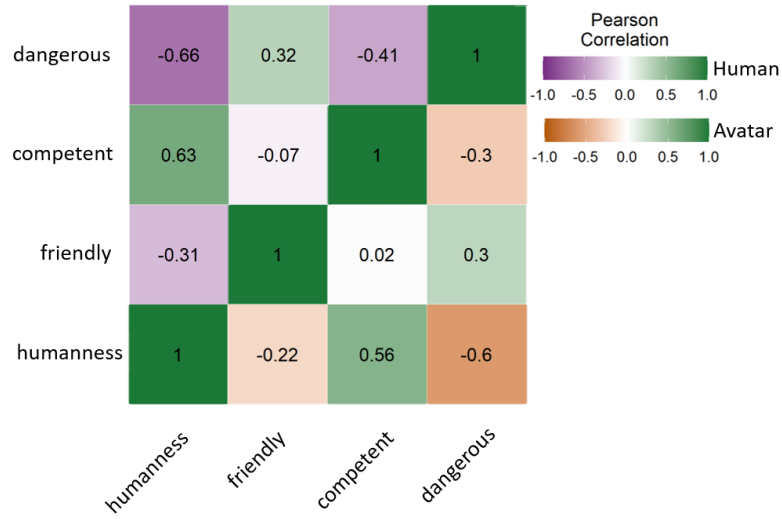


Figure 5.8: Pearson correlation coefficient matrix between humanness, competent, friendly, and dangerous. The avatar face correlations are in the top left triangle and the human faces are in the bottom right.

mans are perceived by AR users as more transparent in the same environments in which White virtual humans are seen as less transparent. We tested whether opacity impacted perceptions of *humanness* broadly, and as a function of virtual human race. We chose humanness as our outcome measure because of the strong existing literature on the importance of humanness in virtual technology (e.g., Uncanny Valley research [123]) and the historic and ongoing role of dehumanization in perpetuating violence and exploitation of racial groups [69, 87].

Overall, across both static human and avatar faces, we found that opacity affected humanness. As opacity increased, perceptions of humanness increased. Women perceived avatar faces to be less human than men. Notably, we did not find that that race moderated this effect. In other words, regardless of human and avatar race, more transparent figures were perceived as less human. However, current OST-HMDs render lighter versus darker skin tones at different levels of transparency. While certain ambient lighting conditions can make all virtual avatars more or less transparent,

transparency varies as a function of color (HSL) and therefore skin tone. Regardless of ambient lighting, virtual humans depicting White individuals will be rendered as more opaque than virtual humans depicting Black individuals. Assuming that the same dehumanization effects transfer from static images to OST-HMDS, Black avatars will effectively be perceived as less human compared to their White counterparts.

Dehumanization has been shown to contribute to antipathy and violence across many different domains. For example, dehumanization of women is associated with men's willingness to rape and sexually harass women [141], dehumanization of Japanese and Haitian victims of natural disasters predicts less willingness to provide aid [5], and Christian participants' dehumanization of Muslims predicts increased willingness to torture Muslim prisoners of war [157]. On the positive side, perception of humanness is thought to be a powerful social cognition that can contribute to ameliorating racial prejudice [2,3]. Even considering across species, research shows that highlighting the similarity of animals to humans (i.e., humanizing animals) increases moral concern for the welfare of animals, as well as moral concern for the welfare of marginalized human groups [7]. The importance of humanness for intergroup relations is clear, suggesting that the selective alteration of humanness through differential opacity rendering of certain racial groups by OST-HMDs is an issue worthy of continued study.

Although we initially included measures of emotion to control for potential variation in facial expression across the avatar and human images, we observed interesting interactive patterns of emotion and opacity in our data. Among human faces, anger tended to be associated with lower perceptions of humanness, particularly at lower levels of opacity. As opacity decreased, the slope between anger and dehumanization increased for both the White and Black faces. Further exacerbating the issue, Black Americans are viewed as angrier than White Americans [73, 83] a pattern that was replicated in our data. The Black human was perceived as angrier than the other humans which has a greater dehumanizing affect at lower opacity levels (see figure 5.6). Current display

technology is limited to displaying virtual humans at the lowest opacity level we tested (35 Opacity); assuming that the results transfer to OST-HMDs, this suggests that Black virtual humans may be more dehumanized than virtual humans with lighter skin-tones.

We also found that the relationship between ratings of happiness and ratings of humanness varied as a function of opacity. For most levels, we observed a negative relationship such that happier-looking avatars appeared less human; however, this pattern was reversed for the most transparent avatars. Altogether, these exploratory findings regarding emotionality indicate that researchers and developers of AR technology need to consider how emotion is created within the virtual world. In the present study we utilized static images of faces. Cross-cultural research demonstrates remarkable consistency in the ability to detect and portray basic emotions through facial expressions [36]. However, emotion can be conveyed through a variety of methods, including voice, postural shifts, and gestures; an understanding of these factors leads to better development of embodied conversational agents or virtual avatars [109]. The findings of the present study suggest that AR researchers and developers should consider opacity when designing virtual agents meant to convey particular emotions.

5.2.3.1 Implications

AR is already in use in a variety of real-world applications, from medical training, to vehicle repair, to educational settings. Our results demonstrated that perceived humanness of static images was affected by opacity. Future work is needed to demonstrate that these results transfer to OST-HMDs. Assuming they transfer, the implications of opacity differences become important. Consider that a physician may use AR technology to assist in visualizing a patient's anatomy during surgery [59]. If the patient is Black and therefore their skin tone renders as partially transparent, results of the present study suggest that they may be viewed as less human. This perception could potentially

exacerbate pre-existing tendencies to view Black individuals as having higher pain tolerances than other racial groups [152], resulting in less pain medicine being administered.

In the present study, we observed that perceived humanness was negatively correlated with perceived dangerousness and positively associated with perceived competence. Though exploratory, these relationships could suggest clear implications for real-world AR use. AR and VR technology is already being used by, and marketed toward, military and police for use in both training and live patrol. In the United States, Black men are associated with danger [33], and Black suspects are disproportionately shot and killed by police [144]. AR training applications that render Black virtual humans as more transparent than White virtual humans may reinforce military and police trainees' perceptions of danger through lowered humanness. In a lower stakes, yet still important, use of AR, colleagues may engage in AR remote collaborations [165]. If the virtual human representing a Black collaborator is more transparent, that may reduce perceptions of competence due to lowered humanness.

These implications are in need of direct testing. However, the results of the present study suggest multiple downstream consequences of dehumanization based on virtual human transparency in AR.

5.2.3.2 Limitations & Future Directions

The present study represents the first test of the effects of differential transparency on the perception of virtual humans' humanness. As such, all measures and materials were created for this research. The faces that were utilized were all young men, and only one specific face was used to represent each human/avatar \times race category. Additionally, the Zombie human face was likely identifiable as a White man with zombie makeup. Future research should test the effects of opacity on multiple virtual human faces from each racial category as well as varying identities and

varying representations (e.g., full body, interactive). Future research should also consider whether user identity characteristics impact perceptions based on transparency including cross-gender and cross-race interactions; VR researchers have called for increased research on diverse representation [132] that should apply to both users and avatars.

It is also important to consider the measurement of humanness. In the present study we combined multiple facets of humanness such as the experience of pain or the possession of emotions into one overarching measure of humanness. Measuring humanness in relation to the uncanny valley [79] especially concerning avatars should be further investigated. There may be important distinctions between these factors that have differential implications based on opacity. Future research should operationalize humanness in varied ways. A critical next step is to go beyond self-report measures of humanness and assess behavior directly, possibly including social presence [76]. Future research could investigate if opacity affects use of force among police officers, physician behavior toward virtual humans, or workplace team outcomes such as work product quality or efficiency. Directly assessing the behavioral implications of opacity and race within virtual humans in AR will determine the importance and urgency of developers creating designs to counteract these effects.

Since a single testing environment was used in the study, we cannot produce a model of how the results of the study would change with respect to factors such as background color and environment illuminance. However, previous research has already examined the effects that such factors have on user perception in OST AR displays [42, 64]. Based on the existing research, increasing environment illuminance or lightening the background color would have the effect of reducing the contrast between the background and virtual human in the image, thus appearing more transparent, which would likely result in the images overall being rated as less human. Decreasing environment luminance or darkening the background color would have the opposite effect, increasing contrast and thus reducing transparency, likely producing results that were overall rated as being more human. Changing the hue of the background color would in turn shift the hues that comprise the

virtual human towards that of the background color. This would result in virtual humans that appear more red, green, or blue than intended. Since it is not usual to see virtual humans of such hues, it is likely that overall the results would shift and be rated as less human overall. Future work could verify that this is the case by incorporating environment illuminance and background color as independent variables in a similar user study involving virtual humans.

The varying levels of opacity used in the study conditions show interesting implications for future work in AR displays, in that even if the contrast between the virtual imagery and the user's physical environment is increased, the imagery will still appear transparent due to the issue of color blending between virtual imagery and the user's physical environment. While progress is being made to understand and resolve the limitation of color blending [64,66], we are likely years away from being able to adjust the coloration of virtual imagery in real time in order to correct for its effects. Further, even when color blending is resolved and solutions are integrated into consumer AR displays, we will still be faced with the issue that OST displays cannot present imagery that is darker than the user's physical environment without blocking incoming light from the environment. This means that in certain conditions, such as bright settings or settings with high lightness colors, the appearance of a dark virtual image may still appear transparent. Research into techniques to correct this issue for OST displays is still at an early stage. This is the case for both OST *head-mounted* displays as well as *heads-up* displays, such as those integrated into a driver's windshield in the automotive industry [65]. As a result, it seems likely that the implications of the study presented here will continue to remain a relevant problem for the foreseeable future.

5.3 Subjective Qualities of Self-Avatar Appearance in OST AR

In this study, we task users to use an OST AR display to create avatars that closely resemble themselves in two different environment lighting conditions.

5.3.1 Experiment

5.3.1.1 Participants

For this study, we recruited 20 participants: 13 male and 7 female, aged 18 to 55. Four of them self-identified as Black/African American, four as Asian, and twelve as Caucasian. The participants were students or members of the local university community. All participants were in STEM disciplines. The participants had normal or corrected-to-normal vision. When asked to assess their AR experience, five said they had none, nine said they had some, and six said they had a lot.

5.3.1.2 Materials

The physical setup is shown in figure 5.9. We performed our experiment in a $2.1\text{ m} \times 2.1\text{ m}$ isolated room.

Participants wore a Microsoft HoloLens 2 for the AR visual stimulus presentation (see figure 5.9). The HoloLens 2 is an OST-HMD with a field of view of circa 54 degrees diagonally, a resolution of 47 pixels per degree of sight, and a refresh rate of 120 Hz. We used a UV CleanBox⁴ to sanitize the equipment between use.

We placed a $0.6\text{ m (wide)} \times 1.8\text{ m (high)}$ mirror in the experimental room in the center of one of the walls. By standing on a marked position slightly off-center in front of the mirror, participants could see their own reflection, while at the same time their avatar was presented via the OST-HMD in the mirror as if it was standing next to them in the experimental room.

The visual stimuli are shown in figure 5.10. The environmental light in the experimental room

⁴<https://cleanboxtech.com>



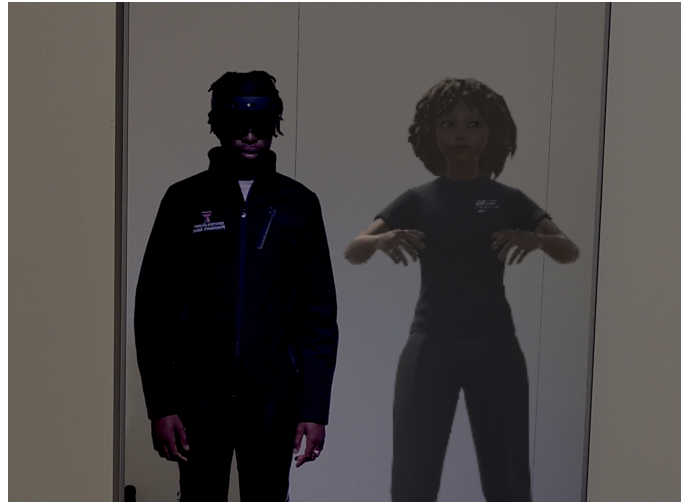
Figure 5.9: Experimental setup: Person standing on a marked location in the experimental room in front of the mirror wearing a HoloLens 2.

could be varied. We used an Urceri MT-912 light meter to calibrate two environmental lighting levels (200 lux and 2,000 lux). This light meter is reported to have an accuracy of $\pm 3\%$ of the measured value, and can make measurements between one and 200,000 lux, as reported by the Urceri website⁵.

The AR avatars were created in real time using the ReadyPlayerMe application and Unity API developed by Timmu Töke⁶. Skin colors and hair colors were adjusted by participants by requesting relative changes from their experimenter, i.e., increasing or decreasing the lightness of the colors using the palette shown in figure 5.10.

⁵<https://www.urceri.com/mt-912-light-meter.html>

⁶<https://readyplayer.me>



(a)



(b)



(c)



(d)

Figure 5.10: Experimental stimuli: Screenshot of a person's view through the HoloLens 2 taken with Microsoft's HoloLens Mixed Reality Live Preview (i.e., close but not a completely accurate representation of participants' view) in the (a) 200 lux condition and (b) 2,000 lux condition, and the color palettes for (c) skin color and (d) hair color.

5.3.1.3 *Methods*

In this experiment, we used a mixed design with one within-subject factor and two between-subject factors. Our within-subject factor was *environment lighting*, which had two levels: The amount of environment light was either 200 lux, which represents the amount of light in a common dark indoor office environment, or 2,000 lux, which represents dim outdoor lighting such as on a cloudy day. We further considered the between-subject factor *skin color*, which, due to the limited diversity of our participant sample, we simplified for our analysis to just two levels: dark and light. Similarly, we considered the between-subjects factor *gender*, which we again simplified to just two levels: female and male. Each participant completed all conditions in random order.

5.3.1.4 *Measures*

In this experiment, we had two main measures:

- *Selected skin color*: When matching the appearance of their avatar to their own reflection in the mirror, participants selected the skin color of the avatar on a ten-level scale (1=dark to 10=light; see figure 5.10 c).
- *Selected hair color*: Participants further selected the hair color of the avatar on a ten-level scale (1=dark to 10=light; see figure 5.10 d).

5.3.1.5 *Procedure*

Prior to the experiment trials, participants gave their informed consent. Participants then received a brief overview of what AR displays are, as well as what their task in this experiment will be. We

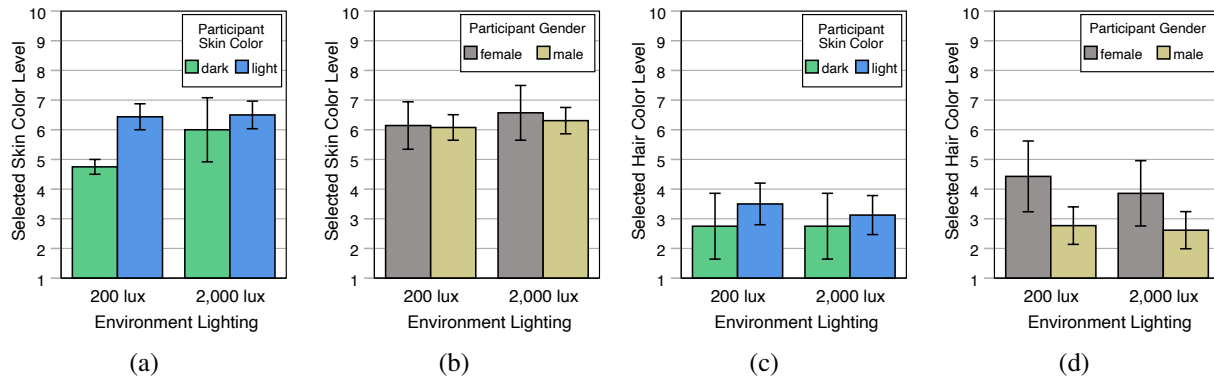


Figure 5.11: Bar charts showing effects of environment lighting on avatar (a+b) skin colors and (c+d) hair colors on ten-level scales (1=dark to 10=light). Plots (a+c) show results for our participants' between-subject factor skin colors, and plots (b+d) show them for participants' gender.

then took a picture of the participants to auto-generate an approximate avatar look-alike using the ReadyPlayerMe software.

Participants were then asked to stand in front of the mirror at the marked location on the floor. Once participants donned the HoloLens 2, an AR avatar appeared next to them in the mirror. Participants were then instructed to adjust the appearance of their avatar. The participants could fine-tune the appearance of their avatar as supported by the ReadyPlayerMe software, e.g., by choosing its clothes and hair style.

The main conditions of the experiment were tested in random order, where the environment lighting was either set to 200 lux or 2,000 lux. In each of these lighting conditions, the participants were asked to adjust their skin color and their hair color to make their avatar match their own reflection in the mirror. The initial skin/hair colors for these trials were randomized between the maximum or minimum lightness levels on the ten-level palette.

5.3.2 Results

Figure 5.11 shows our results for participants' selected skin color and hair color levels for our within- and between-subject variables.

We analyzed participants' selections on the skin color and hair color scales using parametric tests at the five percent significance level, after testing for all assumptions of these tests. For independent variables, we considered the two between-subject variables participants' *skin color* and *gender*, and the within-subject variable *environment lighting*. We only report the significant results.

We first looked at the effects of our between-subject variables using one-way ANOVAs. For participants' *skin color*, we found a trend for the *selected skin color levels*, $F(1, 38) = 2.51$, $p = 0.12$, $\eta_p^2 = 0.06$, and no effect on the *selected hair color levels*, $F(1, 38) = 0.31$, $p = 0.58$, $\eta_p^2 = 0.01$. For participants' *gender*, we found a trend for the *selected hair color levels*, $F(1, 38) = 3.01$, $p = 0.09$, $\eta_p^2 = 0.08$, and no effect on the *selected skin color levels*, $F(1, 38) = 0.08$, $p = 0.78$, $\eta_p^2 = 0.002$. Considering the trends together with the limited between-subject samples in our experiment, which suggest that the statistical power may not have been high enough to show significant effects, we decided to perform our further analysis by modeling these between-subject factors separately.

First, looking at *skin* colors, we analyzed our results with a two-way mixed ANOVA model with the within-subject factor *environment lighting* and participants' between-subject *skin color*. We observed a trend though no significant interaction effect between *environment lighting* and participants' *skin color* on the *selected skin color levels*, $F(1, 18) = 4.13$, $p = 0.057$, $\eta_p^2 = 0.19$. We found a significant main effect of *environment lighting* on the *selected skin color levels*, $F(1, 18) = 5.04$, $p = 0.038$, $\eta_p^2 = 0.22$.

Second, looking at *hair* colors, we analyzed our results with a two-way mixed ANOVA model with the within-subject factor *environment lighting* and participants' between-subject *gender*. We

observed no significant interaction effect between *environment lighting* and participants' *gender* on the *selected hair color levels*, $F(1, 18) = 1.52$, $p = 0.23$, $\eta_p^2 = 0.08$. We found a significant main effect of *environment lighting* on the *selected hair color levels*, $F(1, 18) = 4.58$, $p = 0.046$, $\eta_p^2 = 0.20$.

