Supra-threshold Contrast Perception in Augmented Reality

DONGYEON KIM, University of Cambridge, UK
MALIHA ASHRAF, University of Cambridge, UK
ALEXANDRE CHAPIRO, Reality Labs, Meta, USA
RAFAŁ K. MANTIUK, University of Cambridge, UK

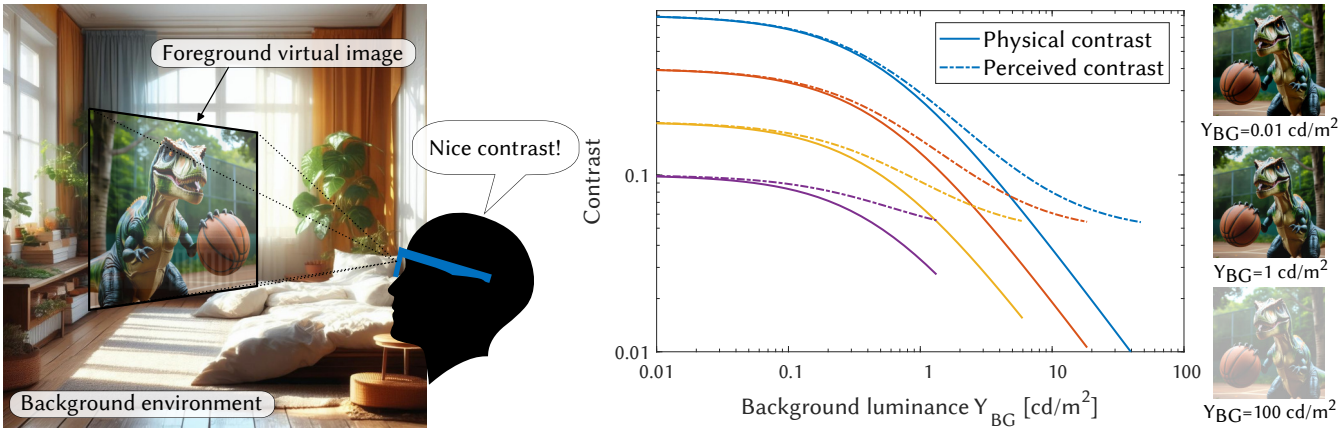

Fig. 1. **Supra-threshold contrast perception in optical see-through augmented reality.** The light from the environment severely reduces the physical contrast of an AR display. This is shown as continuous lines (colors indicating different reference contrast levels) in the plot. These lines drop as the luminance of the background environment increases when a 100 cd/m^2 luminance AR image is displayed. Yet, the perceived contrast modeled in this work, shown as the dotted lines in the plot, is higher than expected. We show that this effect can be explained by supra-threshold models of contrast perception. The images were generated with Stable Diffusion.

When an image is seen on an optical see-through augmented reality (AR) display, the light from the display is mixed with the background light from the environment. This can severely limit the available contrast in AR, which is often orders of magnitude below that of traditional displays. Yet, the presented images appear sharper and show more details than the reduction in physical contrast would indicate. In this work, we hypothesize two effects that are likely responsible for the enhanced perceived contrast in AR: background discounting, which allows observers focused on the display plane to partially discount the light from environment; and supra-threshold contrast perception, which explains the differences in contrast perception across luminance levels. In a series of controlled experiments on AR high-dynamic-range multi-focal haploscope testbed, we found no statistical evidence supporting the effect of background discounting on contrast perception. Instead, the increase of visibility in AR is better explained with models of supra-threshold contrast perception. Our findings can be generalized to incorporate an image input and this model serves to design better algorithms and hardware for display systems affected by additive light, such as AR.

Authors' Contact Information: Dongyeon Kim, dk721@cam.ac.uk, University of Cambridge, Cambridge, UK; Maliha Ashraf, ma905@cam.ac.uk, University of Cambridge, Cambridge, UK; Alexandre Chapiro, alex@chapiro.net, Reality Labs, Meta, Sunnyvale, USA; Rafał K. Mantiuk, rafal.mantiuk@cl.cam.ac.uk, University of Cambridge, Cambridge, UK.

SA Conference Papers '25, Hong Kong, Hong Kong

ACM ISBN 979-8-4007-2137-3/2025/12
https://doi.org/10.1145/3757377.3763824

CCS Concepts: • **Computing methodologies** → **Perception**; **Mixed / augmented reality**; • **Human-centered computing** → *Displays and imagers.*

Additional Key Words and Phrases: optical see-through augmented reality, contrast perception, contrast matching, human visual system, scission

ACM Reference Format:
Dongyeon Kim, Maliha Ashraf, Alexandre Chapiro, and Rafał K. Mantiuk. 2025. Supra-threshold Contrast Perception in Augmented Reality. In *SIGGRAPH Asia 2025 Conference Papers (SA Conference Papers '25), December 15–18, 2025, Hong Kong, Hong Kong*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3757377.3763824

1 Introduction

Optical see-through augmented reality (AR) displays bring advantages in terms of natural real-world interactions and immersion. However, the content shown on these displays is heavily affected by light from the environment, which is mixed in an additive manner with the displayed image. As a result, the physical contrast of the content is compromised (Fig. 1). To address this, several ambient dimming approaches have been introduced (see Sec. S1), incorporating active layers [Cakmakci et al., 2004, Kiyokawa et al., 2000, Magic Leap, Inc.] and static filters, such as sunglasses [Microsoft Corporation]. While effective, these solutions increase design complexity and reduce flexibility. An AR display with good visibility without any dimming requires a strong display engine and high-throughput optics, which significantly increases power demands, posing challenges for wearable devices [Chen et al., 2024].

Recent works have challenged the validity of traditional appearance and contrast vision models in AR settings. Notably, studies on brightness-matching [Murdoch, 2020] and artifact visibility [Chapiro et al., 2024] have hypothesized that the background may be subject to a perceptual discounting effect (sometimes termed scission) when users are focused on the virtual content. Here, we aim to test these new hypotheses in the context of supra-threshold contrast perception. To do this, we conducted three psychophysical experiments using a high-dynamic-range multi-focal haploscope testbed. First, we did not find the effect of contrast discounting when contrast was matched across displays affected by different amounts of ambient light. Second, we found that the observed differences in contrast perception can be explained by the models of supra-threshold contrast perception, which attribute the effect to the raise of contrast sensitivity with luminance. Finally, in the third experiment, we validated that our findings for Gabor patches generalize to complex images.

Combining these learnings, our generalized contrast matching model can accurately predict the visibility of complex images in AR. We demonstrate the utility of the model with an application: an auto-brightness method that preserves the perceived contrast across the luminance levels. While our focus in this work is AR, these models are general and can be applied to any display affected by additive ambient light, such as a projector or an emissive display affected by reflections.

The main contributions of our work are:

- Tested two competing hypotheses explaining AR contrast perception: background discounting and supra-threshold contrast perception,
- Evaluated supra-threshold contrast visibility models for both achromatic and color conditions, and identified the one that best explains both existing and newly collected data,
- Proposed a new contrast matching model generalizing to complex images, supported by a validation study, and,
- Demonstrated an application of our model for auto-brightness.

The code, data, and the supplementary document can be found at the project page[1].

2 Contrast in AR

AR displays combine display color with light from the environment, reducing the physical contrast of the display and detail visibility in bright ambient settings. Below, we model this phenomenon and then introduce the theories of background discounting.

We will assume Michelson's definition of contrast. The base contrast, as seen on an AR display in a completely dark room, is:

$$c_{\text{base}} = \frac{Y_{\text{max}} - Y_{\text{min}}}{Y_{\text{max}} + Y_{\text{min}}}, \tag{1}$$

where, Y_{max} and Y_{min} are the maximum and minimum luminance of a contrast pattern, such as a sinusoidal grating. If the same display is used in a bright environment, which introduces background luminance Y_{BG}, the contrast on the display becomes:

$$c_{\text{AR}} = \frac{(Y_{\text{max}} + Y_{\text{BG}}) - (Y_{\text{min}} + Y_{\text{BG}})}{Y_{\text{max}} + Y_{\text{min}} + 2\,Y_{\text{BG}}} = \frac{Y_{\text{max}} - Y_{\text{min}}}{Y_{\text{max}} + Y_{\text{min}} + 2\,Y_{\text{BG}}}. \tag{2}$$

To simplify notation, we will introduce the mean luminance of the foreground (AR display) and denote $Y_{\text{FG}} = 1/2(Y_{\text{max}} + Y_{\text{min}})$. Then, the physical contrast loss due to background luminance can be expressed as:

$$\frac{c_{\text{AR}}}{c_{\text{base}}} = \frac{Y_{\text{FG}}}{Y_{\text{FG}} + Y_{\text{BG}}}. \tag{3}$$

The loss of physical contrast due to background luminance, Y_{BG}, is shown as continuous lines in the plot in Fig. 1.

