Supra-threshold Contrast Perception in Augmented Reality - Supplementary Material

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

ACM Reference Format:
Dongyeon Kim, Maliha Ashraf, Alexandre Chapiro, and Rafał K. Mantiuk. 2025. Supra-threshold Contrast Perception in Augmented Reality - Supplementary Material. In *SIGGRAPH Asia 2025 Conference Papers (SA Conference Papers '25), December 15–18, 2025, Hong Kong, Hong Kong.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3757377.3763824

This document provides supplementary material for the paper 'Supra-threshold Contrast Perception in Augmented Reality'. The content includes additional details of experiments methodologies, setup, data analysis, and models referenced in the main paper.

Table of Contents

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

S1 Related Works

S1.1 Optical see-through AR displays

AR displays primarily consist of a display unit and combiner optics, which serve to display the virtual image and overlay it on the real-world background, respectively. Optical see-through designs have the advantage of allowing users to see the natural world simultaneously with the virtual content, which brings advantages in terms of natural real-world interactions and immersion. However, the light from the real-world background interferes with the virtual image being displayed.

Several approaches have been proposed in commercial AR devices to preserve contrast. Dimming ambient light using an active per-pixel dimming layer [Cakmakci et al. 2004; Kiyokawa et al. 2000; Magic Leap, Inc.] provides good control, but also increases the complexity of the display design. As a consequence, some commercial solutions [Microsoft Corporation] choose to adopt static dimming filters similar to sunglasses. While simple, this approach can degrade the user's experience - static dimming may become excessive in dark environments, but insufficient in overly bright ones. Finally, an AR display that has good visibility without using any dimming would require a strong display engine paired with high-throughput imaging optics, which significantly increases the required power draw, which is challenging for a wearable device [Chen et al. 2024].

S1.2 Contrast-based rendering and assessments

Unlike machine vision cameras, the human visual system responds differently based on the change of the visual signals over the background and the spatial frequency of such stimuli. Modern displays leverage this perceptual behavior to develop advanced rendering techniques such as foveated rendering [Guenter et al. 2012; Tursun et al. 2019] and tone mapping [Reinhard 2020; Tumblin and Rushmeier 1993], which spatially and photometrically redistribute rendering resources to preserve visibility while operating within limited computational resources.

As a counterpart, visual difference predictors [Daly 1992; Mantiuk et al. 2024], built upon contrast sensitivity functions, have been widely used to evaluate the performance of the rendering techniques. However, since these metrics are primarily tailored for conventional direct-view displays, there was a recent observation that they may not fully capture or predict content visibility shown in AR environments [Chapiro et al. 2024], highlighting the importance of low-level investigations revisiting the contrast perception in such viewing environments.

S2 Experiment stimuli

Below, we provide a detailed description of the stimuli used in Experiments 1 and 2. In both our experiments, the reference stimulus was composed of foreground and background images, split over two focal planes. Sec. S2.1- S2.2 explain how the reference foreground and background images were generated, respectively. The test stimuli in both experiments consisted of a fused image displayed on a single plane. The optical models used to create these stimuli are explained in Sec. S2.3. The full matrix of stimuli used in Experiment 1 and 2 are presented in Tables S2- S3.

S2.1 Foreground image

For each contrast value, defined as the modulated signal over the average signal, a sinusoidal grating representation can be generated, modulated in the spatial domain ($Y(\vec{u}) \in \mathbb{R}^{H\times V}$):

$$Y_{\text{FG,sine}}(\vec{u}) = Y_{\text{FG}} + \Delta Y \cos(2\pi\rho u), \tag{S1}$$

where, $\vec{u}$ denotes the two-dimensional spatial coordinate (u, v) of the visual angles in degrees, ρ denotes the spatial frequency given in a unit of cpd, and the ΔY is the modulation of the grating. Here, we formulate the Weber contrast $c = \Delta Y / Y_{\text{FG}}$.

The area of the grating that covers the visual field influences the detection of the contrast [Rovamo et al. 1993]. Thus, we limit the size of the sinusoidal grating and simultaneously smooth the edges by applying a spatial aperture. A conventional way is to modulate the contrast by a radial Gaussian envelope so that the resulting stimulus forms a Gabor patch:

$$Y_{\text{FG,Gabor}}(\vec{u}) = Y_{\text{FG}} + \Delta Y \cos(2\pi\rho u) \exp\left(-\frac{u^2+v^2}{2\sigma^2}\right) \tag{S2}$$

where, σ adjusts the size of the Gaussian aperture in a unit of visual degrees.