5.3.3 Discussion

Overall, our results give interesting insights into the limitations of OST-HMD technologies. In particular, our results indicate that for some participants the amount of light present in the physical environment affected how they perceived their own avatar in AR, causing them to adjust their avatar's appearance.

When asked to match their avatar's skin color to their own, we found a significant effect of environment lighting on the skin colors participants selected when creating their avatar. Our results (figure 5.11 a) indicate that participants with a dark skin color selected a dark skin color for their avatar when the physical environment was dark (200 lux), but they had to make their avatar's skin color lighter when the physical environment was well-lit (2,000 lux). On average, their selected avatar skin colors in the well-lit physical environment were very close to those selected by participants with a light skin color. In contrast, we did not observe such an effect for participants with a light skin color—their avatar's skin color remained largely the same between the two physical lighting levels.

This effect may be explained by a major limitation of current OST display technologies, including the HoloLens 2 we used in our experiment. With OST displays based on the additive light model, light can be added to the user's view of their physical environment, but not reduced. Hence, it is not possible to present anything on an OST display that is darker than the physical background seen through the display. If a participant's skin color was darker than the physical background,

they had the choice to either appear “transparent” [131] or to increase their skin color to improve contrast between their avatar and their physical environment. While the latter ensured that they would be visible on the display, this caused the appearance of all participants’ avatars to converge on the same (light) skin colors, independently of what skin color they really have. These results are concerning, considering that the display technology may greatly reduce the diversity of avatars presented on those displays in future applications where multiple users and avatars occupy the same AR space.

We observed another interesting effect when we asked participants to match their avatar’s hair color to their own. Our results (figure 5.11 d) indicate that in particular female participants gave themselves a darker hair color in the well-lit environment compared to the dark environment. In contrast to our results for skin colors, the selected hair colors overall were comparatively dark.

As the direction of the effect for hair color is opposite to the effect for skin color, this effect could potentially be explained by participants trying to maximize the light differential between their skin color and their hair color. As for skin colors, when asked to match their appearance, participants had the choice to either have dark colors appear transparent or to increase their lightness. With participants’ hair colors starting off as comparatively dark, a slight increase in lightness would not have effected a major change in transparency. It stands to reason that participants would not have given their avatar a very light hair color (e.g., blonde) just to make their hair more visible on the display. Instead, it appears that they rather tried to make it as dark as possible (i.e., transparent) considering the amount of environment light defining a natural lower bound to its appearance. That this effect was more pronounced for our female participants may be due to women often having longer hair than men, filling a larger portion of their view on the display.

5.3.3.1 Implications

As OST displays begin to reach consumers for general usage, users will find themselves in many different environments ranging from dim nighttime or interior lighting conditions to outdoors in direct sunlight. Until the displays are capable of automatically adjusting factors, such as luminance output and attenuation, to maintain the perceived appearance of virtual imagery in different environment conditions, virtual imagery, including the user's representation as a virtual avatar, will need to be able to adapt appropriately to best maintain the user's desired representation. Existing AR applications involving avatars typically use a one-off avatar creation process that does not consider how user perception of the avatar will change according to the observer's viewing conditions. Our results indicate that users may wish to have the option to manually specify how they will be represented by their avatar for different viewing conditions, or may wish to have the system automatically adjust the appearance of their avatar, potentially interpolating between different user created or system generated versions of their avatar.

5.3.3.2 Limitations

Our experiment has different limitations. First, our participant sample ($N=20$) was not large enough to provide us the required statistical power to show all between-subject effects. Future work should focus on very large and diverse participant samples to elucidate the underlying demographic and appearance-related factors. Second, our results are specific to the HoloLens 2 HMD. While current developments among commercial OST-HMDs aim to increase the maximum luminance of the display, this will not change the underlying problem that colors darker than the user's environment cannot be displayed due to the additive light model. Future work should focus on the development of prototypical mechanisms such as pixel-wise light subtraction/attenuation for OST-HMDs (e.g., [85]). Third, our experiment was limited to only two levels of environment lighting (200 lux and

2,000 lux). Future work should consider dynamic ranges from dark indoor environments to sunny outdoor lighting, which can reach upwards of 100,000 lux, which would likely amplify the issues observed in our experiment.

5.4 Summary

This chapter investigated users' subjective perception of virtual avatars and images of real people shown on OST AR displays. As previously demonstrated in chapter 3, changing the environment lighting conditions in the user's environment affects the user's perception of the virtual imagery, in this case resulting in decreased perceptions of humanness of virtual humans when lighting conditions are increased. This result supports thesis statement one, by demonstrating that users tend to misperceive the virtual humans shown on the OST display when bright lighting conditions cause the imagery to appear transparent. Together, chapters 3 through 5 have demonstrated several similar effects, where in brighter environment lighting conditions the user's AR experience is negatively impacted in several different manners, including reducing the legibility and usability of UIs, potentially decreasing the effectiveness of attention cues, and reducing the perceived humanness of virtual humans.

In study two, users chose to adapt the appearance of their own avatars in response to changes to their viewing conditions. This demonstrated that users were aware that the perceived appearance of their avatar had shifted due to the change in lighting conditions. By choosing to adapt their avatar's appearance, they are providing support for the idea that AR systems should be able to automatically adapt to maintain the intended appearance of the virtual imagery. Such a system could be used to mitigate the negative effect observed in these chapters including, improving the usability of UIs, improving the legibility of text, improving the effectiveness of attention cues, and increasing the perceived humanness of virtual humans shown on the OST AR display.

In the next chapter, I lay the groundwork for future adaptive AR systems by demonstrating how the viewing conditions of a user of an OST AR display are related in terms of luminance or illuminance contrast, and presenting an extended contrast model capable of describing the user's AR experience.

CHAPTER 6: ESTABLISHING A MODEL

In the previous three chapters, I demonstrated that the user's environment appearance, environment lighting conditions, and the intended appearance of AR imagery each significantly affect the perceived appearance of imagery shown on the AR display. Together, these three terms describe the user's *viewing conditions* at a particular point in time. I additionally demonstrated that changes to any of these factors (changes to a user's viewing conditions) can have negative effects on the user's AR experience, such as by decreasing the legibility of text in UIs, decreasing the effectiveness of attention cues, and decreasing perceived humanness in virtual humans and avatars. In order to predict when the user is likely to experience these effects, and to inform future AR systems that adapt to maintain or improve the perceived appearance of virtual imagery, it would be useful to have a model capable of quantifying the perceived appearance of AR imagery, given parameters that describe the user's viewing conditions.

In this chapter, I demonstrate that the user's environment, their AR display, and the imagery shown on the display are relatable when described in *photometric terms* [150], such as luminance or illuminance. I also demonstrate that existing models used to describe an observer's perception of visual stimuli, such as Michelson contrast [122], can be extended to include these terms for usage with OST AR displays. This extended model, based on Michelson contrast, is the primary contribution of this chapter.

Following the derivation and introduction to the new extended model, I demonstrate how photometric measurements can be made on the AR display and within the user's environment to characterize these factors for use with the model. I also discuss several manners in which the intended appearance of virtual imagery shown on the AR display can be quantified for use with the model. Finally, I present a user study that tasks participants to perform a visual search task on an arrangement

of letters in varying levels of contrast, manipulated through changes to the lighting conditions in the environment and through changes to the intended color of virtual imagery shown on the AR display. Through the results of this study, I demonstrate that user performance in the search task, confidence in their responses, and their perception of the task difficulty can each be accurately predicted according to the luminance contrast of the study stimuli, as measured by my extended contrast model.

Portions of this chapter have been previously published in the following papers:

- [46] Austin Erickson, Kangsoo Kim, Gerd Bruder, and Gregory F. Welch. 2020. Exploring the Limitations of Environment Lighting on Optical See-Through Head-Mounted Displays. In Proceedings of the 2020 ACM Symposium on Spatial User Interaction (SUI '20), pages 1–8, 2020.
- [39] Austin Erickson, Gerd Bruder, and Greg Welch. Adapting Michelson Contrast for use with Optical See-Through Displays. In 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), pages 409-410, 2022.

The work in this chapter is primarily used to support thesis statement two, however the results of the user study presented in this chapter are also used to partially support thesis statement one:

- **TS1 - Identify Problems:** In increasingly common viewing conditions, AR users are susceptible to misperceiving or completely missing information shown on their AR display. Sometimes the user may not even be aware that their AR experience or their performance is being affected.
- **TS2 - Establish a Model:** The discrepancy between the intended and perceived appearance of AR imagery can be modeled in terms of luminance contrast, as a function of parameters

specific to the user's viewing conditions. These parameters include the user's environment lighting conditions, the characteristics of their AR display, and the characteristics of the AR imagery.

6.1 Modeling Contrast for OST Displays

In this section, I derive an extended version of Michelson's contrast model for use with OST AR displays. The extended model outputs contrast values describing the relationship between the AR imagery and the user's environment. The model is a function of the user's viewing conditions, and takes parameters that describe the user's environment, the characteristics of their AR display, and the intended appearance of the AR imagery being shown. In general, the contrast values obtained from the model partially describe how easily a particular image can be distinguished by the user. For contrast values below a particular threshold (determined by the contrast sensitivity of the observer), the user will not be able to recognize the presence of the image on the display. As contrast increases above this threshold, the difficulty the user will experience in attempting to distinguish the image is reduced.

For OST AR displays, AR imagery blends in appearance with the user's environment, and contrast can be measured to describe the perceived appearance of the user's physical-virtual environment (see figure 6.1). For example, a contrast measurement could be used to describe how apparent an image shown on the display is compared to the user's view of their physical environment (*physical-virtual contrast*). Additionally, contrast measurements could be used to describe how apparent a particular feature within a virtual image is compared to its surroundings in the image, while considering how the virtual imagery blends with the user's physical environment (*virtual contrast*). Finally, contrast measurements could also be used to describe how apparent a particular feature or region of the user's physical environment is compared to its surroundings (*physical contrast*).

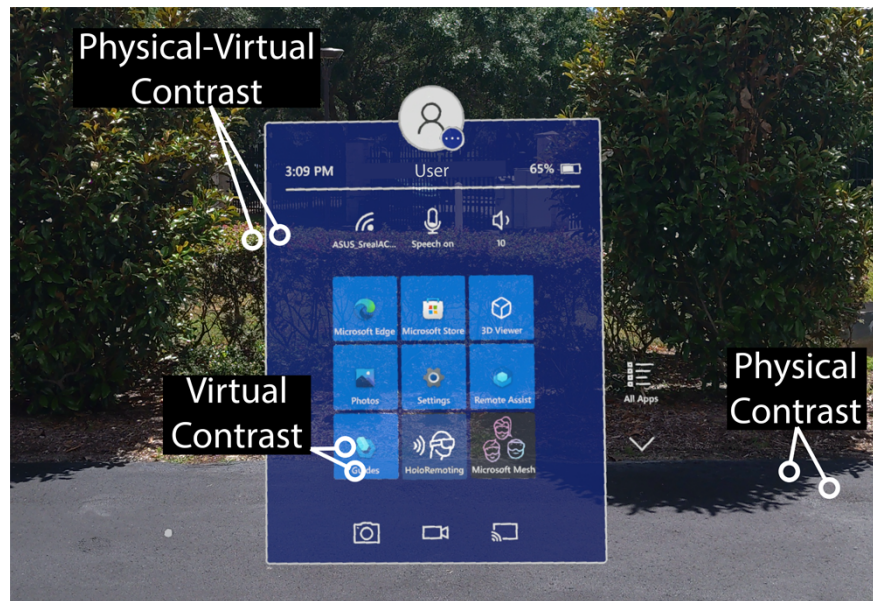


Figure 6.1: Three types of contrast comparisons can be made from the observer’s perspective within an OST AR display. Physical-virtual contrast compares between a point in the AR imagery and a point in the observer’s physical environment, in this case between the floating panel and the bushes behind it. Virtual contrast compares between two points within the AR imagery, in this case between the white text and blue background. Finally, physical contrast compares between two points in the observer’s physical background.

Examples of all three of these cases can be seen in figure 6.1. The extended contrast model derived in this section is primarily used to calculate physical-virtual contrast, and virtual contrast. Following these definitions, traditional contrast models are typically used for physical contrast as they are not typically considering how virtual imagery blends with respect to the observer’s physical environment.

This chapter makes heavy use of photometric terminology, particularly luminance, illuminance, and contrast. If unfamiliar with these terms, I recommend reading their corresponding section of chapter 2. Typically, contrast is calculated using parameters specified in units of luminance, and since contrast is expressed a ratio, output values are unitless. However, due to the relationship between luminance and illuminance, contrast can be calculated in terms of either. The choice to

derive the extended contrast model in terms of luminance or illuminance has several potential trade offs that will be discussed later in section 6.1.2. However, one advantage of the illuminance-based model derived in this section, is that it can also be easily used with input terms in units of luminance (while the same is not necessarily true for illuminance units used in a luminance-based model). In this section, I present the equations that describe the contrast model and demonstrate how it can be used with units of either illuminance or luminance. To show how such a model is derived, first examine Michelson's contrast equation:

$$\textbf{Michelson Contrast: } \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \quad (6.1)$$

Michelson contrast compares two terms based on the perceived intensity of light at two specific positions, when used in terms of luminance. A simple example of this would involve calculating the contrast of the text in this dissertation, where the I_{max} term would be the luminance of the white background behind this text, and the I_{min} term would be the luminance at the position of one of the black letters. However, when used in terms of illuminance, the equation can also be used to compare the perceived intensity of light reaching the user's eye from multiple light sources, such as the perceived intensity of light emitted by the display compared to the perceived intensity of light originating from other sources within the user's environment. In such a scenario, the I_{min} term is the illuminance at the user's eye position within the AR display, while the display is powered off or otherwise not emitting light. The I_{max} term is the illuminance at the user's eye position within the AR display, while the display is rendering AR imagery. If the display is not rendering AR imagery, the I_{min} and I_{max} terms are equal, and the numerator of the equation becomes zero, thus making contrast zero. This trivial case shows that the user cannot see any AR imagery, since nothing is being shown on the display.

Since OST AR displays are additive, the I_{max} term for the equation can be considered as the sum

of two illuminance values: one value solely characterizing the light emitted by the AR display, and one value characterizing the light at the user's eye position from other sources in their environment. In this case, the former term can be measured as an illuminance value from the user's eye position in a pitch dark environment with the AR display rendering the imagery of interest. The latter term is the same as the I_{min} term previously described above, and can be measured from the user's eye position when the display is off. In this case, Michelson's contrast equation becomes:

$$I_{max} = I_{img} + I_{env} \quad (6.2)$$

$$I_{min} = I_{env} \quad (6.3)$$

Substituting equations 6.2 and 6.3 into 6.1:

$$\text{Global Michelson Illuminance Contrast: } \frac{I_{img}}{I_{img} + 2I_{env}} \quad (6.4)$$

In general, this equation forms the basis for the extended contrast model for OST AR display, but in this form it is limited in several ways. While it is somewhat straightforward to establish an illuminance value for the I_{env} term, for instance it could be looked up from a table of common lighting conditions, it is less straightforward to establish the I_{img} term, which is different for every potential image shown on the AR display of interest. It would be more helpful to be able to predict the value of this I_{img} term by introducing several additional parameters specific to the AR display of interest, which I will demonstrate shortly. Also, in this form, the contrast value is global, in that it is outputting a single contrast value that describes the perceived appearance of entire AR image being displayed based on the user's view of their physical environment behind it. In this form, two points within a single AR image cannot be easily compared, such as the contrast between virtual

text and a virtual background color shown on the AR display. For these reasons, I continue to expand this equation throughout this section.

To begin with, I introduce two terms related to the AR imagery being displayed, I_a and I_b . These terms represent two separate illuminance measurements at the observer's eye position within the OST display where the entirety of the display's field of view is rendered in color a and then in color b respectively. Similar to the above equation, these illuminance values are summed with the I_{env} term to factor in how the light from the display blends with the ambient environment light. Additionally, max and min functions are introduced since a and b are arbitrary colors, and the color with the higher luminance is not always intuitively known.

$$I_{max} = \max(I_a, I_b) + I_{env} \quad (6.5)$$

$$I_{min} = \min(I_a, I_b) + I_{env} \quad (6.6)$$

By substituting equations (6.5) and (6.6) into equation (6.1) and simplifying, I obtain the *local* version of the contrast equation shown below, which calculates contrast between two colors shown on the AR display, considering the user's environment lighting conditions. Notice that, for example color a could be black, in which case I_a is equal to zero since no light is emitted by the AR display, and thus the equation simplifies into the form previously shown in equation 6.4. The same thing would occur if color b were set to black. Thus, this version of the contrast equation can be used to calculate contrast between any two colors on the AR display, or to calculate global contrast between the entire image and the user's view of their environment.

$$\text{Local Michelson Contrast: } \frac{\max(I_a, I_b) - \min(I_a, I_b)}{I_a + I_b + 2I_{env}} \quad (6.7)$$

This expression can be simplified by introducing absolute values to eliminate the min and max expressions. Additionally, as mentioned previously in chapter 2, many OST displays have a tinted visor incorporated into their design to attenuate light from the user's environment as it blends with the light emitted by the display. To consider this attenuation of environment light, I introduce an attenuation factor, $k \in \mathbb{R} : 0 \leq k \leq 1$, which scales I_{env} :

$$C = \frac{|I_a - I_b|}{I_a + I_b + 2kI_{env}} \quad (6.8)$$

While certainly an improvement, this equation requires knowing I_a and I_b , the illuminance values of two arbitrary colors, a and b , shown on the display. Since these values are not usually known, I need to establish a function that outputs the illuminance value for an input color according to its RGB value, where $r, g, b \in \mathbb{R} : 0 \leq r, g, b \leq 1$. The illuminance of any color shown on the display is the sum of the illuminance occurring from each of the primary color channels of the display, which each have maximum illuminance values of I_r , I_g , and I_b for the red, green, and blue color channels respectively. Therefore, it is intuitive to think that the function would be a linear combination of maximum illuminance values and RGB values in the following form:

$$I(r, g, b) = I_r r + I_g g + I_b b \quad (6.9)$$

However, in practice, gamma correction is typically applied in the form of a transfer function so that step-wise changes in RGB values occur in perceptually equal intervals of lightness [11]. This correction is typically in the form of a power function I will call G:

$$G(x) = ((x + c_1)/c_2)^\gamma \quad (6.10)$$

The c_1 , c_2 , and γ values are defined according to the color and image encoding standards being used by the specific display, such as standard red, green, blue (sRGB) or the BT.709 standard [4, 139]. Therefore, to factor gamma correction into the equation, I pass r , g , and b through the function G :

$$I(r, g, b) = I_r G(r) + I_g G(g) + I_b G(b) \quad (6.11)$$

This equation can be simplified using vector notation, as a scalar product between two vectors, where the maximum illuminance of the primary color channels are expressed as a row vector \mathbf{a} , where $\mathbf{a} = [I_r, I_g, I_b]$, and the gamma corrected RGB values are specified as a row vector \mathbf{x} , where $\mathbf{x} = [G(r), G(g), G(b)] = [r', g', b']$. Here, T is used to represent the transpose of the vector it is applied to.

$$I(\mathbf{x}) = \mathbf{a}\mathbf{x}^T \quad (6.12)$$

This equation can be substituted back into equation (6.8) to yield a new equation of our extended model of illuminance contrast for OST displays:

$$C = \frac{|\mathbf{a}(\mathbf{x} - \mathbf{y})^T|}{\mathbf{a}(\mathbf{x} + \mathbf{y})^T + 2kI_{env}} \quad (6.13)$$

The different parameters of the equation are defined as follows:

- $\mathbf{a} = [I_r, I_g, I_b]$: A row vector representing the illuminance from the observer's eye position

of the red, green, and blue color channels respectively, at maximum intensity and in a pitch dark environment;

- $\mathbf{x}, \mathbf{y} = [G(r), G(g), G(b)] = [r', g', b']$: Row vectors containing the gamma corrected RGB values, r' , g' , and b' , of the two colors being compared;
- k : The attenuation factor of the display;
- I_{env} : The illuminance from the observer's eye position contributed by the lighting in the observer's physical environment.