3 Hypothesis 1: Background discounting

AR displays provide a transparent overlay over real environments. When a surface is seen as transparent, it is interpreted differently from a simple mixture of foreground and background colors. It is believed that the visual system can decompose an image into multiple contributions [Singh and Anderson, 2002], such as foreground AR image and the background of the environment. Indeed, several works on the perception of brightness [Murdoch, 2020] and chromaticity [Zhang et al., 2021] in AR found discrepancies between matching data and simple color mixture. Zhang et al. [2021] wrote that "the AR as an additive medium is perceived in a different way than traditional media". They further show that their measurements can be explained with a simple model:

$$C_{\text{eff}} = \alpha C_{\text{FG}} + \beta C_{\text{BG}}, \tag{4}$$

where α and β are weighting scalars for the foreground (C_{FG}) and background (C_{BG}) colors in a linear color space (e.g., CIE XYZ). Their color matching results showed a bias towards foreground color ($\alpha > \beta$). Murdoch [2020] found similar bias in a brightness matching task, but only when the foreground overlay was smaller than the background. More recently, a study on the perception of display artifacts in AR [Chapiro et al., 2024] also indicated that the visual system may discount the real environment background when judging the quality of the displayed content.

By incorporating Eq. (4) to Eq. (3) and introducing $\gamma = \beta/\alpha$, we can express the physical contrast with the foreground/background discounting effect as

$$c_{\text{AR}} = \frac{Y_{\text{FG}}}{Y_{\text{FG}} + \gamma Y_{\text{BG}}} c_{\text{base}}. \tag{5}$$

As this potential effect can explain the change in perceived contrast, we first test the hypothesis of foreground/background discounting ($\gamma \neq 1$), i.e. that contrast perception in AR displays differs from that in regular single-plane displays. If this holds true, the contrast on an AR display may appear greater than the physical contrast resulting from optical blending of AR content with the environment.

3.1 Experiment 1: Background discounting

To test the hypothesis of AR background discounting, we ran a study where observers matched the single-plane contrast of a simulated AR display to true AR contrast shown on two focal planes.

[1]Project page: https://www.cl.cam.ac.uk/research/rainbow/projects/ar_contrast/
Code: https://github.com/gfxdisp/ar_contrast_perception

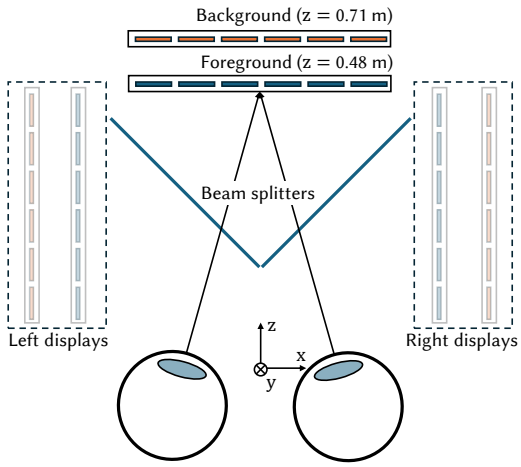

Fig. 2. **Schematic illustration of the dual-layer haploscope.** The haploscope consists of the displays for the left eye and the right eye. The distances are not in scale.

Apparatus of optical see-through AR. Conducting a psychophysical study on commercially available optical see-through AR HMDs is challenging because of confounding factors introduced by optical artifacts (caused by waveguides, diffractive combiners, etc.) To circumvent this issue, we employed a custom-built high-dynamic-range (HDR) haploscope with two focal planes, similar to [Zhong et al., 2021]. Figure 2 illustrates the system setup, designed to approximate the AR system. Foreground displays, optically floated via beam splitters, functioned as AR overlays, while background displays simulated a real environment. The axial distances of the displays were 0.48 m (2.1 D) and 0.71 m (1.4 D), creating a dioptric gap of 0.7 D. This depth difference aligned with the depth of field in stereo displays with multiple planes [Akeley et al., 2004] and the edge of the zone of comfort [Shibata et al., 2011].

Each display plane produced an HDR image by combining a 4K LCD panel (15.6" IPS 3840×2160, LQ156D1JX02) without a backlight and a diffuser screen illuminated with a 2K DLP projector (Acer P5535) as a spatially modulated white backlight. A Fresnel lens was placed between the screen and the LCD to efficiently direct the light beam toward the eye. To eliminate moiré artifacts (common in stacked LCD setups [Matsuda et al., 2022]), diffusers (Acal BFi, L5P1 and L15P1) were placed at the front and back of the Fresnel lens. The display planes were combined using a 50R/50T beam splitter for each eye. Photometric calibration was performed for each display plane using a spectroradiometer (Specbos 1211) positioned at the eye. For the detailed scheme of the testbed, refer to Fig. S6 in the supplementary.

Stimuli and procedure. To test the hypothesis that contrast reproduced in AR differs from that on a single plane, we conducted a contrast matching experiment. The stimulus consisted of reference and test patterns, shown in randomized order — see Fig. 3(A). The reference pattern was shown on two planes of the display: a Gabor patch on the foreground and one of the background patterns on the other plane (see "Conditions" below). The test pattern contained an optical simulation of the reference stimulus and was shown only on the foreground plane. Observers adjusted the test contrast using a knob to match the contrast shown in the reference pattern, with contrast modulated in logarithmic steps.

To ensure observers matched contrast rather than brightness, the test and reference were presented at different luminance levels. The mean luminance of the test pattern was fixed at 30 cd/m^2, while the combined luminance of the reference (foreground + background) was either 60 cd/m^2 or 15 cd/m^2 (1 stop difference), randomly selected in each trial. The randomization was counterbalanced to minimize the impact of luminance on contrast perception (discussed in Sec. 4). The order of conditions and the orientation of the Gabor patches (vertical or horizontal) were randomized. In addition, the initial test contrast value was also counterbalanced with a half or double value of the physical contrast. The stimulus parameters are listed in Table 1.

Table 1. **Specifications of the stimuli used in Experiment 1.** Pixel-per-degree (ppd) denotes a unit of display resolution, and cycle-per-degree (cpd) denotes spatial frequency.

Specification	Value
Display resolution	3840 × 2160 (H × V) FG: 93 ppd; BG: 138 ppd
Stimuli size	8° × 4.5°
Spatial frequency (cpd in Gabor patch)	4 cpd
Envelope of Gabor patch (σ)	1° × 1°
Reference foreground luminance (Y_{FG})	30 cd/m^2 or 7.5 cd/m^2
Reference background luminance (Y_{BG})	30 cd/m^2 or 7.5 cd/m^2
Test luminance (Y_{eff})	30 cd/m^2

Conditions. To represent a range of AR scenarios, we tested five conditions, four of which are depicted in Fig. 3(B) (see supplementary Table S2 for more detail). The control condition, single-plane (SP), displayed a Gabor on a uniform background on the single foreground focal plane for both test and reference stimuli, differing only in luminance. This condition ensured that observers could perform the matching task using only Gabor stimuli, without the influence of single- or dual-plane configurations. Other conditions varied the reference background: dual-flat (DF) used a uniform field, while dual-noise-static (DNS) featured band-limited noise (Fig. 3(B)). To simulate dynamic AR environments, dual-noise-dynamic (DND) introduced movement to the noise. In all these conditions, the test stimulus simulated the optical fusion of foreground and background planes. The final condition, dual-noise-pinhole (DNP), removed defocus blur from the test stimulus to highlight its importance.