S2.2 Background image

The background patterns were generated as follows.

- flat: The ambient light formed by a flat background without texture.

$$Y_{\text{flat}}(\vec{u}) = Y_0. \tag{S3}$$

- bp-noise: To add controllable texture, a bandpass-filtered image is computed as:

$$Y_{\text{bp-noise}}(\vec{u}) = \mathcal{R}\left(\mathcal{F}^{-1}\left(\mathcal{F}\left(Y_{\text{white-noise}}\right)\cdot\mathcal{H}\right)\right). \tag{S4}$$

  where, $Y_{\text{white-noise}}$ is the generated white noise. $\mathcal{F}$ and $\mathcal{F}^{-1}$ are two-dimensional Fourier transform and its inverse, respectively. $\mathcal{R}(\cdot)$ is the real operator with the complex-valued input, and $\mathcal{H}$ is the transfer function defined in the spatial frequency domain corresponding to the bandpass filter. For the bandpass filter, we employ the Butterworth filter defined in the frequency domain ($\vec{\rho}$) as:

$$\mathcal{H}(\vec{\rho}; \rho_0, \sigma_0) = \frac{1}{1 + \left((\sqrt{\rho_u^2 + \rho_v^2} - \rho_0)/\sigma_0\right)^{2n}} \tag{S5}$$

  where, ρ_0 is the central spatial frequency of the filter and σ_0 is the bandwidth of the bandpass filter, which are both in a unit of cpd. In the experiment, we use the fourth order ($n = 4$) Butterworth filter with a central frequency matched to that of the grating, and the filter bandwidth σ_0 is set as 1 cpd.

- bp-noise-dynamic: To understand the effect of relative motion of the background on contrast perception, we additionally formulate a dynamic background as:

$$Y_{\text{bp-noise-dynamic}}(\vec{u}; t) = Y_{\text{bp-noise}}(\vec{u} - \vec{u}(t)) \tag{S6}$$

  where, $\vec{u}(t)$ denotes the additional temporal shift of the background due to potential movement while fixating on the foreground. To define the motion profile, we assume the occurrence of minor head movements, with a speed of 1 degree per second, can still occur even in a constrained environments. The direction of relative motion was chosen as orthogonal to the orientation of the grating provided to the foreground.

S2.3 Optical fusion models

The optical fusion models used to generate test patterns are listed in Table S1, adapted from the recent work of Chapiro et al. [2024]. Among the optical fusion models of foreground and background stimuli, the pinhole model adds the background light without processing. pinhole-diplopic adds the background light while accounting for diplopia, but not for defocus blur. This would be physically accurate if the eyes were assumed to act as pinhole cameras. Lastly, defocus-diplopic simulates both the diplopia and the defocus blur present due to the dioptric distance of two stimuli.

Defocus blur is simulated based on the model described by Cholewiak et al. [2018] given the pupil diameter (p). We convolve the background stimuli with the point spread function ($h(\vec{u})$) applying an

Table S1. **Summary of optical fusion models.**

Fusion model (f)	Equation ($Y_{\text{eff}} = f(Y_{\text{FG}}, Y_{\text{BG}})$)
pinhole (f_P)	$Y_{\text{eff}} = Y_{\text{FG}} + Y_{\text{BG}}$
pinhole-diplopic (f_{PD})	$Y_{\text{eff}} = Y_{\text{FG}} + \frac{1}{2}\sum_{j\in\{\text{L,R}\}} Y_{\text{BG}}(\vec{u} - \vec{u}_j)$
defocus-diplopic (f_{DD})	$Y_{\text{eff}} = Y_{\text{FG}} + \frac{1}{2}\sum_{j\in\{\text{L,R}\}} Y_{\text{BG}}(\vec{u} - \vec{u}_j) * h(\vec{u}; \Delta D, p^*)$

Table S2. **Conditions tested in Experiment 1.** Reference (Ref) and test (Test) patterns used different combinations of foreground (*FG*) and background (*BG*). For the reference, we present two different stimuli for foreground and background and the background textures differ by the test conditions. The effective stimuli (Y_{eff}) are fused with different optical blending models f_i and presented as test stimuli on the foreground.