The final form of the equation can be used to establish contrast values between any pair of RGB color coordinates in the device's color space. Or, as mentioned above, equation 6.8 can be used to establish a global contrast value for an entire image, where the I_a term is set to I_{img} , the illuminance at the user's eye position when the AR image is displayed in a pitch dark setting, and I_b is set to zero.

6.1.1 Use with Luminance Units

As briefly mentioned at the beginning of this section, one advantage of deriving the extended contrast model using units of illuminance is that the resulting equations can easily be used with terms of either luminance or illuminance (but not a combination of the two). To do this, the input terms for the model are similar:

- $\mathbf{a} = [I_r, I_g, I_b]$: A row vector representing the luminance of the red, green, and blue color channels of the AR display respectively, at maximum intensity as measured in a pitch dark environment;

- $\mathbf{x}, \mathbf{y} = [G(r), G(g), G(b)] = [r', g', b']$: Row vectors containing the gamma corrected RGB values, r' , g' , and b' , of the two colors being compared;
- k : The attenuation factor of the display;
- I_{env} : The average scene luminance from the observer's eye position, or a reference luminance describing a particular point of comparison in the observer's physical environment.

6.1.2 Limitations of the Model

As briefly introduced at the beginning of this section, there are trade-offs to using model parameters in illuminance units versus in luminance units. To begin with, illuminance is a more accurate description of the I_{env} term, which describes the light at the user's eye position originating from light sources with their physical environment (other than the AR display). Within the AR display, light is reaching the user's eyes from many different incoming angles, and illuminance considers the cumulative effect of the lighting at the user's eye position. This term is arguably the most important one in the contrast model, as it is used to represent the light adaptation state of the user, which in turn defines how bright AR imagery needs to be to be easily perceivable in that particular environment condition. However, in the illuminance-based model the I_a term is modeled as if the color a was rendered to fill the entire field of view of the AR display, and I_b is modeled similarly for color b . In practice, imagery shown on the AR display consists of many different colors, and there are few cases where a single color would fill the display's field of view in such a manner. Because of this, it is possible that this version of the model would be less accurate at establishing contrast values between two specific regions of the display compared to the luminance-based model, whereas it would be more accurate at establishing global contrast values where the I_{img} term is known.

If I were to switch to a luminance-based version of this contrast model, it would be difficult to characterize the I_{env} term representing the user's environment without making repeated luminance measurements at varying angles for each pixel position on the display. To get around this, a single luminance value could be obtained for each pixel on the display representing the luminance of the user's environment from the user's perspective within that particular region of the display. This approach is less accurate than the illuminance-based approach, as it assumes incoming light from the user's environment is traveling in one particular direction, defined by the vector between the user's eye position and the position of each pixel on the display. However, the luminance-based version of the contrast model has an advantage in the characterization of the I_a and I_b terms of the model. In fact, it's relatively straightforward to establish I_{img} from the global model described above, since the luminance for each pixel in the display's output is known, or is at least estimated, by the display hardware. From the I_{img} term, any pair of luminance values, or even any two regions of the display, can be easily compared.

It is unclear as to which version of the model is "better," and it could likely be that these two variations are equally useful, but for different applications. This particular research question is not addressed in the scope of this dissertation, and I encourage exploration of this issue in future work.

As for other limitations, the model makes several assumptions about the displays it can be used upon. It assumes that adaptive lighting techniques are not already employed by the device to change the intensity or color of imagery shown on the display in response to changing environment lighting conditions (i.e. the intensity of a color displayed in a pitch dark environment is the same as the intensity of the color in a bright environment).

Additionally, I assume that any reduction in environment lighting as it blends with the display lighting, such as due to a tinted visor, happens in a uniform manner across the entire image and is consistent regardless of the intensity of the environment light (i.e. incoming environment light

is reduced by a scale factor due to a tinted visor, and this scale factor is the same for different environment lighting conditions). If this is not the case, the k term would have to be extended or adapted to describe the non-linear effect.

AR displays are also particularly prone to optical mura and non-uniformities in the appearance of imagery shown on the display [106]. These are largely determined by the design of the particular AR display in use, but even then minor imperfections in the mass manufacturing of AR displays could mean that no two displays are identical in terms of optical imperfections. In future iterations of this contrast model, the effects of non-uniformities and other optical mura could potentially be introduced through additional terms to the model, for example in a luminance-based model a matrix could be used to characterize the loss of luminance at each pixel position due to such imperfections.

Finally, the extended contrast model assumes that the gamma function of the AR display is in the form described in equation 6.10. If the display were to use a different gamma function, then this would need to be measured and reflected in the model to remain accurate.

6.2 Photometric Measurements on Optical See-Through AR Displays

This section describes the controlled photometric measurements I made on one particular OST AR Display, the Microsoft HoloLens 2. These measurements demonstrate how to find the terms \mathbf{a} , I_{env} , and k as described in the previous section. While the measurements presented in this section are specific to the HoloLens 2, the same procedure can be repeated for any OST AR display of interest.

6.2.1 Apparatus

I measured illuminance values on the HoloLens 2, which was released in 2019, and features a diagonal field of view of 52 degrees and a resolution of 2048×1080 pixels per eye. All photometric measurements were taken with the HMD powered on and with the brightness control settings of the device turned to 100%. Smaller percentages of display brightness were not tested, since I am primarily concerned with the maximum contrast able to be achieved under each tested lighting condition.

The illuminance measurements were made using an Urceri MT-912 light meter, which is reported to have an accuracy of $\pm 3\%$ of the measured value, and can make measurements between one and 200,000 lux, as reported by the Urceri website¹.

I aimed for illuminance measurements distributed across a wide range of common environment lighting conditions from dim indoor or nighttime lighting levels to bright sunny outdoor lighting levels. To do this, I chose to use a single dimmable Amzcool natural light lamp as our light source, since it could be easily adjusted to create illuminance measurements between 0 to 10,000 lux depending on the distance between the light meter and the light source. The environment in which the measurements were made had a consistent ambient illuminance of less than 1 lux, with no light sources other than the dimmable natural light lamp to influence the measurements. Environment lighting conditions with the lamp powered off were verified to be less than 1 lux.

When performing direct illuminance measurements of the light source, the light meter was positioned to be 8 cm away from the light source and parallel to its surface (see figure 6.2). This distance was chosen because direct illuminance measures taken of the light source at maximum intensity from this distance corresponded to our desired upper limit of 10,000 lux. When taking

¹<https://www.urceri.com/mt-912-light-meter.html>



Figure 6.2: Illustration of the setup used to perform photometric measurements.

illuminance measurements through the display, the display was positioned so that the light meter was at the user's left eye position relative to the display. Relative to the light source, the light meter was in the same position as it was when making direct illuminance measurements of the light source (8 cm away and centered).

For virtual content presentation on the HoloLens 2, I used the Unity game engine. Imagery was streamed from a nearby laptop to the device via the holographic remoting feature. The laptop was positioned away from the measurement area to not affect the measurements. For each measurement, a single color was presented filling the entire field of view of the display. Colors were varied between 14 different RGB values representative of the primary secondary and tertiary colors as well as black and white (see table 6.1).

The primary, secondary, and tertiary colors were displayed at maximum saturation and value. The

Table 6.1: This table shows the RGB values for the primary, secondary, and tertiary colors measured on the HoloLens 2.

	Red	Green	Blue
Black	0	0	0
White	1	1	1
Red	1	0	0
Green	0	1	0
Blue	0	0	1
Yellow	1	1	0
Cyan	0	1	1
Magenta	1	0	1
Orange	1	0.5	0
Chartreuse	0.5	1	0
Spring Green	0	1	0.5
Azure	0	0.5	1
Violet	0.5	0	1
Rose	1	0	0.5

inclusion of the black and white conditions allow us to establish the maximum contrast for each environment lighting condition, while the colors sample even intervals within the display’s color space for single color channels (primary color), and combinations of two color channels (secondary and tertiary colors).

6.2.2 Methods

The illuminance measurements were made under the following controlled lighting conditions:

Environment Lighting (5 levels): The dimmer on the environment light source was varied to achieve illuminance values of 10,000, 1,000, 500, 100, and 0 lux. These light levels respectively correspond to environments with full daylight, overcast daylight, indoor working or retail environments, indoor transitory environments (e.g. hallways), and pitch black environments. I confirmed

the illuminance of each condition with the light meter as described in the next section, and the measured values were found to vary less than 1% away from the intended illuminance value.

Displayed Color (14 levels): The displayed color on the OST-HMD was varied between 14 different colors, including black (appears as transparent [48]), white, the primary colors (red, green, blue), the secondary colors (yellow, cyan, magenta), and the tertiary colors (orange, chartreuse, spring green, azure, violet, and rose). All primary, secondary, and tertiary colors were presented at maximum saturation and value.

6.2.3 Procedure

First, for each targeted environment lighting condition, the light meter was positioned to be 8 centimeters in front of the light source, and the light source was dimmed until the illuminance measurement was representative of the targeted illuminance value. Five illuminance measurements were then recorded of the light source for that environment lighting condition.

Second, the HoloLens 2 was positioned as described in section 6.2.1, and the color to be presented on the display was configured via Unity. A series of five sequential illuminance measurements were then recorded. Following this, the displayed color was changed to the next condition and the process was repeated.

Once all colors had been measured for a particular environment lighting condition, the light source was dimmed according to the next condition and the process was repeated until the measurements for each condition were recorded.

Table 6.2: This table shows the average *illuminance* values, in lux, for the environment, black, white, primary colors, secondary colors, and tertiary colors respectively on the HoloLens 2 across lighting conditions ranging from 0 to 10,000 lux. Color names are abbreviated to their first letter for the primary colors (red, green, and blue), as well as for the secondary colors (yellow, cyan, and magenta). The tertiary colors are abbreviated to two letters representing the primary and secondary colors they fall between in the HSV color space. These tertiary colors are respectively: orange (R-Y), chartreuse (Y-G), spring green (G-C), azure (C-B), violet (B-M), and rose (M-R).

	Black/White		Primary Colors			Secondary Colors			Tertiary Colors					
Env.	Blk	W	R	G	B	Y	C	M	R-Y	Y-G	G-C	C-B	B-M	M-R
0	0	174.9	34.9	96.1	31.3	137.0	134.0	65.7	57.7	110.6	106.0	49.0	36.2	36.8
100	27.0	222.5	61.3	130.5	60.5	182.1	164.6	96.1	92.1	142.3	131.6	84.3	68.6	70.2
500	125.5	275.6	148.7	203.0	154.0	221.7	217.0	166.8	153.0	177.4	176.9	157.0	147.0	141.2
1,000	286.0	473.0	322.9	372.9	318.7	394.0	399.7	324.7	304.6	357.2	352.5	320.5	307.4	307.6
10,000	2665	2720	2625	2778	2747	2761	2743	2687	2651	2667	2663	2677	2674	2637

Table 6.3: This table similarly shows the *Michelson contrast* values between each color and the black level of the display on the HoloLens 2.

	White	Primary Colors			Secondary Colors			Tertiary Colors					
Env.	W	R	G	B	Y	C	M	R-Y	Y-G	G-C	C-B	B-M	M-R
0	0.997	0.985	0.994	0.983	0.996	0.996	0.992	0.991	0.995	0.995	0.989	0.985	0.986
100	0.783	0.388	0.657	0.383	0.742	0.718	0.561	0.546	0.681	0.659	0.515	0.435	0.444
500	0.374	0.085	0.236	0.102	0.277	0.267	0.142	0.099	0.172	0.170	0.112	0.079	0.059
1,000	0.246	0.061	0.132	0.054	0.159	0.166	0.063	0.031	0.111	0.104	0.057	0.036	0.036
10,000	0.010	0.008	0.021	0.015	0.018	0.014	0.004	0.003	<0.001	<0.001	0.002	0.002	0.005

6.2.4 Results and Discussion

In this section, I present the data collected from the photometric measurements as illuminance values and as Michelson contrasts..

6.2.4.1 Illuminance Measurements

Table 6.2 and figure 6.3 depict the mean (M) illuminance measurements made on the HoloLens 2 according to the environment illuminance and displayed color. The results show how the illumi-

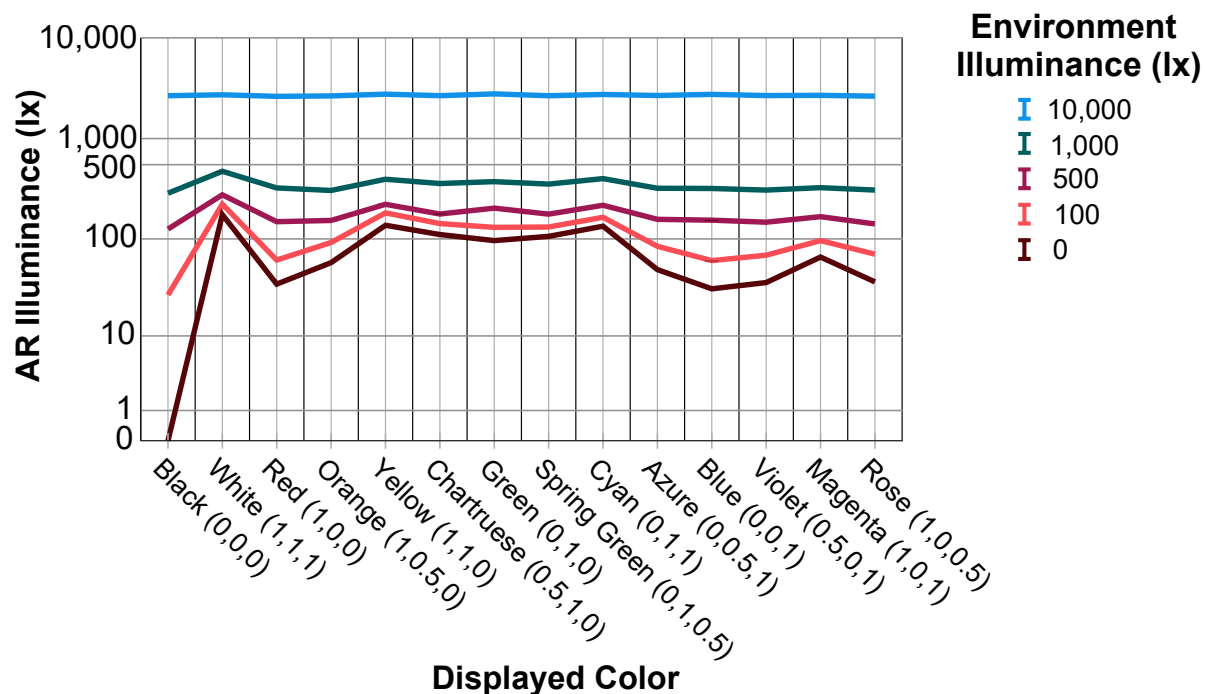


Figure 6.3: This figure depicts the mean illuminance values for each of the 14 tested colors at each of the 5 environment lighting conditions. The y -axis is the illuminance from the user's eye position in the AR display, and is presented in log scale, with one additional label added to depict 500 lux.

nance from the user's perspective changes with respect to the color of the light being emitted by the display and the intensity of the environment lighting. Due to illuminance measurements being weighed according to human color sensitivity, green-hued colors have higher illuminance values than either red or blue hued colors. This decrease in sensitivity is represented in the figure as the troughs that appear between white and yellow (for red), and between cyan and magenta (for blue). The brightest illuminance is achieved at the display's white point, where all three color channels are set to maximum intensity, followed by the secondary colors yellow and cyan, where two color channels are set to maximum intensity, and then green-hued tertiary colors, such as chartreuse and spring green, where the green color channel is set to maximum intensity while another is set to

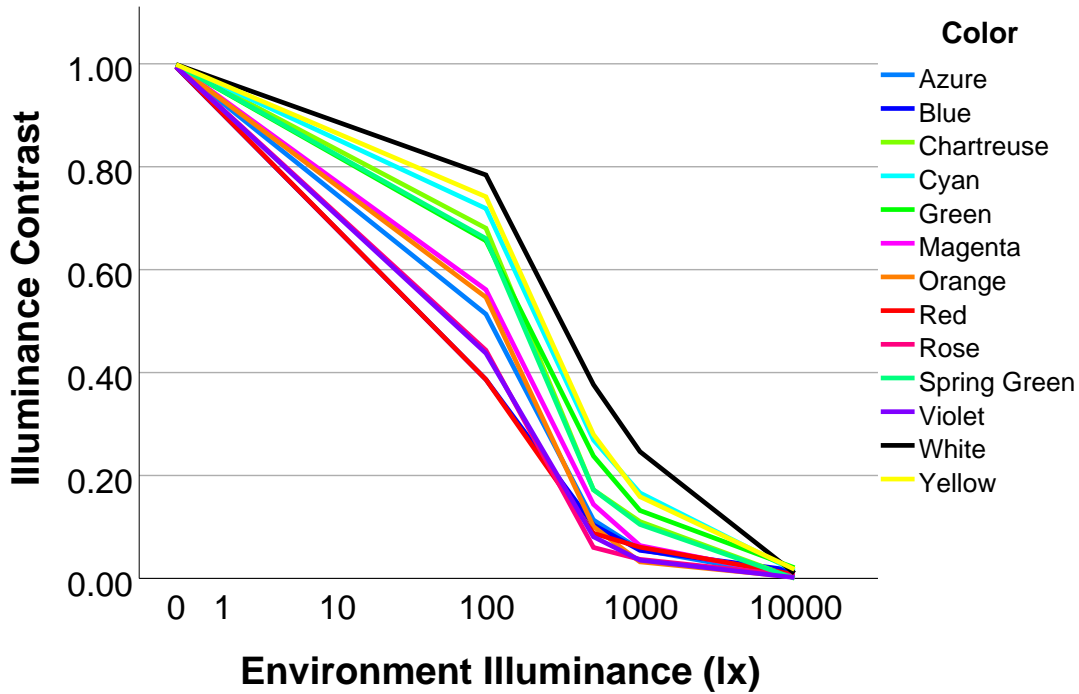


Figure 6.4: This figure depicts Michelson illuminance contrasts of each of the primary, secondary, and tertiary colors, as calculated via illuminance measurements made from the user’s eye position on the HoloLens 2. Imagery was set to fill the entirety of the display. Contrast of AR imagery decreases with increased environment illuminance. However, high luminance colors such as white, green, and yellow are more robust to changes to environment lighting compared to lower luminance colors such as blue and red.

half intensity. The minimum illuminance is always when displaying black, in which no light is emitted by the display, and is thus representative of the attenuated environment light as seen by the observer (kI_{env} , as represented in equation 6.13).