Optical fusion of the test pattern. Our goal was to present a single-plane test stimulus that optically matched the two-plane reference, differing only in luminance. In DNS and DND, we simulated defocus blur assuming eye accommodation at the foreground plane and diplopia caused by interpupillary and dioptric distance (details in supplementary Sec. S2.3). Other conditions did not require defocus blur simulation and used the sum of the foreground and background planes as a test pattern.

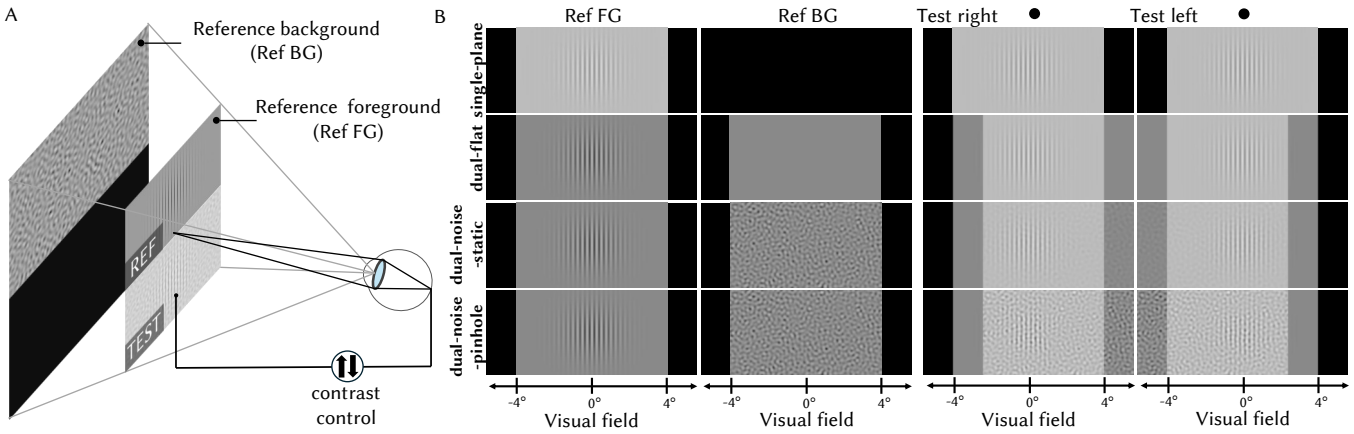

Fig. 3. **Contrast matching experiment in AR.** (A) Participants matched supra-threshold contrast on an AR display (*upper half*, REF) and a simulated single-plane fused image (*bottom half*, TEST). (B) Four experimental conditions from Experiment 1 are shown, excluding dual-noise-dynamic, which involved motion. Reference patterns were displayed on two planes: (first two columns) foreground and background. Test patterns on the foreground plane simulated optical blending, defocus blur, and diplopia of the reference when accommodating on the foreground. The last two columns show right- and left-eye test images (which can be cross-fused). Reference patterns corresponded to the visual field of cyclopean eye, and test patterns matched each eye's visual field. While combined reference luminance was doubled or halved in the experiment, it is shown equal to the test here for better legibility.

Participants. We recruited 17 participants, including 14 naive observers and 3 authors, aged between 14 and 46 (mean age: 28), of whom 7 were female. All participants had normal or corrected-to-normal vision. The participants were screened for stereo vision using Titmus plates (all achieved 60-arcsecond acuity) and for color vision using Ishihara color plates. Participants received compensation for their involvement. Written informed consent was obtained from all participants, and the study was conducted with ethical approval from the host institution.

Training and screening session. Before the main experiments, observers completed a training session to ensure they understood the task of contrast matching. They were shown test and reference grayscale images with six human faces and were instructed to adjust the test image contrast to match the contrast of the reference image. They were instructed to assess the contrast in terms of overall sharpness and the salience of features relative to their respective backgrounds. The test and reference differed in luminance, similar to the patterns in the main experiment, and both were displayed only on the foreground plane. Each observer completed 10 trials. If an observer could not complete the training session with sufficient accuracy, they were disqualified from the main experiment. The results of this session are provided in supplementary Fig. S1.

Defocus blur matching. Defocus blur is often simulated with a fixed pupil size [Cholewiak et al., 2018]. We initially followed this approach and found per-observer pupil size from Watson and Yellott [2012] model (based on stimulus luminance, size, and observer age, binocular adaptation). This unexpectedly produced a stronger blur than what observers perceived in the multi-focal presentation. The discrepancy may be explained by spatial vision calibration to individual retinal blur levels [Sawides et al., 2011], making optically simulated blur—based solely on pupil-size predictions—appear "unnatural" and thus more pronounced.

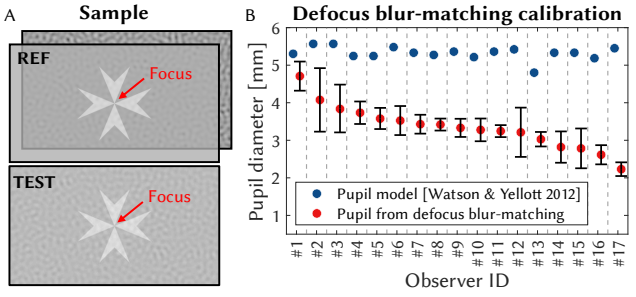

Fig. 4. **Defocus blur-matching calibration.** (A) Stimuli used for the calibration and (B) the comparison between the (*blue*) predicted pupil diameter from the previous work and the (*red*) pupil diameter inversely matched with the calibration. The errorbars denote the standard deviation of the estimated individuals' pupil diameter values.

To customize the defocus blur for each observer, we performed a defocus blur matching experiment before the main session of the experiment. Participants were presented with the test condition identical to DNS as shown in Fig. 4(A), where the grating was replaced with a fixation target. Participants were asked to focus on the center of the fixation target and adjust the blur size of the background in the test pattern to match that of the reference pattern. Each participant performed 10 trials of defocus blur matching, and the median value was used for the blur simulation in the main experiment.

As shown in Fig. 4(B), the pupil diameter predicted by the model [Watson and Yellott, 2012] showed a large discrepancy with the pupil diameter matched in our procedure. A small study was conducted to compare contrast matching when defocus blur was generated using the matched pupil diameter or one predicted by the model. The results showed that observers were unable to match contrast when

the model-predicted pupil was used. Further details are provided in supplementary Sec. S4.

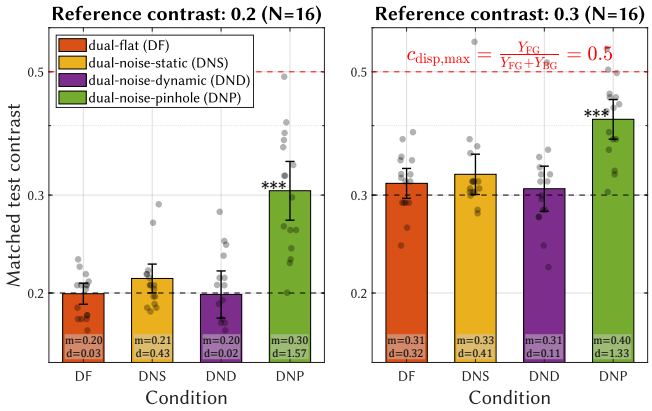

Fig. 5. **Experiment 1 results.** The matched test contrast is compared with the reference contrast of (*left*) 0.2 and (*right*) 0.3 and averaged across 16 observers. The black dots indicate the matched test contrast of individuals, and error bars represent the 95% confidence interval. The dashed lines indicate the physical contrast tested (*black*) and the display's physical contrast limit ($c_{\text{disp,max}}$, *red*), respectively. Each condition is additionally provided with the mean and Cohen's d (*d*). Asterisks indicate the statistical significance (∗ ∗ ∗: p<0.001).

Experiment results. Figure 5 shows the matched test contrast averaged across 16 observers (1 excluded because they failed the matching task in the training session), plotted separately for the reference contrasts of 0.2 and 0.3. Each data point is calculated as the median of log-contrast of ten trials completed by each observer for each condition. These medians are then averaged across all observers.