		FG	*BG*
SP	Ref	$Y_{\text{FG,Gabor}}$	0
	Test	$Y_{\text{eff}} = f_P(Y_{\text{FG,Gabor}}, Y_{\text{BG}})$	0
DF	Ref	$Y_{\text{FG,Gabor}}$	Y_{flat}
	Test	$Y_{\text{eff}} = f_{DD}(Y_{\text{FG,Gabor}}, Y_{\text{BG}})$	0
DNS	Ref	$Y_{\text{FG,Gabor}}$	$Y_{\text{bp-noise}}$
	Test	$Y_{\text{eff}} = f_{DD}(Y_{\text{FG,Gabor}}, Y_{\text{BG}})$	0
DND	Ref	$Y_{\text{FG,Gabor}}$	$Y_{\text{bp-noise-dynamic}}$
	Test	$Y_{\text{eff}} = f_{DD}(Y_{\text{FG,Gabor}}, Y_{\text{BG}})$	0
DNP	Ref	$Y_{\text{FG,Gabor}}$	$Y_{\text{bp-noise}}$
	Test	$Y_{\text{eff}} = f_{PD}(Y_{\text{FG,Gabor}}, Y_{\text{BG}})$	0

achromatic blur of a Gaussian kernel as

$$h(\vec{u}; \Delta D, p) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|\vec{u}\|^2}{2\sigma^2}\right), \tag{S7}$$

where, $\|\cdot\|$ denotes the l-2 norm of the given vector and σ is set as $k\frac{180}{\pi}\frac{p\Delta D}{2}$. ΔD denotes the dioptric difference of the foreground and the background. The value of blur scaling parameter k is chosen as 0.55 from the model of Cholewiak et al. [2018]. This achromatic Gaussian blur simplifies the blur generation.

The simulated blur based on the predicted pupil diameter showed significant discrepancies with the actual blur present in AR scenes in terms of appearance. To address this issue, we calibrated the pupil diameter ($p^* = p/s^*$) by applying a user-specific scaling factor (s^*), determined via a pre-experiment defocus blur-matching calibration (explained later in Sec. S4). Additionally, we estimated the interpupillary distance (ipd) of individuals through interactive calibration using grids presented at two different depth planes for each eye and simulate the diplopic background using the formulation as $\vec{u}_{\mathrm{L/R}} = \left(\tan^{-1}\left(\mp\frac{ipd}{2}\Delta D\right), 0\right)$.

S3 Contrast matching training

Prior to the experiment, all participants were asked to complete a short training session of contrast matching to familiarize themselves with the concept of contrast matching, rather than matching other features such as brightness. Here, the stimuli (Y) were controlled for contrast as follows:

$$Y(c) = (Y_{\mathrm{img}} - Y_0)c + Y_0, \tag{S8}$$

where, Y_{img} represents the luminance of the original image, and $\bar{Y}$ is its mean luminance.

Using an image of six different people's faces, the reference, and the test with different contrast levels (c_{ref}, and c_{test}) and luminance levels (either, Y_0= 30 cd/m^2 or 60 cd/m^2) were presented. The contrast of the reference image was initialized randomly to be between 0.2 and 0.5, and that of the test image to be between 0.1 and 0.9. The reference contrast range was set within the supra-threshold range investigated in the main study.

As in the main experiments, observers adjusted a dial to match the contrast of the test stimulus to that of the reference stimulus. Results from 10 trials were recorded, and Fig. S1 shows the matched contrast ratio ($c_{\mathrm{test}}/c_{\mathrm{ref}}$) for 17 users sorted in a descending order. If the average log10 contrast ratio within an observer's responses exceeded certain thresholds (over 1.1 or under 0.9, experimentally determined) presented as dashed lines in Fig. S1. One observer was classified as an outlier, and their results were thus excluded from the main experiments. The median of the matched contrast ratios, averaged across the remaining observers, was 1.00, indicating that the observers did not provide biased contrast matches in this task.