In general, this pattern in perceived intensity between the different displayed colors, which is particularly apparent for the 0 lux environment lighting plot, is inherent to display technology. This pattern arises due to the manner in which humans are sensitive to colors of different wavelengths [118]. People are most sensitive to green light compared to red and blue, and this sensitivity

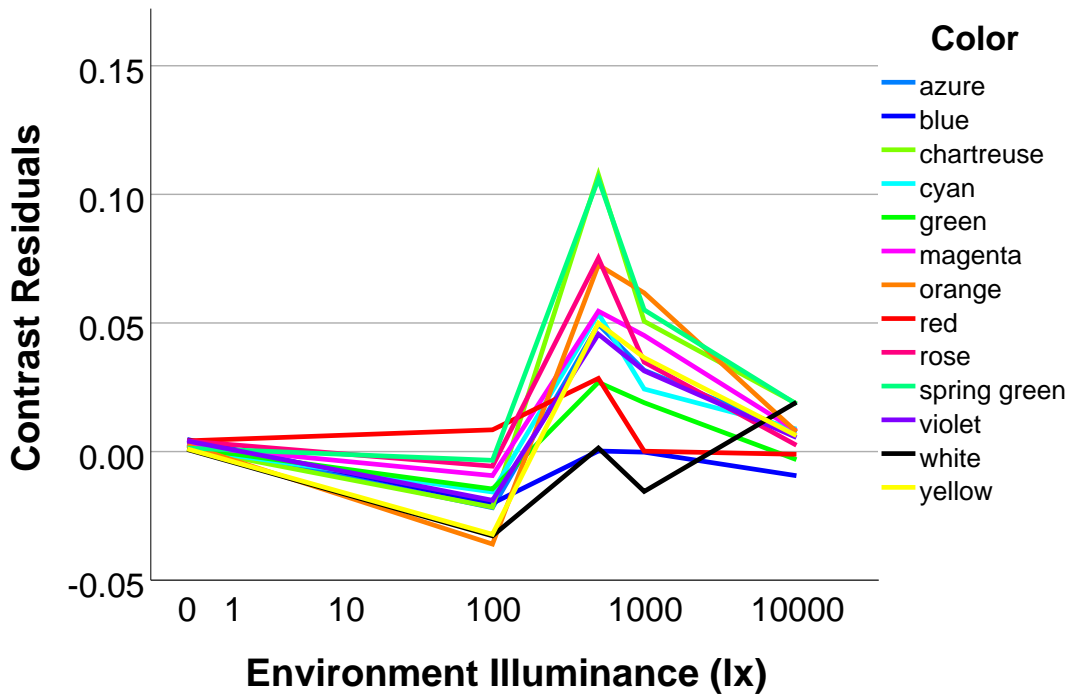


Figure 6.5: This figure depicts the residuals between the extended illuminance contrast values and the Michelson contrast values (extended minus Michelson). In general, modeled contrast closely match measured values for dark environment conditions (0 lux). For dim conditions (100 lux) modeled values tend to be underestimated, whereas for brighter conditions (500-10000 lux) they are overestimated. Residuals for the 500 lux lighting condition are higher than all other lighting conditions, particularly for the tertiary colors.

is inherently factored in for illuminance measurements, therefore increased illuminance for green-hued colors compared to red or blue hued colors are seen in the figure. A change in the maximum illuminance that can be achieved by the display would have the effect of stretching or shrinking the vertical dimension of this plot, meaning that different displays will have slightly different plots, however the relative differences between colors should remain relatively stable, as long as similar primary colors are used to create the color spaces of the displays.

A subset of these tested conditions combine to form the term **a**, as seen in equation 6.13, for

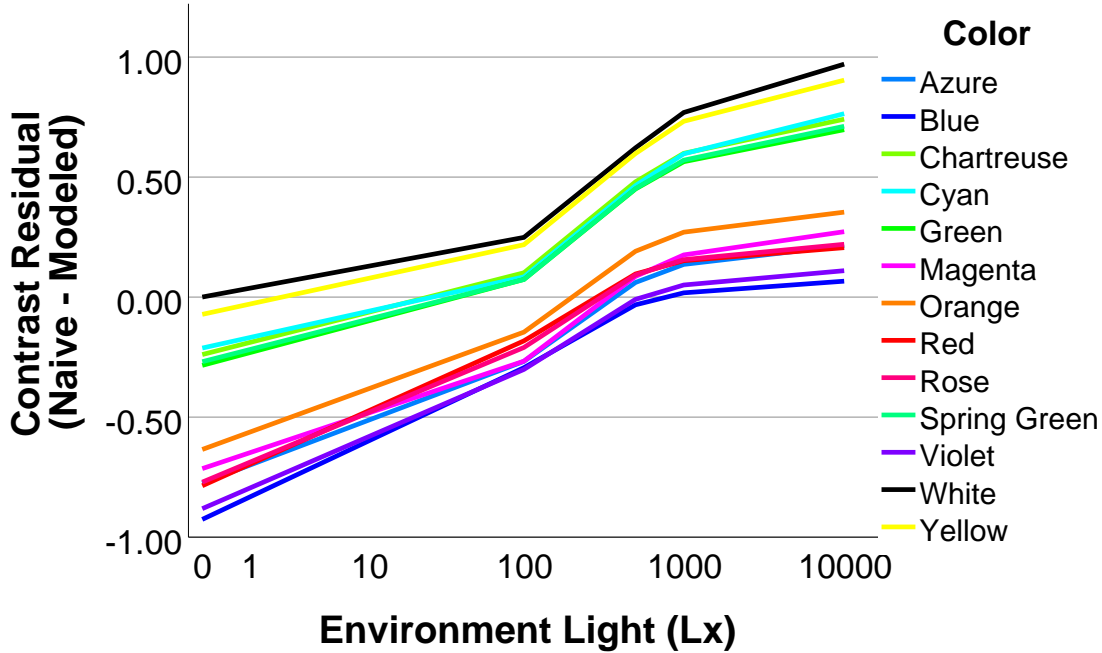


Figure 6.6: This figure depicts the residuals between the modeled contrast values and naive contrasts calculated based on the relative luminance of each of the colors. Naive contrasts consider the expected contrast between each of the colors and black, but do not factor in how environment light impacts perception. Because of this, naive contrasts are underestimated compared to modeled values in dim lighting conditions, and are overestimated in all lighting conditions greater than 500 lx.

the HoloLens 2. These are the measurements of the illuminance contributed from the red, green, and blue colors at the 0 lx environment lighting level. As shown in section 6.1, these particular measurements represent the amount of light contributed solely by the display and can be used to model the illuminance for any displayed color when used in a manner such as the one shown in equation 6.11. For reference, this value came out to be $\mathbf{a} = [34.9, 96.1, 31.3]$ in units of lux for the HoloLens 2. The value for \mathbf{a} in luminance units was similarly calculated using measurements made on a luminance meter (Konica Minolta CS-100) in a dark environment (< 1 nit) on the HoloLens 2. This value came out to be: $\mathbf{a} = [34.9, 96.1, 31.3]$ in units of nits (cd/m^2).

The term I_{env} , as represented in equation 6.13, is also measured several times in the procedure and forms the left-most column in table 6.2. These represent direct measurements of illuminance from the observer’s perspective in the physical environment prior to any attenuation performed by the display (i.e., without looking through the OST-HMD). Similar measurements could be made of the I_{env} term for luminance based units by taking an average luminance of several measurements made from the user’s perspective in several directions. Additionally, this could be gathered from a single luminance measurement if there is a particular point of reference in the physical environment that the luminance of the virtual imagery would be compared to.

The term k , as represented in equation 6.13, can be calculated based on the ratio of the measurements made of the display rendering black/transparent to the I_{env} measurement for that particular lighting condition ($k = I_{black}/I_{env}$). Based on the average of these ratios, gathered from columns 1 and 2 of table 6.2, I calculated that $k = 0.268$ for the HoloLens 2. As described in section 6.1, this term describes how the illuminance of the environment light is reduced after passing through the visor of the OST display. This value should be similar when calculated for luminance based units, however as mentioned in the limitations section (6.1.2) repeated luminance measurements would need to be made from many different directions from the user’s eye position and averaged to achieve an accurate measurement from luminance measurements.

6.2.4.2 Contrast Calculations

From the illuminance measurements gathered in the previous sections, I calculated illuminance contrast in several different manners:

- **Michelson Illuminance Contrast:** These contrast values are calculated using Michelson’s contrast equation (equation 6.1). For the parameters of the equation, I used two photometric

illuminance measurements made from the user's eye position within the AR display: one made with the display powered off, and one made with the display powered on and rendering the desired color.

- **Extended Illuminance Contrast:** These contrast values are calculated using the extended contrast model presented in section 6.1. Contrast values were calculated using the extended model as shown in equation 6.13. The I_{env} parameter for this equation was found using direct illuminance measurements of the user's physical environment. The attenuation factor, k and \mathbf{a} parameters were set according to the values obtained above, in section 6.2.4.1. The RGB value of the color shown on the display was used for the \mathbf{x} , and the RGB value for black was used for \mathbf{y} . Finally, the gamma correction function G was defined using a standard function of $G(x) = x^{2.2}$.
- **Naive Contrast:** These contrast values are calculated using Michelson's contrast equation with parameters defined by the *relative luminance* of the displayed color and the relative luminance of black (which is always zero). While these contrasts are effective for designing the appearance of imagery shown on displays in environments with fixed or controlled lighting conditions, they do not reflect how changes to the user's viewing conditions affect the perceived contrast of the virtual imagery.

A plot of the extended contrast values as a function of environment illuminance can be seen in figure 6.4. In general, contrast reduces with increased environment illuminance, highlighting the importance of considering the user's viewing conditions (particularly the lighting conditions in their environment) for AR applications. However, the rate at which contrast decreases with increased environment illuminance is different depending on the specific color of virtual imagery being shown on the AR display.

In figure 6.4, the contrast of red or blue colored imagery reduces faster compared to green or yel-

low colored imagery, which in turn reduces faster compared to white imagery. In practice, this phenomena can be exploited to improve the contrast of imagery shown on the AR display, such as by selectively changing the color of virtual imagery based on the lighting conditions of the user's environment. However, this approach is somewhat limited, which can be observed in the vertical distribution of contrasts by color in the figure. For instance at 100 lux of environment illuminance, shifting from red to white imagery could increase contrast from 40% to approximately 80%, whereas at 1000 lux of environment illuminance a similar color shift is less effective increasing contrast from approximately 5% to 25%. This effect demonstrates that future AR systems that adapt to maintain or improve the contrast of imagery will have to do more than simply change the color of AR imagery in response to changing environment conditions. This particular point will be explored in more detail in chapter 7.

A plot of the residuals between the extended illuminance contrast values and the Michelson illuminance contrast values can be seen in figure 6.5. This figure demonstrates that the contrasts predicted by the extended model (solely using photometric measurements of the user's environment) are accurate in that they closely resemble the contrasts calculated using the original Michelson's contrast equation with parameters directly measured from the user's eye position within the AR display. For most environment illuminance levels, the residuals fall within a range of plus or minus five percent. However, the residuals for the 500 lux environment lighting condition exceed beyond this range (up to 12%), particularly for the tertiary colors. This could imply that the AR display (HoloLens 2) is using a gamma correction function slightly different than the standard power function with an exponent of 2.2. Measurement of the gamma function specific to the AR display of interest and incorporation of this function into the contrasts calculated by the extended model could result in lesser residuals in a similar plot.

A plot of the residuals between the extended illuminance contrast values and the naive contrast values can be seen in figure 6.6 (naive contrast minus extended contrast). Naive contrasts solely

consider the relative luminance of the colors of interest, and ignore other factors such as the lighting in the user's environment and parameters associated with the user's AR display. Because of this, a graph of naive contrasts as a function of environment illuminance would result in horizontal plots. In figure 6.6, the grouping of colors by their luminance can again be observed, where plots for red and blue colored imagery are lowest on the vertical axis of the figure, followed by green, yellow, and white colors respectively. This figure demonstrates that naive contrasts are underestimated compared to the extended contrasts for low illuminance lighting conditions, and are overestimated for lighting conditions with illuminance greater than 500 lux.

6.3 Effects of Contrast on AR Search Task Performance

In the previous sections, I defined an extended model of luminance contrast that predicts contrast values as a function of terms representing the user's environment, their OST AR display, and the intended appearance of imagery being displayed. As mentioned in chapter 2, luminance contrast has been shown to directly affect the legibility of text and user performance in visual search tasks. In this section, I present a user study that tasks participant's to perform visual search tasks on UIs shown on an OST AR display, the Microsoft HoloLens 2. In this study, the extended contrast model is used to establish contrast values for each study trial experienced by participants. I demonstrate that these contrast values effectively predict user performance, user confidence in their responses, and user perception of task difficulty.

6.3.1 *Participants*

For this study, I recruited 9 participants (7 male, 2 female) ages 23-57 ($m=31.8$, $\sigma = 11.4$), from our local university community. All participants indicated that they had normal or corrected to

normal visual acuity and no known visual or neurological impairments. Participants provided their informed consent to participate in the study and the study procedure was approved by the IRB of our local university.

6.3.2 Methods

The main independent variable of our study is the luminance contrast between the text shown on the OST display and the area immediately surrounding the text. As previously discussed, luminance contrast is influenced by several factors related to OST displays: the feature color, the background color, the attenuation introduced when looking through the display, the environment lighting, and the display's maximum luminance. As the attenuation and luminance factors related to a particular display are typically fixed for current consumer OST displays, I chose to vary luminance contrast through manipulation of the virtual text color, virtual billboard color shown around the text, and participants' environment lighting conditions. Thus I introduced the following sub-variables in our study design:

- **UI Style** The style of the UI was randomly varied such that half of conditions were *Text Only*, consisting solely of randomly-colored text, while the other half of conditions consisted of pseudo-randomly generated *Text and Billboard* combinations, where text was presented in front of a background of a uniform color.
- **Environment Lighting** The lights in the testing environment were configured to provide three different levels of luminance from the participants perspective: 0.5, 170, and 190 lux (0.8, 100, and 180 nits).
- **Text and Billboard Colors** For *Text and Billboard* conditions, the text color was pseudo-randomly generated to fall into one of ten different luminance contrast bins in relation to

the billboard color. These bins occurred at intervals of 0.1 for Michelson contrast values between 0 and 1. Contrasts were calculated using the sRGB relative luminance of the text color and billboard color, ensuring an approximately uniform distribution of tested contrasts. For *Text Only* conditions, when no billboard was present, text color was randomly assigned without constraint to a particular contrast bin, ensuring a uniform distribution of tested colors with respect to the display's color space.

6.3.2.1 Study Task

The study task was adapted from the task used by Gabbard et al. [63]. Participants were presented with a single target letter, shown horizontally centered near the top of the background poster. Below the target letter were three rows of ten randomly-generated letters from the restricted alphabet of C, K, M, O, P, S, U, V, W, X, Z. Within these rows, five occurrences of the target letter were randomly positioned. Participants were instructed to search these three rows from left to right, top to bottom, as quickly as possible to find the third occurrence of the target letter. Once this letter was identified, the participant was instructed to press 1, 2, or 3 on the keyboard to indicate their level of confidence that they had chosen the correct letter. Participants were then instructed to use the arrow keys to position a cursor over the first three occurrences of the target letter, pressing a button to confirm on each occurrence. Participants were instructed during this phase to avoid correcting themselves if at this point they noticed an occurrence of the letter that they previously overlooked or if they mistook a different letter as the target letter. Upon inputting what the participant thought the first three occurrences of the target letter were, the participant was prompted to rate the difficulty of the task on a scale of 1 to 5 based on the lighting and color combination of the UI. From there, the task procedure repeated.

6.3.3 Materials

Virtual stimuli were presented to participants on the Microsoft HoloLens 2 via the holographic remoting feature of the unity engine (version 2019.4.26), which ran on a nearby laptop (Windows 11 pro v10.0.22000, Intel core i7-8750, 32 GB RAM, NVIDIA GTX 1070 GPU). A wireless Logitech USB keyboard paired to the laptop was used to gather participant responses during the experiment.

Text shown on the display was in unity's default font for text mesh pro (TMP) game objects (liberationSans SDF), and was sized at 72 point font (7.2cm square), which corresponds to a visual angle of approximately 2.06 degrees at 2 meters distance.

Participants were seated two meters away and perpendicular to a vertical surface such that perceived depth of virtual imagery matched the focal depth of the display as well as the physical distance between the participant and vertical surface. A uniform grey background poster covered the working area of the surface for which virtual annotations were superimposed. This poster was printed on archival matte paper with a pixel value of (128,128,128) and a size of approximately one square meter.

Two standing LED lamps were positioned immediately to the left and right sides of the participant's seat, which were used to illuminate the background poster. Additionally, three DMX overhead lights were positioned above the participant and angled towards the poster to increase the range of lighting conditions that could be configured for the study.

In order to establish parameters for the extended contrast model, the illuminance and luminance measurements from section 6.2.4.1 were used for the parameters: \mathbf{a} , and k . For model parameters \mathbf{x} and \mathbf{y} , the RGB values of the displayed text and billboard color were input as r , g , and b respectively to the gamma correction function defined by sRGB color standards: $G(x) = ((x + c_1)/c_2)^\gamma$, where

$c_1 = 0.055$, $c_2 = 1.055$, and $\gamma = 2.4$.

6.3.4 Procedure

Upon arrival, participants read an informed consent form and provided verbal consent to continue with the study procedure. They were given a brief overview of how to safely put on and take off the HoloLens 2, and were shown to their seat in the testing area. Participants then put on the HoloLens 2 and faced towards the background poster directly in front of them. Participants were then given time to familiarize themselves with the study procedure by performing several practice trials of the task. Once participants were comfortable with the task procedure, the lighting was set to the first lighting condition (chosen in a counter-balanced order to avoid sequencing effects). Participants completed 50 trials, which typically took about 10 minutes to complete, after which the environment lighting was set to the next lighting level. Participants were given several minutes (typically 2-3 minutes) between lighting conditions to adapt to the change in lighting. Finally, participants performed another 50 trials in the last remaining environment lighting condition. Prior to departing, participants completed a demographics survey and received monetary compensation for their participation.

6.3.5 Measures

Three primary measures were collected as the participants completed the study task:

- **Participant Confidence** Participants indicated their confidence in their task performance on a three point likert scale for each task trial.
- **Perceived Difficulty** Participants indicated the level of difficulty of the task on a five point likert scale for each task trial.