We found that a portion of naive observers always selected higher test contrast to match the reference (their results were biased towards higher test contrast, details in the supplementary Fig. S5). For that reason, we used the results of the SP condition to compensate for the bias of the individual observer.

To test the background discounting hypothesis, we ran two-tailed t-tests with H_0 that the matched contrast was different from the physical contrast (0.2 or 0.3). This is equivalent to testing for $\gamma \neq 1$ in Eq. (5). To control for multiple comparisons, Holm-Bonferroni correction was applied. We also report Cohen's effect size d in Fig. 5. Full details of the statistical analysis are listed in Table S4 in the supplementary. Only the DNP condition showed a statistically significant difference (p<0.001) and the effect size greater than 1. This demonstrates that defocus blur has a significant effect on contrast perception, and it must be correctly simulated to produce matching stimuli. The small effect sizes for other conditions and p>0.05 indicate that we have no statistical evidence to confirm the existence of the background discounting effect on contrast perception in AR. It must be noted, however, that our findings are applicable to contrast perception and may not generalize to the perception of brightness and uniform color patches, for which significant effects were reported [Murdoch, 2020, Zhang and Murdoch, 2021, Zhang et al., 2021].

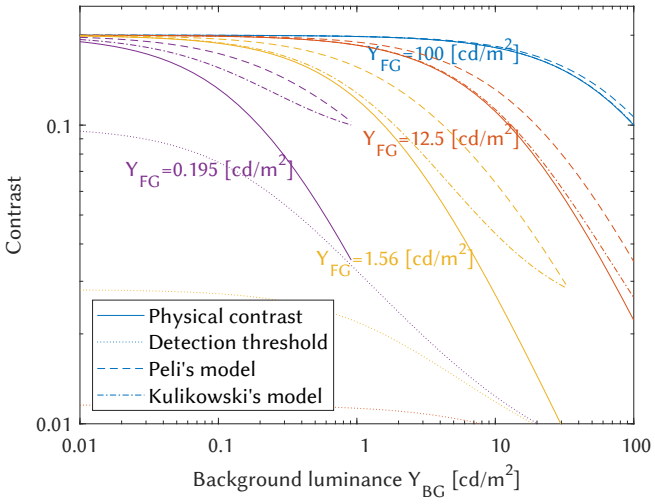

Fig. 6. **The effect of ambient luminance on contrast in AR.** The continuous lines show how the physical contrast is reduced (from c_{base} of 0.2) when the background light is added. The dashed and dot-dashed lines represent matched contrast according to two different models. The dotted lines show the detection thresholds used in the models (for 4 cpd patterns).

4 Hypothesis 2: Supra-threshold contrast perception in AR

With no evidence for the existence of the discounting effect on contrast perception, we focus now on an alternative explanation based on the supra-threshold models of contrast perception. In Sec. 2, we explained how background luminance affects physical contrast in AR. This is shown as continuous lines in Fig. 6, in which physical contrast, c_{AR}, decreases as the background luminance increases, with the strongest reduction at the lowest display luminance levels (e.g., $Y_{\text{FG}} = 0.195\,\text{cd/m}^2$). While Eq. (3) explains the behavior of physical contrast on an AR display, it does not account for perceived contrast. To address this, we introduce two competing models for supra-threshold contrast perception.

Kulikowski [1976] proposed that perceived contrast can be matched across luminance levels by subtracting the threshold contrast for each luminance:

$$c_{\text{test}} - t(Y_{\text{test}}) = c_{\text{ref}} - t(Y_{\text{ref}}), \tag{6}$$

where c_{test} and c_{ref} is the contrast of patterns presented at different luminance levels of Y_{test} and Y_{ref}, and $t(Y)$ are the contrast detection thresholds, which represent the smallest detectable contrast at luminance level Y. Such a contrast matching model has an intuitive explanation: any contrast under $t(Y)$ is below the noise threshold of the visual system. As there is no advantage in preserving noise, the visual system "subtracts" the amplitude of noise from the signal.

Peli [1990] demonstrated that their contrast matching data is better explained by a multiplicative relationship:

$$\frac{c_{\text{test}}}{t(Y_{\text{test}})} = \frac{c_{\text{ref}}}{t(Y_{\text{ref}})}, \tag{7}$$

where the symbols are the same as in Eq. (6) above. This relationship can be explained by the gain control found at the different stages of visual processing [Bonin et al., 2005].

Now, we can combine the physical contrast in AR from Eq. (3) with either Eq. (6) or Eq. (7) to find the lines of matching contrast across background luminance levels. Assuming that our base contrast, c_{base}, is shown at luminance Y_{FG} on an AR display affected by the background luminance Y_{BG}, its equivalent test contrast seen on the same display but with no background light is:

$$c_{\text{test}} = c_{\text{base}} \frac{Y_{\text{FG}}}{Y_{\text{FG}} + Y_{\text{BG}}} - t(Y_{\text{FG}} + Y_{\text{BG}}) + t(Y_{\text{FG}}), \tag{8}$$

assuming Kulikowski's contrast matching model from Eq. (6) and that $c_{\text{ref}} = c_{\text{AR}}$ (see Eq. (3)). A similar equation can be derived for Peli's model from Eq. (7). Matching contrast lines, estimated via both models, are shown as dashed and dot-dashed lines in Fig. 6. Fig. 1 shows similar lines but for the variation in c_{base} rather than Y_{FG}, and only for Kulikowski's model. Detection thresholds were predicted using castleCSF [Ashraf et al., 2024] and are shown as dotted lines in Fig. 6. Dashed lines of matching contrast in Fig. 6 indicate that perceived contrast in AR appears higher than physical contrast (continuous lines) due to increased luminance. As luminance rises, detection thresholds fall (dotted lines in Fig. 6, $t(Y_{\text{FG}} + Y_{\text{BG}}) < t(Y_{\text{FG}})$), leading to higher perceived contrast under both models.

Besides the two models presented above, other models of supra-threshold contrast perception have been proposed [Ashraf and Mantiuk, 2024, Foley and Legge, 1981, Georgeson, 1991]. However, since these are variations of either Kulikowski's or Peli's models, we do not discuss them here. Instead, we evaluate these models in Sec. 4.2. The supra-threshold contrast perception models plotted in Fig. 6 are hypothetical — we do not know if either one can explain contrast perception in AR. We test both models in Experiment 2 below and further validate on several external datasets in Sec. 4.2.

4.1 Experiment 2: Perceived contrast in AR

In Experiment 2, we will test whether the models of supra-threshold contrast matching, introduced in Sec. 4, can explain contrast perception in AR. We used the same apparatus and similar stimuli as in Experiment 1. Below, we explain how the stimuli and the procedure differed between the experiments.

Stimuli and procedure. The stimuli were designed to simulate viewing contrast on an AR display under different ambient luminance levels. Similar to Experiment 1, the reference pattern was presented on two focal planes (see Fig. 3(A)). To simulate the ambient light of a real environment, the background plane of the reference was a uniform field of luminance Y_{BG}. The foreground plane of the reference pattern showed a Gabor patch with a mean luminance Y_{FG}, which simulated a contrast shown on an AR display at that luminance level. The test pattern simulated the same AR display as the reference pattern with a mean luminance level of Y_{FG}, but as seen in a completely dark environment, with the background plane blank. The observers were asked to adjust the contrast of the test pattern so that it matched the perceived contrast of the reference pattern.

Because we wanted to know whether the supra-threshold contrast models from Sec. 4 are also applicable to color vision, we measured both achromatic (Ach) and chromatic contrast patterns. For chromatic patterns, we selected two principal dimensions of the DKL color space [Derrington et al., 1984], red-green (RG) and yellow-violet (YV), the same as the ones modeled in castleCSF [Ashraf et al., 2024]. Two reference contrast levels were selected separately for both achromatic and chromatic patterns. Those levels were selected to ensure they can be matched across the range of background luminance levels. We measured the spatial frequencies of 2 cpd and 4 cpd for achromatic and 2 cpd for chromatic contrast.

Table 2. **Specifications of the stimuli used in Experiment 2.** The display resolution (ppd) and stimuli size were the same as in Experiment 1 (Table 1).