S4 Defocus blur matching

In the paper (Sec. 3.1), we explained the need of using individualized defocus blur parameters for each observer. This parameter is then used to create the optically-fused single-plane stimuli as explained in Sec. S2.3. Here we provide details of the method used to perform this individual calibration, the results, and comparisons with existing models from the literature.

Table S3. **Conditions tested in Experiment 2.** We opted to measure 2 different spatial frequencies for the achromatic stimuli, but only the lower frequency (2 cpd) for the chromatic ones. This is because the sensitivity of the chromatic mechanisms of the visual system peaks towards lower spatial frequencies. The contrast levels c_{base} for each color stimulus were chosen in cone contrast units to be within the display gamut. The foreground of the reference stimuli and the single-plane test stimuli had the same luminance for each trial. The background luminance of the reference (Y_{BG}) were chosen as described in Sec. 4.1 in the paper.

Color direction	*SF* (cpd)	c_{base}	Y_{FG} / Y_{test} (cd/m^2)	Y_{BG} (cd/m^2)
Achromatic	2	0.3	1	2, 4.8
			10	12, 55.7
			100	442.7
		0.8	1	2, 15.3
			10	13, 160.4
			100	98
	4	0.3	1	2, 5.8
			10	14, 80
			100	108, 587
		0.8	1	2, 21.4
			10	15, 225
			100	112
Red-green	2	0.07	1	1
			10	14
			100	101
		0.1	1	2
			10	15
			100	104
Yellow-violet	2	0.3	10	12
			100	96
		0.8	1	2
			10	15
			100	109

Detailed procedure. We collected 10 trials per observer. Observers were instructed to match the contrast of the noisy background on the test to the contrast of the noise pattern of reference while focusing at the fixation target shown on the front plane. The defocus blur size of the test was adjusted with a controller similar to the main experiment. Rather than measuring the size of defocus blur directly, we estimated the parameter rescaling the pupil diameter predicted by the model [Watson and Yellott 2012], allowsing us to employ this measured value for different luminance levels.

Although the reference included a fixation target in the foreground, the accommodation state naturally shifted to the background as observers examined the noise pattern to make their decisions. This natural response in focal shift made the task difficult for the untrained observers. If the median value of the measured

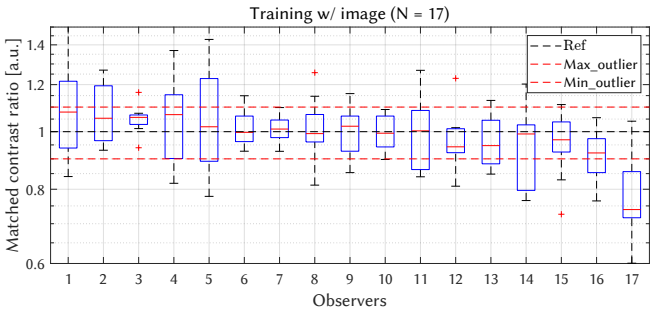

Fig. S1. **Training result of 17 observers recruited for the experiments.** One of the observers (observer 17) was disqualified with the training results and data was excluded. Observers are sorted by the matched contrast ratio values, thus not following the observer ID presented in the paper.

scaling parameter exceeded 2, we repeated the session, allowing for up to three cycles. If the measured scaling parameter still remained over 2, we used the measured value in the main experiment.

Results. The defocus blur-matching results are shown in Fig. S2. It can be observed that the experts who participated in the experiment show consistent results over the trials, but a portion of naïve observers performed the task relatively inconsistently. This result indicates a potential cause of the larger variance in Experiment 1 for the conditions involving defocus blur (DNS and DND) as compared to the conditions that did not involve defocus blur (DF).