- **Response Time** The time it took the participant to identify the third occurrence of the target letter. The timer began when the UI first appeared to the participant, and ended when the participant indicated their confidence score.
- **Error Rate** Participants' performance in the task was measured in the number of errors made based on the selections the user made after indicating their confidence rating. One error was counted for each selection that was a letter other than the target letter (e.g. mistaking a letter 'U' for a letter 'V'). One error was also counted for each target letter that was overlooked (e.g. if the participant selected the second, third, and fourth occurrences of the target letter, this indicated that they overlooked the first occurrence, and thus one error was counted).

6.3.6 Hypotheses

Given that previous work has established a relationship between the luminance contrast and the legibility of text [134], the main hypotheses of the study were:

- **H1** The *contrast values* predicted by our model can be used to predict users' *objective performance* in AR tasks.
- **H2** The *contrast values* predicted by our model can be used to predict users' subjective perception of the *difficulty* of AR tasks.
- **H3** The *contrast values* predicted by our model can be used to predict users' *subjective perception* of their performance in AR tasks.

6.3.7 Results

In this section, I present the results of the user study described in the previous section.

6.3.7.1 Linear Regressions

Using the model parameters described in section 6.3.3, environment lighting, text color, and billboard color were combined using the contrast model from section 6.1 to yield a single luminance contrast value comparing the text color and background color for each trial of the study. For text only conditions, these contrast values compare the luminance of the text (luminance of the text color summed with the attenuated environment luminance) to the luminance of the area in the study environment surrounding the text (attenuated environment luminance). For text and billboard conditions, these contrast values compare the luminance of the text (luminance of text color summed with the attenuated environment luminance) to the luminance of the billboard (luminance of the billboard color summed with the attenuated environment luminance). Plots of contrasts effect on each of the dependent variables are shown in figure 6.7.

Linear regressions were performed in SPSS on the aggregated participant data to test if the modeled luminance contrast values predicted participant's error rates, response times, perceived task difficulty, and confidence. For error rates, the fitted regression model $y = -0.009x + 0.14$ was found to be statistically significant ($R^2 = 55.1\%$, $F(1, 8) = 9.84$, $p = 0.014$). For response time, the fitted regression model $y = -0.069x + 4.3$ was found to be not statistically significant ($R^2 = 36.6\%$, $F(1, 8) = 4.608$, $p = 0.064$). For perceived task difficulty, the fitted regression model $y = -0.14x + 2.8$ was found to be statistically significant ($R^2 = 74.7\%$, $F(1, 8) = 23.593$, $p < 0.001$). Finally, for participants' confidence, the fitted regression model $y = 0.046x + 2.45$ was found to be statistically significant ($R^2 = 65.9\%$, $F(1, 8) = 15.468$, $p = 0.004$).

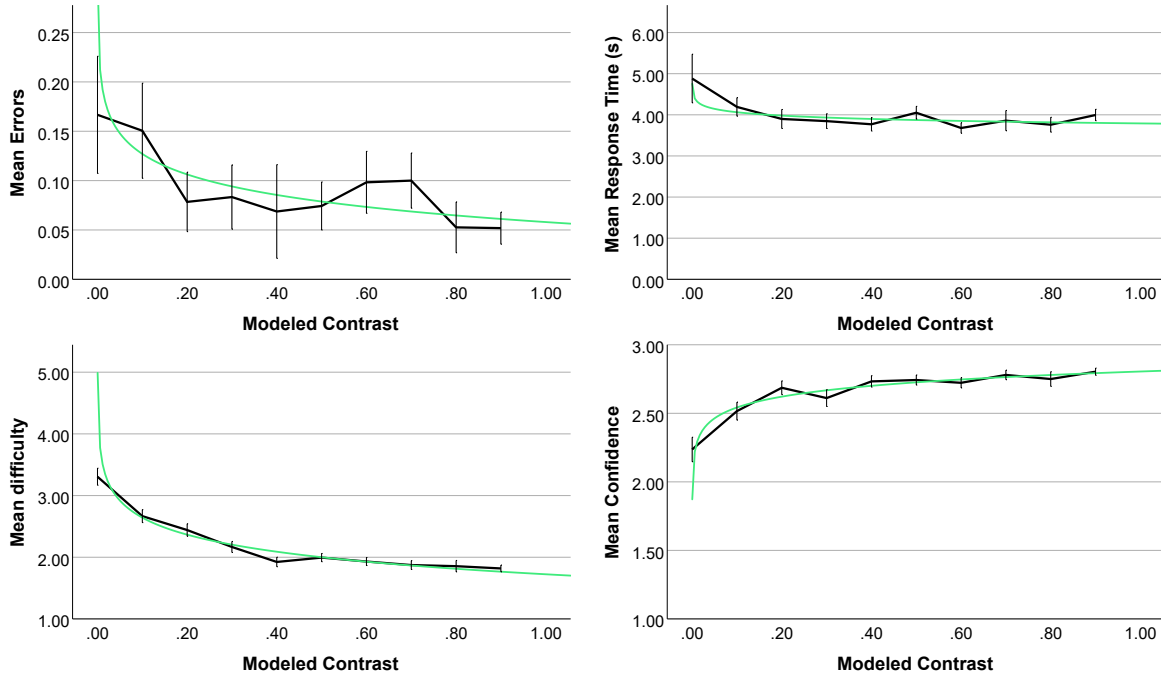


Figure 6.7: These figures show the effects of modeled luminance contrast on: participants' error rate (top-left), response time (top-right), confidence in their responses (bottom-right), and perceived task difficulty (bottom-left). Error bars represent a standard error with a multiplier of one. Contrast values are grouped into bins at intervals of 0.1. The fitted regression models from non-linear regressions are shown in green.

6.3.7.2 Non-Linear Regressions

In examining figure 6.7, contrast appeared to have non-linear effects on each of the dependent variables in the study. Thus, I proceeded to perform non-linear regressions for modeled luminance contrast on the aggregated participant data for each of the four dependent variables. Each of the dependent variables in figure 6.7 appear to be logarithmic functions in the form of $f(x) = a + b \cdot \log x$. Thus, I proceeded to use these functions as the kernels to our non-linear regressions performed in SPSS.

For error rate, the fitted regression model $f(x) = 0.058 - 0.03 \log x$ was found to have an R^2

value of 50.7%. For response time, the fitted regression model $f(x) = 3.791 - 0.118 \log x$ was found to have an R^2 value of 27.7%. For task difficulty, the fitted regression model $f(x) = 1.724 - 0.399 \log x$ was found to have an R^2 value of 94.3%. For user confidence, the fitted regression model: $f(x) = 2.805 + 0.114 \log x$ was found to have an R^2 value of 82.6%. These regression models can be seen in green along with the mean participant data in figure 6.7.

6.4 Discussion

In this section, I discuss the results of the user study presented in the previous section.

6.4.1 Contrast Predicts User Performance

In the analysis, the linear regressions of contrast predicting error rate was statistically significant with $R^2 = 55.1\%$ while the regression of contrast predicting response time was not, with $R^2 = 36.6\%$. The non-linear regressions performed reduced the R^2 values compared to the linear regressions, showing that luminance contrast explained 50.7% of the variance observed in the aggregated error rate data and 27.7% of the variance observed in the response times. These results partially support hypothesis **H1**, suggesting that contrast values from our model are somewhat effective at predicting user error rate in AR tasks, but not necessarily response time. From this, I partially accept hypothesis **H1**.

Since the regression model for response time was not significant, this suggests that there were other factors in our study design that were at least partially responsible for the variance in response time data. In debriefing participants after their completion of the study, some participants indicated that they thought that some trials were easier to complete than others due to the placement of the target letters within the three lines of text as well as the specific letter that was randomly chosen to be the

target letter for each trial. Participants indicated that in some trials, the subsequent occurrences of the target letter were grouped relatively close together, which made the task easier to perform. The placement of these target letters was chosen randomly and I did not record these letter positions in our data, so I cannot determine the extent to which this factor influenced user response times. In the 2007 study by Gabbard et al. they performed a post-hoc analysis on the effects of target letter on user response times, where they found that it had an Cohen's D effect size of $d = 0.07$. Although our task was modified from the task used in the Gabbard et al. study, the two tasks share the same restricted alphabet, and it is possible that the target letter in our task had a similar effect on user response times. This would at least partially explain the comparatively low predictive power of contrast on user response times.

While the regression model for user errors was significant, it only explained approximately half of the variance observed in the aggregated participant data (55.1% for the linear regression model). Similar to the response time data, I believe that other factors related to the study task, such as target letter choice and letter placement, may have at least somewhat affected the frequency of errors made by our participants. The visual search task used in this study was somewhat binary in terms of the frequency of user errors, where in the vast majority of trials (even at low contrast levels), users made no errors at all. Mean user errors decreased from approximately 0.17 in trials with contrast between 0 and 0.1, to approximately 0.5 in trials with contrast values between 0.9 and 1. Perceived difficulty similarly decreases from approximately 3.3, a medium difficult score in the lowest contrast conditions, to approximately 1.8, a relatively low difficulty score in the highest contrast conditions. From this, I can see that the task itself is fairly easy to perform, even in the worst of low contrast conditions. For future work, it would be interesting to evaluate several different AR tasks of varying degrees of difficulty as well as at varying levels of contrast to see if the patterns observed in our data here can be generalized across different types and difficulties of AR tasks.

The greatest change in mean errors occurs in the domain of contrast values between 0 and 0.2, after which errors somewhat stabilize. A similar pattern can be observed in the other three dependent variables, where in general less change is observed for contrast values greater than 0.2 compared. If attempting to establish a minimum recommended threshold contrast value from our data alone, it appears as though a contrast of 0.2 may be an effective lower bound, since all dependent variable seem to somewhat stabilize at greater values. However, this should be confirmed in future work, since variations in the study task itself could potentially influence this threshold.

One interesting thing to note in the error rate data is the apparent spike in errors made between contrast values of 0.5 and 0.8. For this specific region, aggregated participants' ratings of task difficulty are relatively low with little variance. Similarly, the participants' confidence ratings are relatively high within this range with little variance. This suggests that users were not aware of the errors that they were making for trials within this region. Since the variance in the error rate data was relatively high, it is hard to tell if this range of contrasts is actually problematic for user task performance, or if is simply an artifact in the data. Future work investigating the effects of contrast on user performance should verify this potentially interesting phenomena, as confirmation of its presence could have interesting applications for OST AR applications. For instance, if the contrast of AR imagery is detected to be in this undesirable range, then of course the ideal resolution is to improve conditions to achieve better contrast values. However, since this may not always be possible, it may instead be beneficial to lower contrast out of this range or otherwise notify the user aware that they may be more prone to make mistakes when performing AR tasks with imagery in this contrast range.

6.4.2 Contrast Predicts Subjective Task Perceptions

In the analysis, the linear regressions of contrast predicting users' confidence in their performance ($R^2 = 65.9\%$) and their perception of the task difficulty ($R^2 = 74.7\%$) were both significant. Additionally the non-linear regressions performed improved the R^2 values compared to the linear regressions, showing that luminance contrast explained ($R^2 = 82.6\%$) of the variance observed in the participant confidence data and ($R^2 = 94.3\%$) of the variance observed in the task difficulty data. These results support hypothesis **H2** and **H3**, suggesting that contrast values from our model are effective at predicting subjective qualities associated with AR visual tasks, thus I accept **H2** and **H3**.

This result suggests that the extended contrast model may provide an effective way to evaluate how confident users feel conducting visual tasks on OST AR displays as well as a way evaluate how difficult visual tasks are on these displays. Again, these results should be confirmed in future work, as variations to the AR task could potentially influence the patterns that arise in the perceived difficulty of the task with respect to contrast as well as in the users' level of confidence in their responses.

6.5 Summary

This chapter primarily supports thesis statement two, by presenting an extended contrast model for OST AR displays that can be used to evaluate the particular set of viewing conditions the user is in. The extended model adapts Michelson's contrast equation to include additional terms that represent the user's physical environment, their AR display, and the imagery being displayed. I demonstrated how photometric measurements can be made to establish the model parameters specific to the user's environment and AR display, and demonstrated that this extended contrast

model is capable of accurately predicting user performance and subjective perceptions of an AR search task.

In examining the errors made by participants in the search task for viewing conditions with intermediate contrast levels, an increase in errors was observed while no decrease to user confidence or task difficulty was observed for the same set of conditions. This particular result supports the latter part of thesis statement one, by demonstrating that in certain conditions, users may not even be aware that their viewing conditions are affecting their performance and/or AR experience.

In the next chapter, I demonstrate the use cases and applications of this extended contrast model.

CHAPTER 7: MODEL APPLICATIONS

In the previous chapter, I demonstrated how several different factors related to a user's viewing conditions are related in terms of the photometric measures of illuminance or luminance: the user's environment lighting conditions, characteristics of their AR display, and characteristics of the imagery being displayed. I also extended an existing contrast model to include these terms and demonstrated that the model makes accurate predictions of contrast compared to direct photometric measurements and is able to accurately predict user performance, and their subjective perception of a search task on an OST AR display.

In this chapter, I switch focus from establishing this extended contrast model to demonstrating its use cases and applications for the field of AR. The research in this chapter is used to address thesis statement three:

- **TS3 - Apply the Model:** The model from TS2 can be used to evaluate the user's viewing conditions, simulate changes to their viewing conditions, and mitigate negative effects the user is prone to experience through changes made to the AR imagery, to the characteristics of the user's AR display, or to the user's environment.

In the following sections, I will provide details on how to use the extended contrast model to inform adaptive AR systems, and I provide guidelines for designing imagery for use with OST AR displays that is robust to changes in the user's environment conditions.

7.1 Model Use Cases

The extended contrast model can be used internally within AR systems to improve the perceived contrast of AR imagery or to reduce the discrepancy between the perceived and intended appearance of virtual imagery shown on the display. Additionally, it can be used to simulate the perceived appearance of AR imagery for given sets of environment conditions and sets of display parameters. This section provides details and examples of how the extended contrast model can be used in these manners.

7.1.1 Adaptive AR Systems

The extended contrast model can be used to inform the AR system of the expected perceived contrast of AR imagery given the user's viewing conditions as well as to determine how large of a discrepancy there is between the perceived and intended appearance of virtual imagery. Following this, the model can be used to inform the actions taken by the system to improve contrast or reduce this discrepancy. In order to do this, the system needs to:

1. Establish Model Input Parameters
2. Evaluate if Adaptations are Needed
3. Decide Which Adaptations to Perform

In this section, we will discuss at a high level as to how each of these steps can be accomplished, and in some cases demonstrate how they have been handled in existing literature.

Ideally, the AR system would make use of the contrast model internally and directly control contrast enhancement methods through manipulation of the rendered imagery, parameters of the AR

display, or through control over devices within the user's physical environment. However, such a system could also potentially function through indirect control, such as by providing feedback to the user on how to improve their viewing conditions, reducing the discrepancy between the intended and perceived appearance of presented imagery. When using the model in this manner, it is useful to derive the model's equations using matrices to represent the parameters related to the imagery to be displayed and the user's environment conditions. With these matrices, rather than comparing between two points or two colors, \mathbf{x} and \mathbf{y} , in the virtual image, we are instead calculating the contrast between each pixel in the AR imagery and the environment directly behind it as observed by the user.

The equation would take the following form, where contrast \mathbf{C} , is a vector with length equal to the number of pixels in the AR image n , \mathbf{X} is a matrix of size n by 3 representing the AR image to be displayed, and \mathbf{i}_e is also a vector of length n representing the luminance of the user's environment behind each pixel in the AR image. In this equation, the division operation should be performed element wise (Hadamard division).

$$\mathbf{C} = \frac{\mathbf{X}\mathbf{a}}{\mathbf{X}\mathbf{a} + 2k \circ \mathbf{i}_e} \quad (7.1)$$

When arranged in this manner, k can either be a scalar representing a uniform decrease in environment light caused by attenuation on the display (such as from a neutral density filter), or it can be a vector of length n representing the attenuation applied to environment light for each pixel in the displayed image (similar to the methods explored by Itoh et al. [85] and Kaminokado et al. [91]). For this reason, we indicate the operation between k and \mathbf{i}_e with the \circ symbol to represent element-wise multiplication. In this form, non-uniform per-pixel attenuation could also potentially be specified and accounted for in the contrast calculation through introduction of another vector term that scales the luminance output of the display to account for non-uniformities in

the luminance output of the display, but this is left for future work. The following subsection will explain the various manners in which the parameters to this model can be found and used by the adaptive AR system.

7.1.1.1 Establishing Model Parameters

In the previous chapter, I demonstrated how the model parameters could be measured and calculated via direct photometric measurements made on the AR display of interest and within the user's environment. These can certainly still be used to gather the model parameters for use with an adaptive AR system, however it has disadvantages in that the I_e parameter would need to be measured frequently, as any changes in the user's environment or in the user's position could potentially change this value. Ideally, the AR system should be able to make such measurements on its own, when needed.

As the terms related to the particular AR display of interest, \mathbf{a} and k are typically fixed, these values can potentially be measured by the manufacturer or developer and saved for reference when needed. Additionally, the parameter \mathbf{X} describing the AR imagery is essentially a texture representing the AR imagery to be displayed at a particular frame, thus this parameter is known as long as adaptation strategies are being handled as a part of the rendering pipeline of the AR display or system. With these known values, the remaining parameter to be found is I_e describing the lighting conditions in the user's environment.

Some existing methods have demonstrated that external RGB cameras fixed to the AR display can be used to establish luminance and chromaticity values for discrete points in the user's physical environment, which in turn can be used to inform adaptive contrast methods or color blending correction methods [78, 105]. These methods involve computation of a mapping that transforms the imagery from the RGB camera's perspective to the perspective defined according to the eye

position of the user, where it is converted from RGB values into chromaticity values and/or the luminance values needed to establish the vector term \mathbf{i}_e . It should be noted that this approach in establishing the vector term \mathbf{i}_e is limited in that it assumes the light reaching the camera's position (also considering attenuation) is comparable to the light reaching the user's eye position. In practice, this may not always be the case, since the user's gaze direction changes, introducing changes to the amount of light reaching their eyes compared to if they were looking straight ahead. It is possible that eye tracking could be employed to take the user's eye gaze direction into account for this calculation, but the intricacies of such an approach are left to consider in future work.

Using equation 7.1, the AR system has the capability to calculate a vector of contrast values \mathbf{c} , that represents the contrast at each pixel on the display for a given frame. Unless the system is informed of a particular feature or region of interest within this vector, a global measure of contrast can be calculated by taking the mean of all values in the vector. For such mean contrast calculations, it is important to consider that large portions of the field of view of the display may not contain AR imagery, which means that contrast calculations performed on these "empty" pixels would result in contrast values of one. This would weigh the mean contrast value up (towards one), making it seem as though viewing conditions were better than they actually are. To prevent this, the alpha channel of the pixels in the output imagery should be checked in order to verify if the pixel was intended to appear black or if the pixel is otherwise not a part of the AR imagery being presented. Pixels with a zero alpha value should not be counted toward this global mean contrast calculation.

7.1.1.2 Evaluating if Adaptations are Needed

There are several potential methods that can be used to determine whether or not the system should adapt to change the perceived appearance of the virtual imagery. In this section, I will provide details on using a contrast threshold and using a characterization of the intended image appearance to

determine if adaptations are needed. The pros and cons of these methods will be briefly described in this section.