Specification	Value
Color direction	Ach, RG, and YV
Spatial frequency (cpd in Gabor patch)	2, 4 cpd
Reference foreground luminance (Y_{FG})	1, 10, and 100 cd/m^2
Reference background luminance (Y_{BG})	1-587 cd/m^2
Test luminance (Y_{eff})	1, 10, and 100 cd/m^2

Table 3. **Goodness of fit for models tested in Experiment 2.** The values report the reduced chi-square statistic assuming zero degrees of freedom (no parameters). Smaller values indicate a better fit.

Model	Achromatic	Red-green	Yellow-violet
Physical — Eq. (3)	2.45	1.56	1.28
Peli's — Eq. (7)	3.01	1.15	1.63
Kulikowski's — Eq. (6)	**1**	**0.351**	**0.81**

The background luminance levels were selected to answer two questions: (a) does the matching contrast follow the physical contrast or the perceived contrast; and (b), in the latter case, which of the two models, Kulikowski's or Peli's, better explains the perceived contrast. Because the difference between the physical contrast and supra-threshold contrast models is the greatest for the highest background luminance (see Fig. 6), the measurements were performed for the highest possible background luminance that still resulted in detectable reference contrast. To differentiate between the models, we selected the luminance at which Peli's and Kulikowski's models resulted in the most different predictions. The stimuli conditions are summarized in Table 2 and listed in supplementary Table S3.

Participants. We recruited the same group of observers as for Experiment 1. No training session preceded Experiment 2. The experiment was split into two 1 h sessions as the number of stimuli was larger than in Experiment 1. The conditions were split into sessions according to their luminance. The order of the sessions was counterbalanced across the participants.

Experiment results. The results of Experiment 2 are shown in Fig. 7 together with the lines of physical contrast (Eq. (3)), and the predictions of Kulikowski's (Eq. (6)) and Peli's (Eq. (7)) models. The plots follow the format used in Fig. 6 but show the data separately for each spatial frequency, base contrast, and color direction. The two models diverge the most at lower luminance levels, at which the detection thresholds are higher and, therefore, have a stronger influence on the model predictions. The goodness-of-fit for the three considered models is reported in Table 3 as the chi-squared reduced

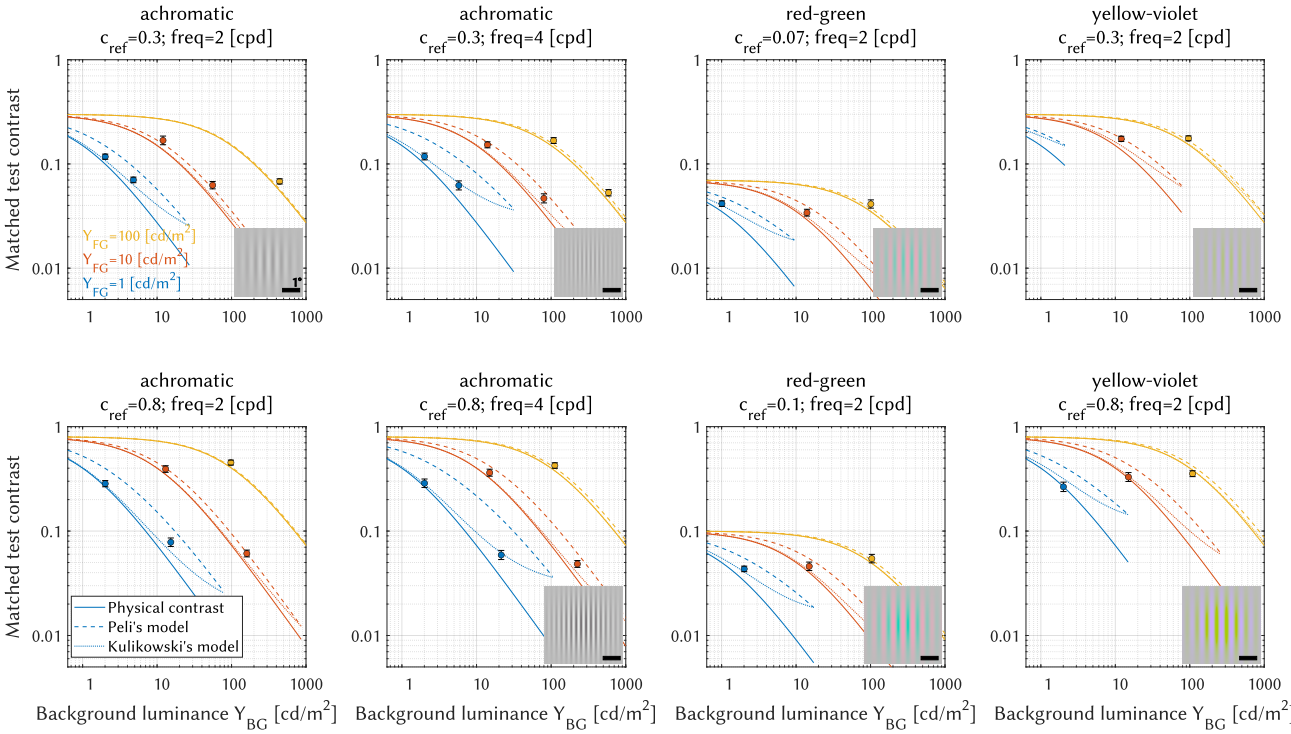

Fig. 7. **The contrast matching results of Experiment 2.** The plots follow the convention of Fig. 6 and show the results for achromatic and chromatic patterns of different spatial frequencies (columns) and reference contrast (rows). Colors denote different luminance levels of an AR display ((*blue*) 1 cd/m^2, (*orange*) 10 cd/m^2, and (*yellow*) 100 cd/m^2) and line patterns different models (see the legend). Each plot shows the cropped stimulus that corresponds to the conditions, and the black bar denotes the visual angle of 1°. The error bars denote 95% confidence interval.

statistics. Both the plots and reported statistics clearly indicate that Kulikowski's model of supra-threshold contrast perception provides the best explanation for our experimental data. Kulikowski's model is particularly important for modeling supra-threshold contrast for darker AR displays observed at higher background light levels. For these conditions, the added background light enhances contrast sensitivity the most and improves the perception of small contrast.

4.2 Validation on other contrast matching datasets

We further validate supra-threshold contrast perception models on several datasets collected from the literature and listed in Table S6. We performed this validation to confirm our findings apply across various measurements, not just Experiment 2 data. In addition to the models explained in Sec. 4, we experimented with additional models from the literature: Foley and Legge's model [1981], Georgeson's model [1991] and Ashraf and Mantiuk's model [2024]. The models are explained in more detail in the supplementary section S7.

To operate on both achromatic and chromatic contrast, we converted the contrast measured in each dataset to cone contrast units. Cone contrast is defined as the magnitude of the three-dimensional differential contrast vectors in LMS (long, medium, and short cone responses) color space:

$$c = \frac{1}{\sqrt{3}}\sqrt{\left(\frac{\Delta L}{L_0}\right)^2 + \left(\frac{\Delta M}{M_0}\right)^2 + \left(\frac{\Delta S}{S_0}\right)^2}, \tag{9}$$

where, L_0, M_0, and S_0 are the color coordinates of the background and ΔL, ΔM and ΔS are the differential color coordinates of the stimuli with respect to the background. We used Stockman and Sharpe [2000]'s cone fundamentals to compute the LMS responses.

To avoid overfitting, we optimized a common set of parameters for all datasets. Parameters values were estimated using a non-linear optimization method (`fminunc` in MATLAB). In addition to the models' intrinsic parameters, we introduced a per-dataset sensitivity correction factor, which accounted for the variability in stimuli and methodologies between the datasets. The optimization minimized the loss defined as the sum of squared differences of log contrasts between the predicted ($\tilde{c}$) and measured test contrasts (c):

$$\underset{\theta, s_d}{\operatorname{argmin}} \sum_i \sum_d \left(\log_{10} c_{d,i} - \log_{10} \tilde{c}_\theta(s_d t^{(i)}_{\text{test}}, s_d t^{(i)}_{\text{ref}})\right)^2, \tag{10}$$

where, θ are the parameters of the model, d is the dataset index, i is the measurement index. $t^{(i)}_{\text{test}}$, and $t^{(i)}_{\text{ref}}$ are, respectively, the threshold contrast predicted by the castleCSF model for the test and the reference condition, and s_d are the per-dataset CSF adjustment factors.