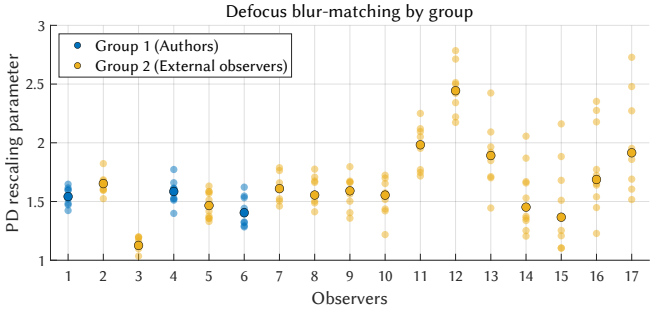

Fig. S2. **Individiual results of defocus blur-matching calibration.** The pupil diameter (PD) rescaling parameters were estimated 10 times for individual users and sorted by the standard deviation (The observer ID does not match the one presented in the paper). The plots display the parameter estimated across 10 trials, with the median value highlighted by a black edge. Observers were categorized into two groups for analysis. (*blue*: Group 1 (authors), *red*: Group 2 (external observers))

Pilot study. In Experiment 1, the defocus blur generated by the predicted pupil diameter model exhibited a significant difference compared to the pupil diameter interactively matched using the defocus blur matching method, as shown in Fig. 6(B). To examine the effect of defocus blur on the perceived contrast of the foreground scene, we conducted a pilot study with a subset of observers participated the Experiment 1. This additional study included test conditions dual-noise-static-Watson (DNS-W) and dual-noise-dynamic-Watson (DND-W), which simulated the defocus blur using the predicted pupil diameter model. Note that the dual-noise-static and dual-noise-dynamic in the main paper were presented as dual-noise-static-Matched (DNS-M) and dual-noise-dynamic-Matched (DND-M) to rule out potential confusion.

Figure S3 shows the results of the pilot study conducted with four observers (3: authors and 1: naïve observer). These results indicate that the simulated defocus blur, generated using the model [Watson and Yellott 2012], blurs the textures more than the actual condition. Note that the rescaling parameters shown in Fig. S2, are above 1. Consequently, the observers adjusted the test contrast to a lower level than the reference contrast as the masking from the background textures diminish in the test stimuli.

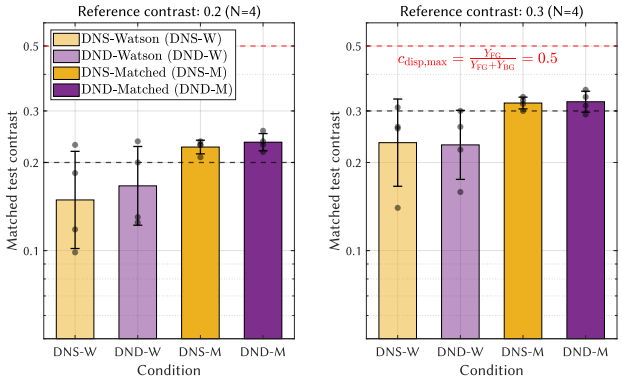

Fig. S3. **Pilot study of Experiment 1.** The test contrasts of four conditions (*light orange*: DNS-W, *light purple*: DND-W, *orange*: DNS-M, and *purple*: DND-M) are matched with four observers (black dots) and the 95% confidence interval is provided as errorbar.

Pupil diameter measurements. For DNS and DND conditions, which incorporated defocus blur to provide the test stimuli, we used the pupil diameter obtained through defocus blur matching. This approach was necessary because our testbed lacked sufficient space to place an eye tracker, and we aimed to maintain the size of the defocus blur consistently throughout the experiment. In practice, however, even with the same luminance level of stimuli presented, the eye's pupil diameter fluctuates over time, and measurements are often affected by eye blinking, which let the interactive defocus blur rendering challenging.

In Fig. S4, we present the pupil diameter data expanded with measurements from a portion of observers. An eye tracker from Pupil Labs was used to record the pupil diameter while observers were exposed to the stimuli employed in defocus blur matching calibration. The pupil diameter of the observer's right eye was recorded for a duration of one minute, which is sufficiently long to allow the pupil to adapt to the luminance level. Portions of the data corrupted by eye blinking were excluded to ensure data accuracy. A reference scale bar was placed directly beneath the observer's eye to scale the recorded pupil diameter to physical size, and captured simultaneously. The measured pupil diameter showed a large per-subject variation, and the discrepancy with the pupil diameter inversely

matched with the appearance supports the necessity of the defocus blur matching task preceding Experiment 1.