One option for determining whether an adaptation method is needed is to look at the mean contrast value obtained from the model compared against a threshold value, representing a minimum acceptable contrast, where if the mean contrast from the model is found to be less than the threshold, potential adaptation methods are explored. Accessibility guidelines made by the W3C recommend a minimum simple contrast ratio of 4.5 to 1 for tasks such reading web-based virtual content. When converted to Michelson contrasts similar to the ones output by the extended contrast model, this is a contrast value of 64%, which is quite a bit higher than the threshold observed in the data from the study presented in chapter 6, where it would appear that contrast values near 20% could potentially be considered for this threshold. Since the W3C threshold is the higher of the two mentioned here, we recommend to use the W3C recommended minimum contrast of 64% until such a time where future work has confirmed this or has established a different threshold specifically for OST displays.

The downsides of using this threshold-based approach to determining whether adaptation strategies are employed is that it treats all types of imagery that can be displayed on the AR display as equal, and it is not comparing against the intended appearance of the AR imagery. Regarding the former point, it may be acceptable to use an adaptation strategy that alters the perceived color of a UI to improve the perceived contrast, for instance changing the foreground color of text or other UI features from blue to white. However, similar methods employed on other types of imagery, such as images or representations of virtual humans, may have undesirable effects. For instance, the appearance of a virtual human could be shifted to where it appears they are wearing different colored clothes or even have a different skin tone or complexion. Similarly, colors used to encode information could be shifted in a manner that makes it difficult for the user to follow the encoding scheme, which may make the user prone to misinterpreting the encoded information. Additionally,

while this approach is useful for maximizing contrast, it may be case that the developer/designer intended for the imagery to appear with low contrast, perhaps for aesthetic or stylistic reasons, and changing the perceived appearance of the imagery comes at some sort of detriment to the AR experience.

An alternative to the contrast threshold-based approach is to compare the mean contrast value obtained by the model to a characterization of the intended appearance of the AR imagery. Such a characterization of the imagery's intended appearance could be calculated in many different manners, and once again, a threshold value would be needed to determine whether the discrepancy between this value and the mean contrast from the extended model is sufficiently large enough to warrant use of an adaptation strategy. For instance, the relative luminance of each pixel in the AR imagery can be easily computed, and the mean could be taken, characterizing the intended appearance of the AR imagery as a mean relative luminance. The difference between this and the mean contrast obtained from the extended model can be used to consider the amount of discrepancy between the intended and perceived appearance of the AR imagery, where if the discrepancy is found to be greater than a threshold value, adaptation methods are explored. For this type of evaluation, it is less intuitive as to the threshold value that should be used in the comparison, and it is likely that different characterizations of the imagery's intended appearance may need different values to perform in a similar manner. Future work should investigate how the use of adaptation strategies, informed by different characterizations of the imagery's intended appearance, affect the user's performance and their subjective perception of the imagery over time.

7.1.1.3 Decide Which Adaptations to Perform

When it is determined that adaptation strategies are needed, the following question arises: What can we do to improve the perceived contrast of the AR imagery? One of the primary benefits of

using the extended contrast model is that it allows us to compare between different potential adaptation methods that are available at a particular time. With the model, we can evaluate how much change in perceived contrast is expected for each of the potential adaptation methods available to the system.

The AR system has potential adaptation methods that can be applied to change different parameters of the contrast model: adaptations to AR imagery, adaptations to the AR display parameters, and adaptations to the user's physical environment. The change in contrast that can be gained is different for each of three groups, where changes to AR imagery are somewhat limited in what they can achieve compared to changes made to the display or the user's environment.

For example, let us consider a set of conditions in which the mean contrast of the scene has been calculated to be below the desired minimum threshold. The system wants to evaluate potential methods of increasing the contrast above the threshold, and starts by examining adaptations that can be applied to the output color of the AR imagery. The maximum increase in contrast that can occur when solely adapting the color of AR imagery happens by shifting the colors to appear white instead of their intended color. Since white is the maximum luminance light the display is capable of emitting, this provides an upper bound with which we can evaluate how well AR imagery adaptations can potentially perform in the current viewing conditions defined by a set of static display parameters and static environment luminance values. If the mean contrast calculated using the extended model with all-white imagery is not higher than the minimum threshold contrast value, then adaptations to the display or to the user's environment must be considered alternatively or in addition to the adaptations made to the AR imagery.

A similar procedure can be used to evaluate how changing the attenuation of the display or the luminance within the physical environment would affect the perceived contrast of imagery on the OST HMD. In this case, the terms defining the AR imagery and luminance of the display are

held fixed in addition to either the attenuation or environment luminance terms. Then either the attenuation or environment luminance terms are adapted and the effects on contrast are observed. In this manner, the system can evaluate how much attenuation would need to be applied to the display or how much of a reduction in environment lighting conditions would need to occur to bring the contrast value to the target value.

Once an adaptation strategy, or set of adaptation strategies has been identified, the system can either directly enact the change or make suggestions to the user on how to adjust the display or their environment to suit their needs in their AR experience.

7.1.2 Simulating the Perceived Appearance of Imagery on Optical See-Through AR Displays

Another general use case of the extended contrast model presented in chapter 6 is that it can be used to simulate the perceived appearance of virtual imagery in different conditions. The extended model separates its input parameters into three main groups: parameters that describe the user's physical environment, parameters that describe the user's AR display, and parameters that describe the imagery being displayed. A simulator can take all of these parameters as input, and allow the user to examine how changes made to any subset of the parameters affect the contrast and the subjective quality of the imagery. Such a simulator can be useful for:

- Evaluating the designs of virtual imagery in varying environments and/or on varying AR displays
- Evaluating the design of new AR displays or modifications made to a particular AR display of interest
- Evaluating the design of physical environments in which AR technology will be used

The remainder of this subsection describes the how such a simulator can be created and used.

7.1.2.1 Methods

The most basic requirement of the simulator is that it needs to be able to render a first person view in a 3D virtual environment with AR virtual imagery superimposed over the camera's view. For this reason, game engines such as the Unity engine and the Unreal engine are particularly well suited for the creation of the simulator. In order to control the blending of the AR imagery with the virtual environment, it is desirable to use a game engine with a rendering pipeline that is capable of working in physical lighting units, such as the high definition rendering pipeline (HDRP) of unity. This feature allows you to specify the intensity of all lights in the virtual environment in units of either luminance or illuminance. As lighting calculations are performed in the rendering pipeline, photometric luminance values are accessible through custom shaders for each pixel in the camera's field of view. Prior to the tone mapping step in the rendering pipeline, accessing these values will yield a non-normalized luminance value for the pixel in the scene, as if you had measured it manually with a luminance meter. For the purposes of the simulator, these are the values that are needed to calculate the blending of AR imagery with the camera's view of the scene. After the tone mapping step, these values are compressed to fall in the range of 0 to 1, and the luminance data is essentially lost.

The key component in the simulator is a custom shader that is called just prior to tone mapping in the game engine's rendering pipeline. For the unity engine, this involves creating a custom pass shader that gets injected at "BeforePostProcess" injection point. This shader needs a reference to a texture representing the AR imagery to be displayed, as well as a reference to the luminance of the three primary colors of the AR display to be modeled (the \mathbf{a} term of equation 6.13), as well as the attenuation factor (k term).

This shader should first sample the color of the pixel in the camera's render texture, and the luminance component of this color should be scaled by the attenuation factor k term of the AR display being modeled, while the chromaticity of the color is maintained. Then, the shader should sample the texture containing the AR imagery, and calculate the expected luminance of this pixel using equation 6.12 from chapter 6. The color of the AR pixel should then be converted such that the chromaticity of the sampled color is maintained, while the luminance is adjusted to the expected luminance value from the previous step. Following this, a blending of the AR color and scene color can be calculated using the equations specified in ¹, with the colors specified in the CIE 1931 color space. This final blended color should be output to the texture operated on by the shader.

7.1.2.2 *Limitations*

While no simulator is perfect, some are useful, and the simulator described above is no exception. One of the main limitations of simulators similar to the one described above is that the display that the simulated imagery is output to typically has a much lower dynamic range compared to the dynamic range of the human eye. Because of this, the output imagery is compressed, particularly so for bright scenes, and the resulting imagery output by the simulator is not a perfect representation of what the user would see in the similar physical environment.

Game engines such as unity and the unreal engine typically employ a post-processing step after tone mapping where exposure is used to brighten or darken the appearance of displayed colors according to the perceptual state of the observer or the exposure time of the modeled camera. This step is sometimes used to model how an observer's eye adapts according to the lighting conditions in their environment, and thus this post processing step will influence the appearance of imagery

¹<https://www.ledsmagazine.com/smart-lighting-iot/white-point-tuning/article/16695431/understand-color-science-to-maximize-success-with-leds-part-2-magazine>

output by such a simulator. Therefore, it is important to consider the particular exposure model being using in the rendering pipeline, to ensure that it is best simulating the desired perceptual state of the user/observer in the simulated environment.

7.2 Design Guidelines and Considerations

Based on the understanding provided by our model, we have several guidelines for designing UIs and other types of virtual imagery to be viewed on OST displays.

In general, the contrast of any virtual element compared to the observer's view of their environment will be maximized when using combinations of black/transparent and white virtual imagery. As demonstrated above, this combination will result in the widest range of usable environment lighting conditions. For the specific case of UIs, it would seem intuitive that using either black features on a white background or white features on a black background would result in similar user performance and usability of the UI. However, Erickson et al. [48] demonstrated that "dark mode" style UIs with white features on black backgrounds offer several advantages in the form of increased visual acuity, increased usability, and increased visual comfort compared to light mode style UIs with white backgrounds and black features. Therefore, in general we recommend using dark mode style UIs with white features on black backgrounds.

When considering multiple colors, such as for a two-colored UI with a foreground/text color and a background color (see the tile-based UI design in figure 6.1), this pairing of colors should ideally be compared using equation 6.13 to find out if the contrast is acceptable for the desired environment lighting conditions. In order for UI designs to be robust to changing environment conditions, there are several heuristics that can be followed to inform the color choices of foreground and background UI elements:

Increase Foreground Luminance: If considering the UI colors within the hue, saturation, value color space (HSV), decreasing the HSV saturation of the UI foreground color while increasing its HSV value will have the effect of increasing the luminance of the color by shifting the color towards the white point of the display, thus increasing the contrast between the foreground and background UI elements.

Decrease Value of Background: Similarly, the background elements of the UI can be decreased in HSV value. Decreasing the value of the color has the effect of shifting the background color towards black, which increases the contrast between the background and foreground of the imagery. This also has the effect of reducing the opacity of the background.

If applying these techniques to the UI shown in figure 6.1, we would expect to see several changes to the UI's appearance. In general, the foreground colors used to represent text and most symbology in this screenshot are already displayed in white, however the appearance of the blue-tinted symbols, such as the ones on the leftmost column, could be white-shifted to improve the contrast between the symbol and background. Further improvements to contrast could come about through manipulation of the background colors, where the color of the main tile and smaller tiles within could be shifted from blue towards black in order to improve the contrast between the items within the frame and the user's physical environment, although this could potentially interfere with the user's ability use Gestalt principles to group information in the way the designer intended them to [158].

7.3 Summary

In this chapter, I provided details on how the extended contrast model can be used to inform adaptive AR systems and to simulate the perceived appearance of virtual imagery in different

conditions, and I provided guidelines for designing imagery for OST AR displays that is robust to changes in the user's environment conditions.

There is still much work to be done to achieve AR systems capable of adapting in real time to changes in the user's viewing conditions in order to maintain the perceived appearance of virtual imagery shown on the display. However, the work presented in this and the previous chapter lay the groundwork for such systems, and demonstrate the requirements that need to be met to attain them.

CHAPTER 8: CONCLUSION

In this chapter, I summarize the key findings of this dissertation with respect to the thesis statements described in chapter 1, and highlight the opportunities for future work in this research area.

This dissertation involved research addressing three main thesis statements:

- **TS1 - Identify Problems:** In increasingly common viewing conditions, AR users are susceptible to misperceiving or completely missing information shown on their AR display. Sometimes the user may not even be aware that their AR experience or their performance is being affected.
- **TS2 - Establish a Model:** The discrepancy between the intended and perceived appearance of AR imagery can be modeled in terms of luminance contrast, as a function of parameters specific to the user's viewing conditions. These parameters include the user's environment lighting conditions, the characteristics of their AR display, and the characteristics of the AR imagery.
- **TS3 - Apply the Model:** The model from TS2 can be used to evaluate the user's viewing conditions, simulate changes to their viewing conditions, and mitigate negative effects the user is prone to experience through changes made to the AR imagery, to the characteristics of the user's AR display, or to the user's environment.

In the next several sections, I describe how the research in this dissertation has addressed these thesis statements.

8.1 Thesis Statement One - Identify Problems

Portions of chapters 3 through 6 were used to partially support thesis statement one. In chapter 3, I demonstrated that the user's environment lighting conditions affect the legibility of text and usability of UIs, such that when the lighting in the user's environment brightens, the user's ability to read text and use UIs shown on an OST AR display is reduced. A similar effect was found in the results of chapter 5, where brighter lighting conditions caused users to misperceive virtual humans, where users perceived the virtual humans and avatars to be less human like when bright lighting conditions caused them to appear more transparent. While the studies in chapter 4 did not investigate environment lighting as an independent variable, it was found that the effectiveness of the visual attention cues was reduced compared to similar cues tested on more traditional displays, including flat panel displays and VR or VST AR displays. It is possible that environment lighting had a similar effect on the effectiveness of the visual cues, however this should be confirmed in future work. Additionally, chapter 4 revealed that the physical appearance of the users environment has a significant effect on the effectiveness of the attention cues, where cues are less effective in environments with a uniform background appearance compared to a non-uniform appearance.

Taken together, the results of these studies demonstrate that there are many different environments where users are prone to experience negative effects when using an OST AR display, particularly bright environments and environments with uniform appearances. The range of lighting conditions tested in these studies was relatively small compared to the range of lighting conditions that can occur as users start to use OST AR displays in their daily life. Many sunlit environments are brighter in terms of illuminance than the environments tested here, which means the effects observed in these studies will be exaggerated, and users will be prone to misperceive text, UIs, and virtual humans shown on the OST AR display. Additionally, the user study in chapter 6 demonstrated an interesting effect. When the AR user's viewing conditions were described in terms of contrast

using the extended contrast model, viewing conditions with intermediate contrast values (between 0.4 and 0.8) resulted in more errors being made by participants compared to just outside this range. Despite the increase in errors, users did not change their reported confidence and difficulty scores, indicating that they may have been unaware of their decrease in performance, rating their confidence high and task difficulty low. This finding is particularly important, as it demonstrates that in certain viewing conditions users may not recognize that their performance and their AR experience is being affected until those conditions have deteriorated to low contrast ranges (less than 0.2). In these conditions, the user will be making decisions and actions based on their interpretation of the information being shown, and if they misinterpret or miss certain information being displayed, their decisions and actions will suffer as a consequence.

8.2 Thesis Statement Two - Establish a Model

Thesis statement two was primarily addressed in chapter 6, where I demonstrated that an AR user's viewing conditions can be described in terms of photometric measurements, namely luminance or illuminance. I defined the AR user's viewing conditions as the relationship between the user, their physical environment, their AR display, and the particular imagery being shown on the display. Each of these factors can be described in terms of luminance or illuminance, and existing contrast models can be extended to include these terms, allowing for calculation of a contrast that is a description of the user's viewing conditions and their overall AR experience.

This chapter demonstrated that the predictions of the extended contrast model closely resemble contrasts calculated from direct measurements made on an OST AR display. Additionally, the user study in chapter 6 demonstrated that the contrasts from the extended model can significantly predict user performance, and their subjective perceptions of a visual search performed on an OST AR display. While this is just one of many particular tasks that can be performed on OST

AR display, it shows promise that the extended model is useful in describing the user's viewing conditions and their AR experience. Future work should investigate additional types of tasks and applications that can be performed on OST AR displays identify other potential strengths and/or weaknesses of using the extended model to describe an AR user's viewing conditions. Future work should also investigate potential modifications that can be made to the model to improve its ability to describe the experience of users in AR applications, for example factoring terms that describe optical mura such as luminance non-uniformities of the display. Also, additional work could be done to compare the model's predictions when used with input terms specified in luminance versus specified in illuminance.

8.3 Thesis Statement Three - Apply the Model

Thesis statement three is supported by the research presented in chapters 6 and 7. From chapter 6, the results of the user study demonstrate how an AR user's viewing conditions can potentially be split into three different categories:

- High Contrast - Viewing conditions with high contrast (greater than 0.8) that are optimal for AR usage. The participant data for this range had relatively low error rate, high user confidence, and low perceived task difficulty. When working in similar conditions, the user is unlikely to experience significant detrimental effects when using their OST AR display.
- Mid Contrast - Viewing conditions where contrast is sub-optimal for AR usage (between 0.4 and 0.8). The participant data for this range was characterized by an increase in error rate compared to high contrast conditions, while user confidence and perceived task difficulty was mostly unchanged. In these particular conditions, the user may not be aware that their perception, and therefore their decisions and actions, are being negatively affected by their

viewing conditions.

- **Low Contrast** - Viewing conditions where contrast is poor (less than 0.4), and imagery on the OST AR display may be difficult for the user to interpret or even to identify. The participant data for this range was characterized by an increase in errors made, a decrease in user confidence and an increase in perceived task difficulty. In these conditions, the user is aware that their performance is being affected by their viewing conditions, which may cause them to be more careful when making decisions and actions in these conditions.

It should be noted that the specific contrast ranges used to describe these different viewing condition classifications above are based on one particular task performed in OST AR. Future work should evaluate other types of tasks and applications in OST AR to see if the ranges in these descriptions hold for more general AR usage, or if separate ranges are found for other types of tasks and applications.

In chapter 7, I outlined several potential use cases for the extended model introduced in chapter 6. I first demonstrated how the model can be incorporated into future AR displays to allow for control over the AR imagery, characteristics of the AR display, or characteristics of the user's environment, to regulate the perceived appearance of imagery shown on OST AR displays. Such systems could adapt to improve contrast for tasks and applications where changes to the perceived color of virtual imagery are less consequential. Otherwise, such systems could adapt to maintain the intended appearance of the virtual imagery shown on the display in response to changes in the user's viewing conditions. Future work should investigate the potential tradeoffs involved with use of such methods, as maintaining the intended appearance of the virtual imagery could come at a cost of making the perceived appearance of the user's physical environment more dynamic. This could potentially harm user performance in contexts where they must maintain awareness of both virtual imagery and their physical surroundings. The results of studies in this area could be used to

determine when adaptation strategies employed by an OST AR display are particularly effective or ineffective, as well as identify specific manners of enacting the speed and transition of adaptations recommended by the systems.