Table 4. **Fitting error comparison across the datasets and models.** The free parameters of each tested model were fitted for all the datasets combined. The error is the RMSE in log cone contrast units expressed in decibels (dB). The error value for the best-fitted model for each dataset is highlighted in bold. (Y: Luminance, and SF: Spatial Frequency)

Dataset	Kulikowski's model [1976]	Peli's model [1990]	Georgeson's model [1991]	Foley and Legge's model [1981]	Ashraf and Mantiuk's model [2024]
No. of fitted parameters	0	0	2	2	5
[Ashraf et al., 2022]	5.27	5.84	5.79	6.99	**5.19**
[Georgeson and Sullivan, 1975] (SF)	5.58	11.65	5.57	12.32	**5.01**
[Georgeson and Sullivan, 1975] (Y)	**1.83**	33.88	2.17	5.07	1.98
[Kulikowski, 1976]	**0.26**	17.96	1.24	15.20	3.59
[Hess, 1990]	**0.37**	11.61	0.79	13.12	3.48
AR matching (ours)	1.10	2.28	**1.01**	4.57	1.07
All datasets	**4.20**	8.11	4.66	8.60	4.64

Because the datasets were fairly sparse and two key models had no tunable parameters, we used all data for both training and testing. The per-dataset CSF adjustment factors and the detailed plots for each model and dataset are included in the supplementary report.

As reported in Table 4, Kulikowski's additive model had the lowest prediction error for most datasets, closely followed by Ashraf and Mantiuk's and Georgeson's models. It should be noted that Kulikowski's and Peli's models have no optimization parameters and solely rely on the threshold contrast to predict the matched contrasts. The models with tunable parameters did not achieve a better overall fit (across all datasets) than Kulikowski's model with no parameters.

5 Generalization to complex images

The models introduced so far could explain well the perception of basic psychophysical stimuli — Gabor patches. Next, we want to show that these models are also applicable to complex images and have practical importance across many applications.

5.1 Global contrast matching model

First, we will consider a model that can match perceived contrast shown on two displays, each operating at a different luminance level and affected by a different amount of ambient light. In Sec. 6, we will show how such a model can be used to control auto-brightness in a perceptually meaningful way. We want our model to be independent of image content so that it could be precomputed and used at negligible computational cost.

According to Kulikowski's model (Eq. (6)), physical contrast c of frequency f shown at a luminance level Y and affected by ambient (background) luminance Y_{BG} is represented as "equivalent" contrast:

$$C'(Y, Y_{\mathrm{BG}}; c, f) = \max\left(c\frac{Y}{Y + Y_{\mathrm{BG}}} - t(Y + Y_{\mathrm{BG}}, f), 0\right). \tag{11}$$

The max operator ensures contrast below the detection threshold becomes 0 (is invisible). The original contrast c is reduced due to ambient light (Eq. (8)). An image shown on two displays will match in perceived contrast if their equivalent contrasts from Eq. (11) closely match, or the loss is minimized:

$$\begin{aligned}&\mathcal{L}(Y^{\mathrm{test}}_{\mathrm{dmax}}; Y^{\mathrm{ref}}_{\mathrm{dmax}}, Y^{\mathrm{test}}_{\mathrm{BG}}, Y^{\mathrm{ref}}_{\mathrm{BG}}) = \\ &\mathbb{E}_{c\sim C}\,\mathbb{E}_{f\sim\mathcal{F}}\sum_{n=0}^{N-1}\Bigg|\log\left(C'\left(Y^{\mathrm{test}}_{\mathrm{dmax}}\,2^{-n}, Y^{\mathrm{test}}_{\mathrm{BG}}; c, f\right) + \epsilon\right) - \\ &\qquad\qquad \log\left(C'\left(Y^{\mathrm{ref}}_{\mathrm{dmax}}\,2^{-n}, Y^{\mathrm{ref}}_{\mathrm{BG}}; c, f\right) + \epsilon\right)\Bigg|,\end{aligned} \tag{12}$$

where $Y^{\mathrm{test}}_{\mathrm{dmax}}$, $Y^{\mathrm{ref}}_{\mathrm{dmax}}$ are the peak luminances of the test and reference displays, and $Y^{\mathrm{test}}_{\mathrm{BG}}$, $Y^{\mathrm{ref}}_{\mathrm{BG}}$ are the corresponding ambient luminance levels. The summation over n ($N = 11$) is introduced to sample representative display luminance levels. Because of the contrast perception is non-linear [Cannon, 1985], we compare the logarithms of equivalent contrast. $\epsilon = 1e-12$ is a small constant introduced to avoid taking the logarithm of zero. We found that L1 distance best explains our experimental data, provided in Table S7 in the supplementary, which will be introduced in the subsequent subsection.

The expected values in Eq. (12) should be computed for a representative set of contrasts c, and frequencies f. They could be computed as a weighted mean, or by sampling from a non-uniform distribution found by inverse transform sampling. We use the latter as a more efficient option. We assume that low contrast dominates in natural images and that its distribution is well approximated by a probability density function $p(c) \propto 1/c\,(= (\log c)')$, corresponding to a log-uniform distribution of C in Eq. (12), over the range [0.01, 0.9]. The spatial energy spectrum is assumed to follow a power law characteristic of natural images, given by $w(f) \sim 1/f^{1+\alpha}$, ($\alpha = .8$) [Ruderman and Bialek, 1993]. The inverse transform sampling for distribution of $\mathcal{F}$ gives us samples $f = \left[u\left(f_{\max}^{-\alpha} - f_{\min}^{-\alpha}\right) + f_{\min}^{-\alpha}\right]^{-1/\alpha}$ where, u is drawn from a uniform distribution, $u \sim \mathcal{U}(0, 1)$, and $f_{\min} = 1$, $f_{\max} = 30$ cpd (acuity limit of 20/20 vision [Snellen, 1897]) is the range of relevant (visible) frequencies.

5.2 Experiment 3: Supra-threshold contrast matching of complex images in AR

To validate the proposed contrast model for complex images, we run an experiment in which participants match contrast across images shown at different luminance levels and affected by a different amount of ambient light.

Stimuli and Procedure. This study is similar to Experiment 2, but with two main modifications: we displayed complex images instead of Gabors, and observers modified image luminance (exposure) instead of contrast. We measured matching luminance to test the model on a different image modification than that used in Experiment 2, and because it directly links to applications (e.g., auto-brightness). Four images with approximately uniform histograms of pixel values were selected from the DIV2K dataset [Agustsson and Timofte, 2017] (see Fig. S7 in the supplementary). The reference stimulus was shown at either of the two peak luminance levels—200 and 800 cd/m^2—in the foreground, while the flat white background luminance was maintained at 100 cd/m^2. Simultaneously, the test stimulus was presented in the foreground with the initial luminance level randomized in a range of 5–15 cd/m^2 with the background having a flat white with luminance levels of 0.1, 1, and 10 cd/m^2. We followed the configuration illustrated in Fig. 3. Each stimulus was resized to fit the visual field of 5 degrees (H) × 2.8 degrees (V).

Test and reference images could not be presented simultaneously, as glare caused by the much brighter reference image affected the contrast of the test image. We also found that the matching procedure was too difficult if test and reference images were presented sequentially. Instead, we used a haploscopic matching protocol, in which test and reference images were presented to different eyes but without an overlap [Ashraf et al., 2022]. The fused image from both eyes showed two images presented next to each other. The haploscopic presentation allowed each eye to adapt to different luminance. The presentation of test and reference was counterbalanced across the left and right eye to eliminate potential bias due to eye dominance [Porac and Coren, 1976]. The total number of trials was 96: 4 (images) × 2 (peak luminance levels of reference foreground) × 3 (luminance levels of test background) × 2 (left/right eye shown test/reference) × 2 (repetitions).

The participants were given instructions: *use the knob on the controller to adjust the test image at the bottom so that the contrast and the visibility of the details in that image matches the contrast and visibility of details of the reference image at the top. Do not match the brightness of the two images — only the contrast and the visibility of the details should be matched.* Participants could revisit the previous trial if they accidentally proceeded to the next trial.