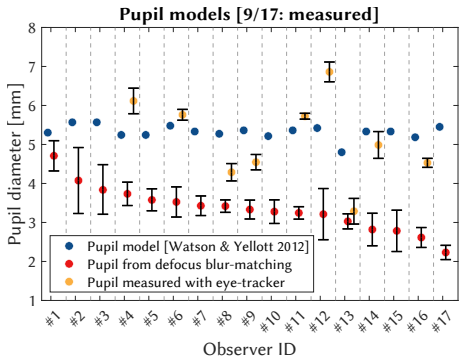

Fig. S4. **Pupil diameter profiles.** We measured the pupil diameter of 9 out of 17 observers who participated in the defocus blur-matching calibration and they are sorted by the estimated size of pupil dimaeter following Fig. 4. The error bars represent the standard deviation of the measurements (*yellow*) and the standard deviation of the estimated values obtained from the calibration (*red*).

S5 Individual bias in contrast matching

In the results of Experiment 1, we found that some naive observers consistently matched towards higher test contrast, even under SP conditions. This bias was often one-sided, as shown in Fig. S5. This outcome was unexpected, as we conducted contrast matching training, and the results of the training, shown in Fig. S1, did not indicate any bias.

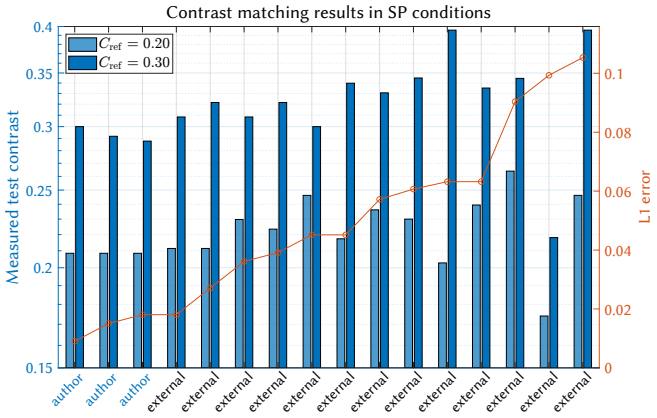

Fig. S5. **Per-individual bias in contrast matching** The plot presents the matched test contrast in SP conditions in two different reference contrast levels (C_{AR}=0.2 (*light blue*), and 0.3 (*blue*)) and the L1 error (*orange*) between the matched test contrast and the reference contrast sorted per individual observers. Observers were anonymized but categorized into two groups (blue: author, and black: external observer).

S6 Data analysis

In both Experiments 1 and 2, each data point was calculated as the median of the log-contrast values across ten trials for each observer in each condition. Using the median minimizes the influence of outliers or extreme values, which can occur due to variability in observer performance or environmental factors during the experiment. The medians from individual observers were then averaged across all participants to compute the final data points for each condition. All further analysis is then performed on either the averaged data across the sample or the individual observers' median data. All analysis was also performed in log contrast units.

In the following sections, we present some additional statistical analyses with the data from Experiment 1 to determine the significance of the background discounting effect in AR.

S6.1 Statistical significance - Experiment 1

In the paper Sec. 3.1, we describe the statistical analyses conducted to test the background discounting hypothesis. The full t-test results are shown in Table S4, including both p-values and Cohen's effect sizes. The results indicate that the contrasts in the dual-flat (DF), dual-noise-static (DNS), and dual-noise-dynamic (DND) conditions were not significantly different from the physical contrast of the stimulus. However, the dual-noise-pinhole (DNP) condition showed a statistically significant difference, highlighting the critical role of simulated blur in achieving contrast matching in AR.

Table S4. **Statistical comparison in Experiment 1** The table reports the p-values (p) and Cohen's d (d).

Comparison	C_{AR}**=0.2**	C_{AR}**=0.3**
DF vs. C_{AR}	$p = 0.904$, $d = 0.031$	$p = 0.215$, $d = 0.324$
DNS vs. C_{AR}	$p = 0.106$, $d = 0.431$	$p = 0.122$, $d = 0.410$
DND vs. C_{AR}	$p = 0.925$, $d = 0.024$	$p = 0.665$, $d = 0.110$
DNP vs. C_{AR}	$p < 0.001$, $d = 1.567$	$p < 0.001$, $d = 1.331$