APPENDIX A: IRB APPROVAL LETTERS



UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

APPROVAL

June 18, 2019

Dear Gregory Welch:

On 6/18/2019, the IRB reviewed the following submission:

Type of Review:	Modification / Update – Addition of study personnel.
Title:	Enhanced Perception and Cognition in Augmented Reality
Investigator:	Gregory Welch
IRB ID:	MOD00000258
Funding:	Name: Office of Naval Research
Grant ID:	ONR award# N00014-17-1-2927
IND, IDE, or HDE:	None
Documents Reviewed:	None

The IRB approved the protocol from 6/18/2019 to 4/15/2020.

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Adrienne Showman
Designated Reviewer



UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board

FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

APPROVAL

March 26, 2020

Dear Gregory Welch:

On 3/26/2020, the IRB reviewed the following submission:

Type of Review:	Modification and Continuing Review
Title:	Enhanced Perception and Cognition in Augmented Reality
Investigator:	Gregory Welch
IRB ID:	MODCR00000365
Funding:	Name: Ofc of Naval Research, Funding Source ID: ONR award# N00014-17-1-2927
Grant ID:	None
IND, IDE, or HDE:	None
Documents Reviewed:	• CITI Training All Members, Category: Training;

The IRB approved the protocol from 3/26/2020 to 3/25/2021.

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system. Guidance on submitting Modifications and a Continuing Review or Administrative Check-in are detailed in the manual. When you have completed your research, please submit a Study Closure request so that IRB records will be accurate.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Adrienne Showman
Designated Reviewer



UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board

FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

APPROVAL

June 18, 2021

Dear Gregory Welch:

On 6/18/2021, the IRB reviewed the following submission:

Type of Review:	Modification / Update to study team members, added survey instruments
Title:	Enhanced Perception and Cognition in Augmented Reality
Investigator:	Gregory Welch
IRB ID:	MOD00001985
Funding:	Name: Ofc of Naval Research, Funding Source ID: ONR award# N00014-17-1-2927
Grant ID:	None
IND, IDE, or HDE:	None
Documents Reviewed:	<ul style="list-style-type: none">• Familiarity with Technology Questionnaire.pdf• Protocol• Single Item Questionnaire.pdf

The IRB approved the protocol from 6/18/2021 to 3/23/2022.

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system. Guidance on submitting Modifications and a Continuing Review or Administrative Check-in are detailed in the manual. When you have completed your research, please submit a Study Closure request so that IRB records will be accurate.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Renea Carver
Designated Reviewer



UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board

FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

APPROVAL

May 31, 2022

Dear Gregory Welch:

On 5/31/2022, the IRB reviewed the following submission:

Type of Review:	Modification / Update
Title:	Enhanced Perception and Cognition in Augmented Reality
Investigator:	Gregory Welch
IRB ID:	MOD00002968
Funding:	Name: Ofc of Naval Research, Funding Source ID: ONR award# N00014-21-1-2578; Name: Ofc of Naval Research, Funding Source ID: ONR award# N00014-17-1-2927
Grant ID:	None
IND, IDE, or HDE:	None
Documents Reviewed:	None

The IRB approved the minor modification from 5/31/2022 to 2/8/2023.

In conducting this protocol, you are required to follow the requirements listed in the Investigator Manual (HRP-103), which can be found by navigating to the IRB Library within the IRB system. Guidance on submitting Modifications and a Continuing Review or Administrative Check-in are detailed in the manual. When you have completed your research, please submit a Study Closure request so that IRB records will be accurate.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Kamille C. Birkbeck

Kamille Birkbeck
Designated Reviewer



UNIVERSITY OF CENTRAL FLORIDA

College of Nursing
Department of Computer Science
Institute for Simulation & Training
12201 Research Parkway, Suite 300
Orlando, FL 32826-2210

Friday, March 31, 2023

To whomever it may concern,

This letter is to confirm that the human-subjects user studies run as a part of Austin Erickson's dissertation research were covered under the IRB-approved study titled: "Enhanced Perception and Cognition in Augmented Reality" (IRB ID: MOD00000258), which was initially approved on November 30, 2018, and currently expires on January 9, 2024. Austin was a Sub-Investigator on the study.

Sincerely,

A handwritten signature in blue ink, reading "Greg F. Welch".

Prof. Gregory F. Welch, PhD, FNAI (Computer Science)
Pegasus Professor and AdventHealth Endowed Chair in Healthcare Simulation
College of Nursing, Academic Health Sciences Center
Computer Science, College of Engineering and Computer Science
Institute for Simulation & Training, School of Modeling, Simulation, and Training
Co-Director of the Synthetic Reality Lab

Prof. Gregory F. Welch | +1 407.796.2823 | welch@ucf.edu | nursing.ucf.edu | sreal.ucf.edu

LIST OF REFERENCES

- [1] H. Adams, J. Stefanucci, S. H. Creem-Regehr, G. Pointon, W. B. Thompson, and B. Bodenheimer. Shedding light on cast shadows: An investigation of perceived ground contact in ar and vr. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2021.
- [2] F. Albarello, R. Crisp, and M. Rubini. Promoting beliefs in the inalienability of human rights by attributing uniquely human emotions through multiple categorization. *The Journal of social psychology*, 158(3):309–321, 2018.
- [3] F. Albarello and M. Rubini. Reducing dehumanisation outcomes towards blacks: The role of multiple categorisation and of human identity. *European Journal of Social Psychology*, 42(7):875–882, 2012.
- [4] M. Anderson, R. Motta, S. Chandrasekar, and M. Stokes. Proposal for a standard default color space for the internet—srgb. In *Color and imaging conference*, volume 4, pages 238–245. Society for Imaging Science and Technology, 1996.
- [5] L. Andrighetto, C. Baldissarri, S. Lattanzio, S. Loughnan, and C. Volpato. Humanitarian aid? two forms of dehumanization and willingness to help after natural disasters. *British journal of social psychology*, 53(3):573–584, 2014.
- [6] R. T. Azuma. A survey of augmented reality. *Presence: teleoperators & virtual environments*, 6(4):355–385, 1997.
- [7] B. Bastian, K. Costello, S. Loughnan, and G. Hodson. When closing the human–animal divide expands moral concern: The importance of framing. *Social Psychological and Personality Science*, 3(4):421–429, 2012.

- [8] D. Bauer and C. R. Cavonius. *Ergonomic Aspects of Visual Display Terminals*, chapter Improving the legibility of visual display units through contrast reversal. London: Taylor & Francis, 1990.
- [9] E. Bayle, S. Hourlier, S. Lelandais, C.-A. Salasc, L. Leroy, J. Plantier, and P. Neveu. Interocular conflict from a monocular augmented reality display: Impact of visual characteristics on performance. *PLOS ONE*, 16(9):1–24, 09 2021.
- [10] S. Benedetto, A. Carbone, V. Draï-Zerbib, M. Pedrotti, and T. Baccino. Effects of luminance and illuminance on visual fatigue and arousal during digital reading. *Computers in Human Behavior*, 41:112–119, 2014.
- [11] M. Bertalmío. Chapter 5 - Brightness perception and encoding curves. In M. Bertalmío, editor, *Vision Models for High Dynamic Range and Wide Colour Gamut Imaging*, Computer Vision and Pattern Recognition, pages 95–129. Academic Press, Jan. 2020.
- [12] V. A. Billock, G. A. Gleason, and B. H. Tsou. Perception of forbidden colors in retinally stabilized equiluminant images: an indication of softwired cortical color opponency? *J. Opt. Soc. Am. A*, 18(10):2398–2403, Oct 2001.
- [13] O. Bimber and R. Raskar. *Spatial augmented reality: merging real and virtual worlds*. CRC press, 2005.
- [14] R. Blake. A primer on binocular rivalry, including current controversies. *Brain and mind*, 2(1):5–38, 2001.
- [15] C. Blehm, S. Vishnu, A. Khattak, S. Mitra, and R. W. Yee. Computer Vision Syndrome: A Review. *Survey of Ophthalmology*, 50(3):253–262, 2005.

- [16] E. J. Borsting, M. W. Rouse, G. L. Mitchell, M. Scheiman, S. A. Cotter, J. Cooper, M. T. Kulp, R. London, and C. Group. Validity and Reliability of the Revised Children Aged 9 to 18 Years. *Optometry & Vision Science*, 80(12):832–838, 2003.
- [17] M. P. Browne, M. Winterbottom, and R. E. Patterson. Performance and comfort of monocular head-mounted displays in flight simulators. In P. L. Marasco and P. R. Havig, editors, *Head- and Helmet-Mounted Displays XV: Design and Applications*, volume 7688, pages 99 – 106. International Society for Optics and Photonics, SPIE, 2010.
- [18] A. Buchner and N. Baumgartner. Text – background polarity affects performance irrespective of ambient illumination and colour contrast. *Ergonomics*, 50(7):1036–1063, 2007.
- [19] A. Buchner, S. Mayr, and M. Brandt. The advantage of positive text-background polarity is due to high display luminance. *Ergonomics*, 52(7):882–886, 2009.
- [20] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. Leonard. Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, 32(6):1309–1332, 2016.
- [21] C. Cajochen, S. Frey, D. Anders, J. Späti, M. Bues, A. Pross, R. Mager, A. Wirz-Justice, and O. Stefani. Evening exposure to a light-emitting diodes (LED)-backlit computer screen affects circadian physiology and cognitive performance. *Journal of Applied Physiology*, 110(5):1432–1438, 2011.
- [22] F. W. Campbell and K. Durden. The Visual Display Terminal Issue: A Consideration of Its Physiological, Psychological and Clinical Background. *Ophthalmic and Physiological Optics*, 3(2):175–192, 1983.

- [23] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic. Augmented reality technologies, systems and applications. *Multimedia tools and applications*, 51:341–377, 2011.
- [24] C. Casco, A. Grieco, E. Giora, and M. Martinelli. Saliency from orthogonal velocity component in texture segregation. *Vision Research*, 46(6):1091–1098, 2006.
- [25] A. H. S. Chan and P. S. K. Lee. Effect of display factors on Chinese reading times, comprehension scores and preferences. *Behaviour and Information Technology*, 24:81–91, 2005.
- [26] M. L. Chenechal, T. Duval, V. Gouranton, J. Royan, and B. Arnaldi. Vishnu: virtual immersive support for helping users an interaction paradigm for collaborative remote guiding in mixed reality. In *2016 IEEE Third VR International Workshop on Collaborative Virtual Environments (3DCVE)*, pages 9–12, 2016.
- [27] Z. Choudhary, A. Erickson, N. Norouzi, K. Kim, G. Bruder, and G. Welch. Virtual big heads in extended reality: Estimation of ideal head scales and perceptual thresholds for comfort and facial cues. *ACM Trans. Appl. Percept.*, 20(1), jan 2023.
- [28] H. D. Crane and T. P. Piantanida. On seeing reddish green and yellowish blue. *Science*, 221(4615):1078–1080, 1983.
- [29] S. de Cunsel. Evaluation of augmented reality (AR) displays performance based on human visual perception. In B. C. Kress and C. Peroz, editors, *Digital Optics for Immersive Displays II*, volume 11350, pages 47 – 53. International Society for Optics and Photonics, SPIE, 2020.
- [30] S. Debernardis, M. Fiorentino, M. Gattullo, G. Monno, and A. E. Uva. 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):125–139, 2014.
- [31] S. Demirel, R. S. Anderson, S. C. Dakin, and L. N. Thibos. Detection and resolution of vanishing optotype letters in central and peripheral vision. *Vision Research*, 59:9–16, 2012.
- [32] P. G. Devine. Stereotypes and prejudice: Their automatic and controlled components. *Journal of personality and social psychology*, 56(1):5–18, 1989.
- [33] P. G. Devine and A. J. Elliot. Are racial stereotypes really fading? the princeton trilogy revisited. *Personality and social psychology bulletin*, 21(11):1139–1150, 1995.
- [34] A. Donval, N. Gross, E. Partouche, I. Dotan, O. Lipman, and M. Oron. Smart filters: operational HMD even at bright sunlight conditions. *Display Technologies and Applications for Defense, Security, and Avionics VIII; and Head- and Helmet-Mounted Displays XIX*, 9086:90860T, 2014.
- [35] M. Doroodchi, P. Ramos, A. Erickson, H. Furuya, J. Benjamin, G. Bruder, and G. F. Welch. Effects of optical see-through displays on self-avatar appearance in augmented reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 352–356. IEEE, 2022.
- [36] P. Ekman, W. V. Friesen, M. O’sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, et al. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712–717, 1987.
- [37] S. Engmann, B. M. ’t Hart, T. Sieren, S. Onat, P. König, and W. Einhäuser. Saliency on a natural scene background: Effects of color and luminance contrast add linearly. *Attention, Perception, & Psychophysics*, 71(6):1337–1352, Aug. 2009.

- [38] A. Erickson, G. Bruder, G. Welch, I. Bynum, T. Peck, and J. Good. Perceived humanness bias in additive light model displays. *Journal of Vision*, 22(14):4201–4201, 2022.
- [39] A. Erickson, G. Bruder, and G. F. Welch. Adapting michelson contrast for use with optical see-through displays. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 409–410. IEEE, 2022.
- [40] A. Erickson, G. Bruder, and G. F. Welch. Analysis of the saliency of color-based dichoptic cues in optical see-through augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 2022.
- [41] A. Erickson, K. Kim, G. Bruder, and G. Welch. Effects of dark mode graphics on visual acuity and fatigue with virtual reality head-mounted displays. In *Proceedings of IEEE Virtual Reality (VR)*, pages 434–442, 2020.
- [42] A. Erickson, K. Kim, G. Bruder, and G. Welch. Exploring the Limitations of Environment Lighting on Optical See-Through Head-Mounted Displays. In *Proceedings of ACM Conference on Spatial User Interfaces (SUI)*, pages 1–8, 2020.
- [43] A. Erickson, K. Kim, G. Bruder, and G. Welch. 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*, pages 1–8, 2020.
- [44] A. Erickson, K. Kim, G. Bruder, and G. F. Welch. A Review of Visual Perception Research in Optical See-Through Augmented Reality. In F. Argelaguet, R. McMahan, and M. Sugimoto, editors, *ICAT-EGVE 2020 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2020.

- [45] A. Erickson, K. Kim, G. Bruder, and G. F. Welch. Dark/Light Mode Adaptation for Graphical User Interfaces on Near-Eye Displays. In A. Kulik, M. Sra, K. Kim, and B.-K. Seo, editors, *ICAT-EGVE 2020 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments - Posters and Demos*. The Eurographics Association, 2020.
- [46] A. Erickson, K. Kim, G. Bruder, and G. F. Welch. Exploring the limitations of environment lighting on optical see-through head-mounted displays. In *Symposium on Spatial User Interaction*, SUI '20, New York, NY, USA, 2020. Association for Computing Machinery.
- [47] A. Erickson, K. Kim, G. Bruder, and G. F. Welch. Beyond visible light: User and societal impacts of egocentric multispectral vision. In *Virtual, Augmented and Mixed Reality: 13th International Conference, VAMR 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings*, pages 317–335. Springer, 2021.
- [48] A. Erickson, K. Kim, A. Lambert, G. Bruder, M. P. Browne, and G. F. Welch. An extended analysis on the benefits of dark mode user interfaces in optical see-through head-mounted displays. *ACM Trans. Appl. Percept.*, 18(3), may 2021.
- [49] A. Erickson, K. Kim, R. Schubert, G. Bruder, and G. Welch. Is it cold in here or is it just me? analysis of augmented reality temperature visualization for computer-mediated thermoception. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 202–211. IEEE, 2019.
- [50] A. Erickson, N. Norouzi, K. Kim, J. LaViola Jr., G. Bruder, and G. Welch. Effects of Depth Information on Visual Target Identification Task Performance in Shared Gaze Environments. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1934–1944, 2020.

- [51] A. Erickson, N. Norouzi, K. Kim, R. Schubert, J. Jules, J. J. LaViola Jr, G. Bruder, and G. F. Welch. Sharing gaze rays for visual target identification tasks in collaborative augmented reality. *Journal on Multimodal User Interfaces*, 14(4):353–371, 2020.
- [52] A. Erickson, D. Reiners, G. Bruder, and G. F. Welch. Augmenting human perception: Mediation of extrasensory signals in head-worn augmented reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 373–377. IEEE, 2021.
- [53] S. Etchebehere and E. Fedorovskaya. On the role of color in visual saliency. *Electronic Imaging*, 2017(14):58–63, 2017.
- [54] F. Ferris, A. Kassoff, G. Bresnick, and I. Bailey. New Visual Acuity Charts for Clinical Research. *American Journal of Ophthalmology*, 94(1):91 – 96, 1982.
- [55] M. Fiorentino, S. Debernardis, A. E. Uva, and G. Monno. Augmented Reality Text Style Readability with See-Through Head-Mounted Displays in Industrial Context. *Presence: Teleoperators and Virtual Environments*, 22(2):171–190, 2013.
- [56] S. T. Fiske, A. J. Cuddy, P. Glick, and J. Xu. A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of personality and social psychology*, 82(6):878–902, 2002.
- [57] I. O. for Standardization (ISO). 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, 12 1994.
- [58] I. O. for Standardization (ISO). 8597:1994 and 8596:2009 - Ophthalmic Optics. Visual Acuity Testing. Standard Optotype and its Presentation, 7 2009.

- [59] H. Fuchs, H. Yang, T. Peck, A. Bulysheva, et al. Methods, systems, and computer readable media for image guided ablation, Feb. 23 2016. US Patent 9,265,572.
- [60] B. V. Funt, M. S. Drew, and J. Ho. Color Constancy from Mutual Reflection. *International Journal of Computer Vision*, 6(1):5–24, 1991.
- [61] J. Gabbard, D. Mehra, and J. E. Swan. Effects of AR Display Context Switching and Focal Distance Switching on Human Performance. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2228–2241, 2019.
- [62] J. Gabbard, J. E. Swan, and D. Hix. The Effects of Text Drawing Styles, Background Textures, and Natural Lighting on Text Legibility in Outdoor Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 15(1):16–32, 2006.
- [63] J. Gabbard, J. E. Swan, D. Hix, S.-J. Kim, and G. Fitch. Active Text Drawing Styles for Outdoor Augmented Reality: A User-Based Study and Design Implications. In *Proceedings of IEEE Virtual Reality*, pages 35–42, Charlotte, 2007. IEEE.
- [64] J. Gabbard, J. E. Swan, J. Zedlitz, and W. W. Winchester. 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)*, pages 79–86, 2010.
- [65] J. L. Gabbard, G. M. Fitch, and H. Kim. Behind the Glass: Driver Challenges and Opportunities for AR Automotive Applications. *Proceedings of the IEEE*, 102(2):124–136, 2014.
- [66] J. L. Gabbard, M. Smith, C. Merenda, G. Burnett, and D. R. Large. A perceptual color-matching method for examining color blending in augmented reality head-up display graphics. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–1, 2020.