Participants. We recruited 14 observers, 10 of whom participated in both Experiment 1 and 2. A 5-minute training preceded an 1-hour session to familiarize them with the haploscopic setup and the task, and we did not screen participants for Experiment 3.

5.3 Results and model validation

We predict the results of Experiment 3 with the contrast matching model from Eq. (12) by optimizing the match:

$$Y^{*\,\text{test}}_{\text{dmax}} = \underset{Y^{\text{test}}_{\text{dmax}}}{\operatorname{argmin}}\, \mathcal{L}(Y^{\text{test}}_{\text{dmax}}; Y^{\text{ref}}_{\text{dmax}}, Y^{\text{test}}_{\text{BG}}, Y^{\text{ref}}_{\text{BG}}). \tag{13}$$

The predictions, together with the measurements, are shown in Fig. 8. Dashed lines show the expected match if observers were matching physical contrast. The results show that the model can explain the data, though with a small error at $Y^{\text{test}}_{\text{BG}} = 0.1$ and 10 cd/m^2. The root mean-squared error (RMSE) between the predicted and measured

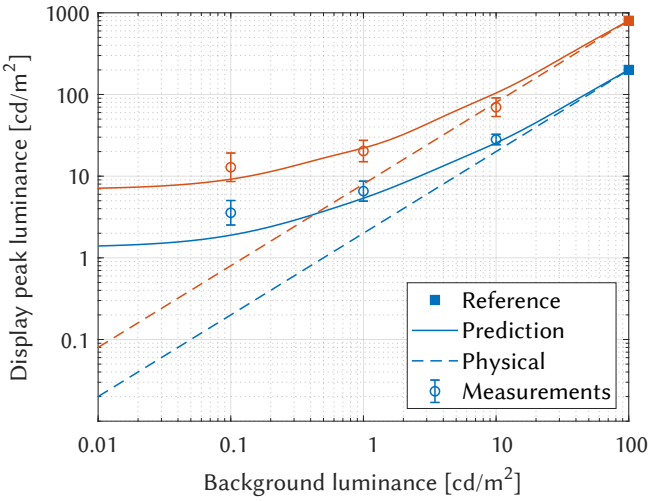

Fig. 8. **Matching contrast of complex images in AR.** The continuous lines represent the lines of matching contrast for a display that is affected by ambient light (x-axis) and has its peak luminance adjusted (y-axis). The circles with error bars (95% confidence intervals) show the results of Experiment 3. The two colors correspond to the two reference conditions (square) we used in Experiment 3 ($Y^{\text{ref}}_{\text{dmax}} = 200$ and $Y^{\text{ref}}_{\text{dmax}} = 800$ cd/m^2). The dashed lines represent matching physical contrast. Per-image results can be found in the supplementary Fig. S8.

peak luminance values, computed in base-10 logarithmic units, was measured as 0.142. As, only one eye could see the test or reference stimulus in Experiment 3, we corrected the sensitivity for monocular viewing (detection thresholds were multiplied by $\sqrt{2}$ [Legge, 1984]). This reduced the validation RMSE from 0.159 to 0.142.

6 Applications

Auto-brightness. Auto-brightness, found in most handheld devices, often disappoints—the image ends up too bright or too dark when moving between bright and dark environments. This is because the methods used are mostly empirical and do not consider contrast perception.

Auto-brightness (AB) methods are meant to (a) preserve image quality across viewing conditions, and (b) strike a balance between image quality and power consumption. The lines of matching contrast in Fig. 8 show the peak luminance that a display should be set to in order to preserve perceived contrast across the ambient light levels (requirement (a)). Which line of matching contrast to use is a user- or system-controlled parameter, which trades image quality for power (requirement (b)).

In Fig. 9, we show AB curves found by solving Eq. (13), referred to as Perceptual AB, for a reference display of varying contrast ratio (from 64:1 to 4:1), seen in a dark room. Therefore, we set $Y^{\text{ref}}_{\text{BG}} = 0$ and $Y^{\text{ref}}_{\text{dmax}} = r\, Y^{\text{ref}}_{\text{black}}$, where r is the contrast ratio and $Y^{\text{ref}}_{\text{black}}$ is the black level, which we set to 0.5 cd/m^2 in our example. The display that maintains its peak on one of the lines delivers a similar quality across ambient light levels. The choice of the line offers a trade-off between quality and power consumption. For reference, we also plot in Fig. 9

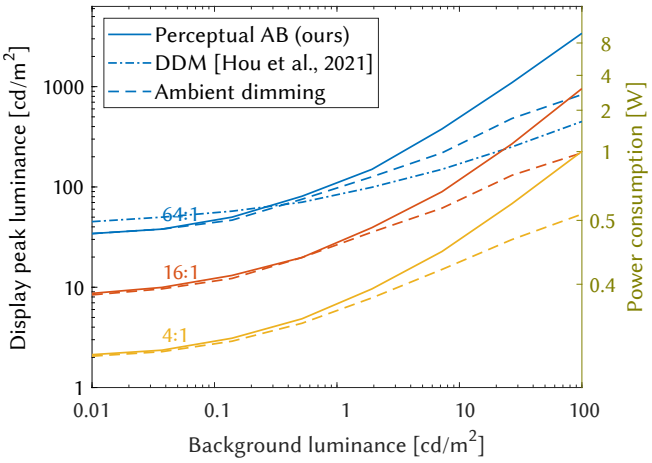

Fig. 9. **Display auto-brightness curves.** Continuous lines (Perceptual AB (ours)) show the lines of matching perceived contrast with respect to a reference display with contrast from 64:1 to 4:1, which is observed with no ambient light. The dotdashed line presents another form of auto-brightness curve investigated with ergonomic study in a driving simulator [Hou et al., 2021]. The dashed lines show the same auto-brightness curves when ambient light dimming is used and the dotted lines follow the physical contrast. The right-hand y-axis shows the expected power consumption of Quest Pro, based on the data from [Chen et al., 2024].

the auto-brightness curve based on the display dimming model (DDM) [Hou et al., 2021], which is derived from an ergonomic study assuming a screen reflectance of 0.05. The DDM curve is shallower and, as shown by our data in Fig. 8, it does not maintain equivalent contrast visibility across ambient light levels.

Ambient dimming. Some AR headsets [Magic Leap, Inc.] offer the ability to control the transmission of the optical component and dim the real environment. We can reformulate the optimization from Eq. (13) to take advantage of such dimming by multiplying the ambient background light $Y_{\mathrm{BG}}^{\mathrm{ad}} = T\, Y_{\mathrm{BG}}$ by the luminous transmission factor of T. We set the transmission factor to change between 0.8 (assumed maximum) and 0.2 (the minimum) as the ambient luminance increases from 1 to 100 cd/m^2 (interpolated on the log-scale). This choice is arbitrary and will vary with the hardware specification. The dashed lines in Fig. 9 show that ambient dimming can save energy while delivering similar contrast.

7 Discussion and conclusions

The optical properties of AR result in reduced contrast in bright environments. Previous works [Chapiro et al., 2024, Murdoch, 2020, Zhang et al., 2021] indicated that the visual system may be able to isolate the AR image and partially discount the background. In Experiment 1, we demonstrated that such discounting does not apply to contrast perception. Instead, we propose that higher visibility of details on AR displays can be explained by supra-threshold models of contrast perception. Our own result of Experiment 2 on a simulated AR display and four other independent datasets demonstrated that Kulikowki's model of contrast matching can explain supra-threshold contrast perception well. We further extended the model to predict contrast matches in complex images and validated those in Experiment 3. We have shown how such a model can be used to better control auto-brightness.

Limitations. We did not investigate higher-level phenomena, such as color appearance, or the effect of ambient light on depth perception. We also do not consider the effect of temporal light- and dark-adaptation [Ferwerda et al., 1996] or local adaptation [Vangorp et al., 2015] on contrast perception. Our model for complex images was intentionally designed to be content-independent, though higher accuracy can be expected if the content is considered. This, however, would require a much more extensive data collection.