S6.2 Power analysis - Experiment 1

We conducted a power analysis to evaluate the sufficiency of our sample size to detect effects with 80% power at a significance level of $\alpha = 0.05$, using one-tailed tests. For large effect sizes ($d > 0.8$, e.g., $d = 1.567$ in condition DNP for $c_{AR} = 0.2$), the current sample size of $n = 16$ was more than adequate, requiring as few as 6 participants. For moderate effect sizes ($d \approx 0.5$, e.g., $d = 0.431$ in condition DNS for $c_{AR} = 0.2$), a sample size of 40∼ 50 observers would be necessary to achieve sufficient power. For small effect sizes ($d < 0.2$), the required sample would exceed 500 observers, making detection of such effects impractical within the constraints of the current study. For our purpose, the study design was appropriately powered to detect large effects.

S7 Supra-threshold contrast models

Section 4 in our paper presented the two notable supra-threshold models from the literature; the additive (Kulikowski) and the multiplicative (Peli) contrast models. In addition to these, we tested our dataset and those from the literature on a few other models as

Table S5. **Post-hoc sample size estimation for Experiment 1** The table reports the estimated sample size (n) for the observed effect sizes in Table S4.

Comparison	C_{AR}=0.2	C_{AR}=0.3
DF vs. C_{AR}	$n = 8,170$	$n = 77$
DNS vs. C_{AR}	$n = 45$	$n = 49$
DND vs. C_{AR}	$n = 13,629$	$n = 651$
DNP vs. C_{AR}	$n = 6$	$n = 7$

well. The supra-threshold contrast response is often characterized as a non-linear contrast transducer function [Foley and Legge 1981; Wilson 1980]. For a specified test and reference stimulus, where the contrast is perceptually matched, the responses of the transducer functions can be assumed to be equal and presented as:

$$k_{\text{test}}\left(\frac{c_{\text{test}}}{t_{\text{test}}}\right)^{\gamma_{\text{test}}} = k_{\text{ref}}\left(\frac{c_{\text{ref}}}{t_{\text{ref}}}\right)^{\gamma_{\text{ref}}}, \tag{S9}$$

where k and γ are the model's gain control (multiplier and exponent respectively) parameters. The equation can be simplified to predict the test contrast as a function of the known variables:

$$c_{\text{test}} = t_{\text{test}}\left(k\frac{c_{\text{ref}}}{t_{\text{ref}}}\right)^{\gamma}. \tag{S10}$$

t_{ref} and t_{test} are the threshold contrast of the reference and test stimuli respectively, and c_{ref} is the contrast of the reference stimulus to be matched.

Another model of supra-threshold contrast matching was proposed by Georgeson [1991], which is a modification on Kulikowski's model. While Kulikowski's model is a simple additive model of the threshold contrasts, Georgesons's model stipulates that for patterns where the threshold contrast is disproportionately lower due to limiting optical factors, neural noise, etc. the perception at supra-threshold levels is not affected as much. A simple additive model such as Kulikowski's, would underestimate the perceived contrast of such patterns. Georgeson proposed a non-linear model where the perceived contrast is boosted when the difference between the thresholds of the test and reference contrast is too high and call it the "overconstancy model". In this model, the response to a specified test contrast can be presented as:

$$R_{\text{test}} = \left(\frac{c_{\text{test}} - t_{\text{test}}}{c_{\text{norm}} - t_{\text{test}}}\right)^{m}, \tag{S11}$$

where c_{norm} is the contrast at which the response of the mechanism saturates, and m is the exponent of the transducer. To predict the contrast value matched to a given reference contrast value, the contrast overconstancy model can be rearranged as follows:

$$c_{\text{test}} = (c_{\text{norm}} - t_{\text{test}})\left(\frac{c_{\text{ref}} - t_{\text{ref}}}{c_{\text{norm}} - t_{\text{ref}}}\right)^{q} + t_{\text{test}}. \tag{S12}$$

q is the combined exponent parameter for both the test and reference contrasts, and can be optimized along with c_{norm}.