- [67] M. Gattullo, A. E. Uva, M. Fiorentino, and G. Monno. Effect of Text Outline and Contrast Polarity on AR Text Readability in Industrial Lighting. *IEEE Transactions on Visualization and Computer Graphics*, 21(5):638–651, 2015.
- [68] A. C. S. Genay, A. Lecuyer, and M. Hachet. Being an avatar ”for real”: a survey on virtual embodiment in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, pages 1–21, 2021.
- [69] P. A. Goff, J. L. Eberhardt, M. J. Williams, and M. C. Jackson. Not yet human: implicit knowledge, historical dehumanization, and contemporary consequences. *Journal of personality and social psychology*, 94(2):292–306, 2008.
- [70] M. Gonzalez-Franco and T. C. Peck. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74, 2018.
- [71] R. Grasset, P. Lamb, and M. Billinghurst. Evaluation of mixed-space collaboration. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR’05)*, pages 90–99, 2005.
- [72] J. Grudin. Rivalry and interference with a head-mounted display. *ACM Trans. Comput.-Hum. Interact.*, 9(3):238–251, sep 2002.
- [73] A. G. Halberstadt, V. L. Castro, Q. Chu, F. T. Lozada, and C. M. Sims. Preservice teachers’ racialized emotion recognition, anger bias, and hostility attributions. *Contemporary Educational Psychology*, 54:125–138, 2018.
- [74] T. Harding, J. Martin, and C. Rash. Using a helmet-mounted display computer simulation model to evaluate the luminance requirements for symbology. *Helmet- and Head-Mounted Displays X: Technologies and Applications*, 5800:159–168, 2005.

- [75] T. Harding, C. Rash, M. Lattimore, J. Statz, and J. Martin. Perceptual issues for color helmet-mounted displays: luminance and color contrast requirements. In *Degraded Visual Environments: Enhanced, Synthetic, and External Vision Solutions*, volume 9839, page 98390E, Baltimore, 2016. SPIE.
- [76] C. Harms and F. Biocca. Internal consistency and reliability of the networked minds measure of social presence. In *Seventh annual international workshop: Presence*, volume 2004. Universidad Politecnica de Valencia Valencia, Spain, 2004.
- [77] S. Higuchi, Y. Motohashi, Y. Liu, M. Ahara, and Y. Kaneko. Effects of VDT tasks with a bright display at night on melatonin, core temperature, heart rate, and sleepiness. *Journal of Applied Physiology*, 94(5):1773–1776, 2015.
- [78] J. Hincapié-Ramos, L. Ivanchuk, S. K. Sridharan, and P. P. Irani. SmartColor: Real-time color and contrast correction for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 21(12):1336–1348, 2015.
- [79] C. Ho and K. F. MacDorman. Measuring the uncanny valley effect. *International Journal of Social Robotics*, 9(1):129–139, 2017.
- [80] J. T. Holladay. Visual acuity measurements. *Journal of Cataract & Refractive Surgery*, 30(2):287–290, Feb. 2004.
- [81] P. Hsieh and P. U. Tse. Illusory color mixing upon perceptual fading and filling-in does not result in ‘forbidden colors’. *Vision Research*, 46(14):2251–2258, 2006.
- [82] L. Huang and H. Pashler. Quantifying object salience by equating distractor effects. *Vision Research*, 45(14):1909–1920, 2005.
- [83] K. Hugenberg and G. V. Bodenhausen. Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science*, 14(6):640–643, 2003.

- [84] Y. Itoh, M. Dzitsiuk, T. Amano, and G. Klinker. Semi-Parametric Color Reproduction Method for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, 21(11):1269–1278, 2015.
- [85] Y. Itoh, T. Langlotz, D. Iwai, K. Kiyokawa, and T. Amano. Light Attenuation Display: Subtractive See-Through Near-Eye Display via Spatial Color Filtering. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):1951–1960, 2019.
- [86] Y. Itoh, T. Langlotz, J. Sutton, and A. Plopski. Towards indistinguishable augmented reality: A survey on optical see-through head-mounted displays. *ACM Comput. Surv.*, 54(6), jul 2021.
- [87] K. Jacoby. Slaves by nature? domestic animals and human slaves. *Slavery and Abolition*, 15(1):89–99, 1994.
- [88] J. Jia, Y. Zhang, X. Wu, and W. Guo. Image-based label placement for augmented reality browsers. In *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, pages 1654–1659, San Francisco, 2018. IEEE.
- [89] C. A. Johnson and E. J. Casson. Effects of luminance, contrast, and blur on visual acuity. *Optometry and vision science*, 72:864–869, 1995.
- [90] J. A. Jones, J. E. Swan, G. Singh, and S. R. Ellis. Peripheral visual information and its effect on distance judgments in virtual and augmented environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization - APGV '11*, page 29, 2011.
- [91] T. Kaminokado, Y. Hiroi, and Y. Itoh. Stainedview: Variable-intensity light-attenuation display with cascaded spatial color filtering for improved color fidelity. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3576–3586, 2020.

- [92] S. Kamkar, H. A. Moghaddam, and R. Lashgari. Early visual processing of feature saliency tasks: A review of psychophysical experiments. *Frontiers in Systems Neuroscience*, 12:54, 2018.
- [93] K. Kiltner, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.
- [94] K. Kim, A. Erickson, A. Lambert, G. Bruder, and G. Welch. Effects of dark mode on visual fatigue and acuity in optical see-through head-mounted displays. In *Symposium on Spatial User Interaction*, SUI '19, New York, NY, USA, 2019. Association for Computing Machinery.
- [95] K. Kim, A. Erickson, A. Lambert, G. Bruder, and G. Welch. Effects of dark mode on visual fatigue and acuity in optical see-through head-mounted displays. In *Proceedings of the ACM Symposium on Spatial User Interaction (SUI)*, pages 1–9, 2019.
- [96] K. Kiyokawa, H. Iwasa, H. Takemura, and N. Yokoya. Collaborative immersive workspace through a shared augmented environment. In B. Gopalakrishnan and S. Murugesan, editors, *Intelligent Systems in Design and Manufacturing*, volume 3517, pages 2 – 13. International Society for Optics and Photonics, SPIE, 1998.
- [97] T. R. Knapp. Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing research*, 39(2):121–123, 1990.
- [98] C. Kniestedt and R. L. Stamper. Visual acuity and its measurement. *Ophthalmology Clinics of North America*, 16(2):155–170, 2003.
- [99] F. L. Kooi and A. Toet. Visual comfort of binocular and 3D displays. *Displays*, 25(2–3):99–108, 2004.

- [100] A. Krekhov, S. Cmentowski, A. Waschk, and J. Krüger. Deadeye visualization revisited: Investigation of preattentiveness and applicability in virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):547–557, 2020.
- [101] A. Krekhov and J. Krüger. Deadeye: A Novel Preattentive Visualization Technique Based on Dichoptic Presentation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):936–945, Jan. 2019. Conference Name: IEEE Transactions on Visualization and Computer Graphics.
- [102] K. Krösl, D. Bauer, M. Schwärzler, H. Fuchs, G. Suter, and M. Wimmer. 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):911–923, 2018.
- [103] W. M. Kuzon Jr, M. G. Urbanchek, and S. McCabe. The seven deadly sins of statistical analysis. *Annals of plastic surgery*, 37(3):265–272, 1996.
- [104] M. Lambooi, M. Fortuin, W. Ijsselstein, B. Evans, and I. Heynderickx. Measuring visual fatigue and visual discomfort associated with 3-D displays. *Journal of the Society for Information Display*, 18(11):931–943, 2010.
- [105] T. Langlotz, M. Cook, and H. Regenbrecht. Real-time radiometric compensation for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(11):2385–2394, 2016.
- [106] J. Lee, T. Zhan, and S.-T. Wu. Prospects and challenges in augmented reality displays. *Virtual Reality & Intelligent Hardware*, 1:77, Feb. 2019.
- [107] S. Lee and J. Kim. An experimental study on the impacts of luminance contrast upon readability in vdt environments. *Journal of the Ergonomics Society of Korea*, 26(2):21–33, 2007.

- [108] A. Leykin and M. Tuceryan. Automatic determination of text readability over textured backgrounds for augmented reality systems. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 224–230, Arlington, 2004. IEEE.
- [109] M. Lhommet and S. C. Marsella. Expressing emotion through posture. *The Oxford handbook of affective computing*, 273:273–285, 2014.
- [110] Y. Li, K. Kim, A. Erickson, N. Norouzi, J. Jules, G. Bruder, and G. F. Welch. A scoping review of assistance and therapy with head-mounted displays for people who are visually impaired. *ACM Transactions on Accessible Computing (TACCESS)*, 15(3):1–28, 2022.
- [111] R. J. Little and D. B. Rubin. *Statistical analysis with missing data*, volume 793. John Wiley & Sons, 2019.
- [112] T. D. Little, T. D. Jorgensen, K. M. Lang, and E. W. G. Moore. On the joys of missing data. *Journal of pediatric psychology*, 39(2):151–162, 2014.
- [113] M. Livingston, J. Barrow, and C. Sibley. Quantification of Contrast Sensitivity and Color Perception using Head-worn Augmented Reality Displays. In *Proceedings of IEEE Virtual Reality Conference*, pages 115–122, Lafayette, 2009. IEEE.
- [114] M. Livingston, J. Gabbard, J. Swan, C. Sibley, and J. Barrow. Basic Perception in Head-Worn Augmented Reality Displays. In *Human Factors in Augmented Reality Environments*, pages 35–65. Springer New York, New York, 2013.
- [115] M. A. Livingston, L. J. Rosenblum, D. G. Brown, G. S. Schmidt, S. J. Julier, Y. Baillot, J. E. Swan, Z. Ai, and P. Maassel. Military applications of augmented reality. *Handbook of augmented reality*, pages 671–706, 2011.
- [116] N. K. Logothetis, D. A. Leopold, and D. L. Sheinberg. What is rivalling during binocular rivalry? *Nature*, 380(6575):621–624, Apr. 1996.

- [117] P. Lubos, G. Bruder, O. Ariza, and F. Steinicke. Ambiculus: LED-based Low-resolution Peripheral Display Extension for Immersive Head-mounted Displays. In *Proceedings of the Virtual Reality International Conference (VRIC)*, pages 13:1–13:4, 2016.
- [118] D. Malacara. *Color Vision and Colorimetry: Theory and Applications, Second Edition*. SPIE, 1000 20th Street, Bellingham, WA 98227-0010 USA, July 2011.
- [119] S. Mayr and A. Buchner. After-effects of tft-lcd display polarity and display colour on the detection of low-contrast objects. *Ergonomics*, 53(7):914–925, 2010.
- [120] B. McElree and M. Carrasco. The temporal dynamics of visual search: evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance*, 25(6):1517, 1999.
- [121] C. Merenda, M. Smith, J. Gabbard, G. Burnett, and D. Large. 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)*, pages 1–6, March 2016.
- [122] A. A. Michelson. *Studies in optics*. University Press, Chicago, 1927.
- [123] M. Mori. Bukimi no tani [the uncanny valley]. *Energy*, 7:33–35, 1970.
- [124] N. Norouzi, A. Erickson, K. Kim, R. Schubert, J. LaViola Jr., G. Bruder, and G. Welch. Effects of Shared Gaze Parameters on Visual Target Identification Task Performance in Augmented Reality. In *ACM Symposium on Spatial User Interaction (SUI)*, pages 12:1–12:11, 2019.
- [125] N. Norouzi, K. Kim, G. Bruder, A. Erickson, Z. Choudhary, Y. Li, and G. Welch. A Systematic Literature Review of Embodied Augmented Reality Agents in Head-Mounted Display Environments. In F. Argelaguet, R. McMahan, and M. Sugimoto, editors, *ICAT-EGVE 2020*

- *International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, pages 101–111. The Eurographics Association, 2020.
- [126] H. Nothdurft. Feature analysis and the role of similarity in preattentive vision. *Perception & psychophysics*, 52(4):355–375, 1992.
- [127] H. Nothdurft. Saliency from feature contrast: additivity across dimensions. *Vision research*, 40(10-12):1183–1201, 2000.
- [128] J. Orlosky, K. Kiyokawa, and H. Takemura. Managing mobile text in head mounted displays: studies on visual preference and text placement. *ACM SIGMOBILE Mobile Computing and Communications Review*, 18(2):20–31, 2014. Publisher: ACM New York, NY, USA.
- [129] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, V. Tankovich, C. Loop, Q. Cai, P. A. Chou, S. Menick, J. Valentin, V. Pradeep, S. Wang, S. B. Kang, P. Kohli, Y. Lutchyn, C. Keskin, and S. Izadi. Holoportation: Virtual 3D Teleportation in Real-Time. In *Proceedings of the Symposium on User Interface Software and Technology*, pages 741–754, 2016.
- [130] C. Owsley. Contrast sensitivity. *Ophthalmology Clinics of North America*, 16(2):171–177, 2003.
- [131] T. C. Peck, J. J. Good, A. Erickson, I. Bynum, and G. Bruder. Effects of transparency on perceived humanness: Implications for rendering skin tones using optical see-through displays. *IEEE Transactions on Visualization & Computer Graphics*, 01(01):1–1, feb 2022.
- [132] T. C. Peck, L. E. Sockol, and S. M. Hancock. Mind the gap: The underrepresentation of female participants and authors in virtual reality research. *IEEE transactions on visualization and computer graphics*, 26(5):1945–1954, 2020.

- [133] D. Pelli and P. Bex. Measuring contrast sensitivity. *Vision Research*, 90:10–14, 2013.
- [134] G. P. J. Penkelink and J. Besuijen. Chromaticity contrast, luminance contrast, and legibility of text. *Journal of the Society for Information Display*, 4(3):135–144, 1996.
- [135] C. Piepenbrock, S. Mayr, and A. Buchner. Positive display polarity is particularly advantageous for small character sizes: Implications for display design. *Human Factors*, 56(5):942–951, 2014.
- [136] C. Piepenbrock, S. Mayr, and A. Buchner. Smaller pupil size and better proofreading performance with positive than with negative polarity displays. *Ergonomics*, 57(11):1670–1677, 2014.
- [137] C. Piepenbrock, S. Mayr, I. Mund, and A. Buchner. Positive display polarity is advantageous for both younger and older adults. *Ergonomics*, 56(7):1116–1124, 2013.
- [138] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billingham. *Mini-Me: An Adaptive Avatar for Mixed Reality Remote Collaboration*, page 1–13. Association for Computing Machinery, New York, NY, USA, 2018.
- [139] C. Poynton. Frequently asked questions about gamma. *Rapport Technique, janvier*, 152, 1998.
- [140] R. T. Pramod and S. P. Arun. Features in visual search combine linearly. *Journal of Vision*, 14(4):6–6, 04 2014.
- [141] L. A. Rudman and K. Mescher. Of animals and objects: Men’s implicit dehumanization of women and likelihood of sexual aggression. *Personality and Social Psychology Bulletin*, 38(6):734–746, 2012.

- [142] R. M. Schiffman, M. D. Christianson, G. Jacobsen, J. D. Hirsch, and B. L. Reis. Reliability and validity of the ocular surface disease index. *Archives of Ophthalmology*, 118(5):615–621, 2000.
- [143] M. Schrepp, A. Hinderks, and J. Thomaschewski. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6):103, 2017.
- [144] K. Scott, D. S. Ma, M. S. Sadler, and J. Correll. A social scientific approach toward understanding racial disparities in police shooting: Data from the department of justice (1980–2000). *Journal of Social Issues*, 73(4):701–722, 2017.
- [145] J. Shen, E. M. Reingold, and M. Pomplun. Guidance of eye movements during conjunctive visual search: the distractor-ratio effect. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 57(2):76, 2003.
- [146] E. Shneor and S. Hochstein. Eye dominance effects in feature search. *Vision research*, 46(25):4258–4269, 2006.
- [147] E. Shneor and S. Hochstein. Eye dominance effects in conjunction search. *Vision research*, 48(15):1592–1602, 2008.
- [148] S. Soãres. *Optical Scatter—Techniques and Analysis*, pages 1–31. Springer, Cham, 2019.
- [149] C. Stangor and J. Walinga. 5.2 Seeing. In *Introduction to Psychology - 1st Canadian Edition*. BCcampus, Oct. 2014.
- [150] A. Stimson. Photometry and radiometry for engineers. *New York*, 1974.
- [151] S. Taptagaporn and S. Saito. How display polarity and lighting conditions affect the pupil size of VDT operators. *Ergonomics*, 33:201–208, 1990.

- [152] S. Trawalter, K. M. Hoffman, and A. Waytz. Racial bias in perceptions of others' pain. *PloS one*, 7(11):e48546, 2012.
- [153] A. Treisman and G. A. Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12:97–136, 1980.
- [154] M. Turatto and G. Galfano. Color, form and luminance capture attention in visual search. *Vision Research*, 40(13):1639–1643, 2000.
- [155] S. Utzig, R. Kaps, S. M. Azeem, and A. Gerndt. Augmented reality for remote collaboration in aircraft maintenance tasks. In *2019 IEEE Aerospace Conference*, pages 1–10, 2019.
- [156] M. Venkatesan, H. Mohan, J. R. Ryan, C. M. Schürch, G. P. Nolan, D. H. Frakes, and A. F. Coskun. Virtual and augmented reality for biomedical applications. *Cell reports medicine*, 2(7):100348, 2021.
- [157] G. T. Viki, D. Osgood, and S. Phillips. Dehumanization and self-reported proclivity to torture prisoners of war. *Journal of Experimental Social Psychology*, 49(3):325–328, 2013.
- [158] J. Wagemans, J. Feldman, S. Gepshtein, R. Kimchi, J. R. Pomerantz, P. A. Van der Helm, and C. Van Leeuwen. A century of gestalt psychology in visual perception: Ii. conceptual and theoretical foundations. *Psychological bulletin*, 138(6):1218, 2012.
- [159] K. Waldow, A. Fuhrmann, and S. M. Grünvogel. Investigating the effect of embodied visualization in remote collaborative augmented reality. In P. Bourdot, V. Interrante, L. Nedel, N. Magnenat-Thalmann, and G. Zachmann, editors, *Virtual Reality and Augmented Reality*, pages 246–262, Cham, 2019. Springer International Publishing.
- [160] H. K. Walker, W. D. Hall, and J. W. Hurst. *Clinical Methods: The History, Physical, and Laboratory Examinations*. Butterworths, Boston, 1990.

- [161] A.-H. Wang, J.-J. Fang, and C.-H. Chen. 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*, 32:93–104, 2003.
- [162] C. Ware. *Information visualization: perception for design*. Morgan Kaufmann, 2019.
- [163] J. M. Wolfe and S. L. Franzel. Binocularity and visual search. *Perception & Psychophysics*, 44(1):81–93, Jan. 1988.
- [164] B. Yoon, H. Kim, G. A. Lee, M. Billinghurst, and W. Woo. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 547–556, 2019.
- [165] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 547–556, 2019.
- [166] L. Zhang and M. J. Murdoch. Perceived transparency in optical see-through augmented reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 115–120, 2021.
- [167] Y. Zhao, M. Hu, S. Hashash, and S. Azenkot. 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*, pages 4170–4181, 2017.
- [168] Y. Zhao, S. Szpiro, and S. Azenkot. ForeSee: A Customizable Head-Mounted Vision Enhancement System for People with Low Vision. *Proceedings of the International ACM SIGACCESS Conference on Computers & Accessibility*, pages 239–249, 2015.
- [169] B. Zou, I. S. Utochkin, Y. Liu, and J. M. Wolfe. Binocularity and visual search—revisited. *Attention, Perception, & Psychophysics*, 79(2):473–483, 2017.