Importance. Our findings have important implications for the design of AR display algorithms. However, these models are also applicable to other problems, such as tone mapping for displays strongly affected by additive light [Pytlarz, 2023] (e.g., projection), maintaining appearance when changing brightness [Wanat and Mantiuk, 2014], or the visibility of compression and other artifacts across displays of varying brightness [Ye et al., 2019]. Our findings can help design methods that better align with the functional properties of the human visual system.

Acknowledgments

We would like to thank Özgür Yöntem, Fangcheng Zhong, Param Hanji, Akshay Jindal, Thomas Bytheway, Marek Wernikowski, and Jinze Sha for their help in building the HDR haploscope used in this work. We thank Romain Bachy and Lili Zhang for helpful discussions regarding AR content visibility and scission models. Also, we are grateful to Joseph March for his help with display calibration. We would like to thank the anonymous reviewers for their thoughtful feedback and constructive suggestions.

References

E. Agustsson and R. Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017.

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

M. Ashraf and R. Mantiuk. Modelling contrast matching across luminance levels. In *32nd Color and Imaging Conference-Color Science and Engineering Systems, Technologies, and Applications, CIC 2024*. Society for Imaging Science and Technology, 2024. doi: 10.2352/CIC.2024.32.1.13.

M. Ashraf, R. K. Mantiuk, J. Martinovic, and S. Wuerger. Suprathreshold contrast matching between different luminance levels. *Color Imaging Conference*, 30(1): 219–224, 2022. ISSN 21692629. doi: 10.2352/CIC.2022.30.1.38.

M. Ashraf, R. K. Mantiuk, A. Chapiro, and S. Wuerger. castleCSF — a contrast sensitivity function of color, area, spatio-temporal frequency, luminance and eccentricity. *Journal of Vision*, 24, 2024. doi: 10.1167/jov.24.4.5.

V. Bonin, V. Mante, and M. Carandini. The suppressive field of neurons in lateral geniculate nucleus. *Journal of Neuroscience*, 25(47):10844–10856, 2005.

O. Cakmakci, Y. Ha, and J. P. Rolland. A compact optical see-through head-worn display with occlusion support. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 16–25. IEEE, 2004.

M. W. Cannon. Perceived contrast in the fovea and periphery. *Journal of the Optical Society of America. A, Optics and image science*, 2(10):1760–8, Oct. 1985. ISSN 0740-3232.

A. Chapiro, D. Kim, Y. Asano, and R. K. Mantiuk. AR-DAVID: Augmented reality display artifact video dataset. *ACM Trans. Graph.*, 43(6), 2024. doi: 10.1145/3687969.

K. Chen, T. Wan, N. Matsuda, A. Ninan, A. Chapiro, and Q. Sun. PEA-PODs: Perceptual evaluation of algorithms for power optimization in XR displays. *ACM Transactions on Graphics*, 43(4):1–17, July 2024. ISSN 0730-0301, 1557-7368. doi: 10.1145/3658126.

S. A. Cholewiak, G. D. Love, and M. S. Banks. Creating correct blur and its effect on accommodation. *Journal of Vision*, 18(9):1–1, 2018.

A. M. Derrington, J. Krauskopf, and P. Lennie. Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol.*, 357(1):241–265, Dec. 1984.
J. A. Ferwerda, S. N. Pattanaik, P. Shirley, and D. P. Greenberg. A model of visual adaptation for realistic image synthesis. In *Proc. of SIGGRAPH '96*, page 249–258, New York, New York, USA, 1996. ISBN 0-89791-746-4. doi: 10.1145/237170.237262. URL http://portal.acm.org/citation.cfm?doid=237170.237262.
J. M. Foley and G. E. Legge. Contrast detection and near-threshold discrimination in human vision. *Vision research*, 21(7):1041–1053, 1981.
M. Georgeson and G. Sullivan. Contrast constancy: deblurring in human vision by spatial frequency channels. *The Journal of physiology*, 252(3):627–656, 1975.
M. A. Georgeson. Contrast overconstancy. *JOSA A*, 8(3):579–586, 1991.
R. Hess. Rod-mediated vision: Role of post-receptoral filters. *Night vision*, pages 3–48, 1990.
D. Hou, W. Xu, S. Jing, and Y. Lin. Display dimming model characterized by three-dimensional ergonomic study. *Optical Engineering*, 60(3):035110–035110, 2021.
K. Kiyokawa, Y. Kurata, and H. Ohno. An optical see-through display for mutual occlusion of real and virtual environments. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000)*, pages 60–67. IEEE, 2000.
J. Kulikowski. Effective contrast constancy and linearity of contrast sensation. *Vision research*, 16(12):1419–1431, 1976.
G. E. Legge. Binocular contrast summation—ii. quadratic summation. *Vision research*, 24(4):385–394, 1984.
Magic Leap, Inc. Magic leap. https://developer-docs.magicleap.cloud/docs/guides/features/dimmer-feature/. Accessed: 2025-05-21.
N. Matsuda, Y. Zhao, A. Chapiro, C. Smith, and D. Lanman. HDR VR. In *ACM SIGGRAPH 2022 Emerging Technologies*, pages 1–2. 2022.
Microsoft Corporation. Microsoft HoloLens documentation. https://learn.microsoft.com/en-us/hololens/. Accessed: 2024-12-23.
M. J. Murdoch. Brightness matching in optical see-through augmented reality. *Journal of the Optical Society of America A*, 37(12):1927, Dec. 2020. ISSN 1084-7529, 1520-8532. doi: 10.1364/JOSAA.398931.
E. Peli. Contrast in complex images. *JOSA A*, 7(10):2032–2040, 1990.
C. Porac and S. Coren. The dominant eye. *Psychological bulletin*, 83(5):880, 1976.
J. Pytlarz. Ambient light compensation through adaptive visual modeling. In *SMPTE 2023 Media Technology Summit*. SMPTE, 2023.
D. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Advances in neural information processing systems*, 6, 1993.
L. Sawides, P. De Gracia, C. Dorronsoro, M. A. Webster, and S. Marcos. Vision is adapted to the natural level of blur present in the retinal image. *PLoS ONE*, 6(11):e27031, Nov. 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0027031.
T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of vision*, 11(8):11–11, 2011.
M. Singh and B. L. Anderson. Toward a perceptual theory of transparency. *Psychological review*, 109(3):492, 2002.
H. Snellen. On the methods of determining the acuity of vision. *Systems of diseases of the eye*, 2:93–98, 1897.
A. Stockman and L. T. Sharpe. The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research*, 40(13):1711–1737, 2000. ISSN 00426989. doi: 10.1016/S0042-6989(00)00021-3.
P. Vangorp, K. Myszkowski, E. W. Graf, and R. K. Mantiuk. A model of local adaptation. *ACM Transactions on Graphics*, 34(6):1–13, Nov. 2015. ISSN 0730-0301, 1557-7368. doi: 10.1145/2816795.2818086.
R. Wanat and R. K. Mantiuk. Simulating and compensating changes in appearance between day and night vision. *ACM Transactions on Graphics*, 33(4):147, 2014. doi: 10.1145/2601097.2601150.
A. B. Watson and J. I. Yellott. A unified formula for light-adapted pupil size. *Journal of vision*, 12(10):12–12, 2012.
N. Ye, K. Wolski, and R. K. Mantiuk. Predicting visible image differences under varying display brightness and viewing distance. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, page 5429–5437, Long Beach, CA, USA, June 2019. IEEE. ISBN 978-1-72813-293-8. doi: 10.1109/CVPR.2019.00558. URL https://ieeexplore.ieee.org/document/8953430/.
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. IEEE, 2021.
L. Zhang, M. J. Murdoch, and R. Bachy. Color appearance shift in augmented reality metameric matching. *JOSA A*, 38(5):701–710, 2021.
F. Zhong, A. Jindal, A. O. Yöntem, P. Hanji, S. J. Watt, and R. K. Mantiuk. Reproducing reality with a high-dynamic-range multi-focal stereo display. *ACM Transactions on Graphics*, 40(6):1–14, Dec. 2021. ISSN 0730-0301. doi: 10.1145/3478513.3480513.

Received 23 Jan 2025; Accepted 17 September 2025