More recently, the model from Ashraf and Mantiuk [2024] evaluated existing models—specifically, the additive model by Kulikowski [1976] and the multiplicative model by Peli [1990]—and identified their limitations in accurately predicting contrast matching over a broad luminance range for the HDR dataset in **?**. To address these shortcomings, they proposed a hybrid model that integrates both additive and multiplicative components. The model accounts for threshold ratios and differences, optimizing its parameters through regression analysis. The model is summarized as follows:

$$c_{\text{test}} = \delta(r^t)(c_{\text{ref}})^{\gamma} + \alpha(\Delta t), \tag{S13a}$$

$$\delta(r^t) = \delta_m r^t + \delta_i, \tag{S13b}$$

$$\alpha(\Delta t) = \alpha_m \Delta t + \alpha_i, \tag{S13c}$$

$$r^t = \frac{t_{\text{ref}}}{t_{\text{test}}}, \tag{S13d}$$

$$\Delta t = t_{\text{ref}} - t_{\text{test}} \tag{S13e}$$

where γ, δ_m, δ_i, α_m, and α_i are the model parameters. γ accounts for the non-linearity between the matched contrasts. The $\delta(\cdot)$ function represents a scaling factor for the reference contrast, dependent on the threshold ratio (r^t) with δ_m, and δ_i as the slopes and the intercepts of this linear relationship. Similarly, the $\alpha(\cdot)$ adjusts the contribution of the threshold difference (Δt) with its linear dependence on α_m, and α_i, which are respectively the slope and intercept of this adjustment.

S7.1 Validation with other datasets

Sec. 4.2 in the paper provides details of the training method used to validate the contrast matching models. This validation is based on the dataset from our study (Experiment 2) as well as additional datasets from the literature. Table S6 provides the dataset used in supra-threshold contrast modeling. From the reported dimensions of the stimuli used in each of the studies, it can be seen that the datasets encompass a wide range of stimuli and experimental conditions. This was done to evaluate the models' generalizability across different supra-threshold contrast perception scenarios.

S8 Additional details

S8.1 Experimental setup

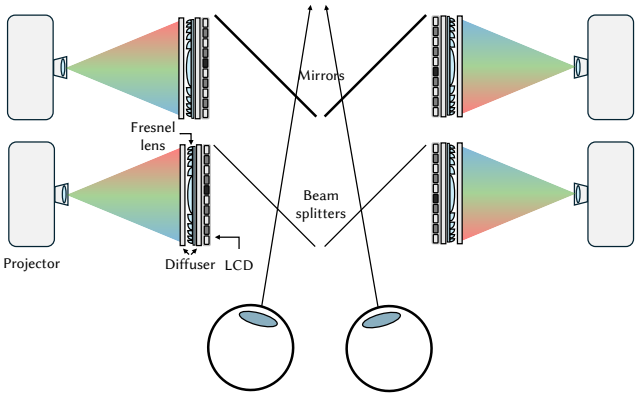

Fig. S6. **Detailed schematic of the experimental setup.**

The dual-layer haploscope, as shown in Fig. S6, is constructed using a stack comprising a projector, diffusers, a Fresnel lens, and an LCD to deliver dual-layer HDR images to individual eyes. The

Table S6. **Summary of datasets used in supra-threshold contrast modeling.** The table includes experimental stimuli, input dimensions, and matching axes (Y: Luminance, SF: Spatial Frequency, and C: Color).

Dataset	Stimulus (Aperture)	Dimensions	Protocol	Matching Axes	No. of data points
[?]	Grating (Gaussian)	Y, SF, C	Matching	Y	505
[Georgeson and Sullivan 1975]	Grating (Rectangular)	Y, SF	Matching	Y, SF	82
[Kulikowski 1976]	Grating (Rectangular)	Y	Matching	Y	46
[Hess 1990]	Grating (Rectangular)	Y, SF	Matching	Y	72
AR matching (ours)	Grating (Gaussian)	Y, SF, C	Matching	Y	32

Fig. S7. **Assets used in Experiment 3.** Mushroom: CC BY-SA 4.0, provided by serlo.org. Coral: CC BY-SA 2.0, sourced from Flickr and was captured by Matt Kieffer. Valley: CC BY-SA 3.0, sourced from Wikimedia Commons, and attributed to Tobi 87. Spider: public domain under the CC0 1.0, Universal license and was obtained from Stockvault.

Fresnel lens (Comar Optics) is positioned at the center of the display stack to efficiently converge the backlight toward the viewing zone, maximizing light usage. It is sandwiched with two diffusers (a product of Luminit) having narrow diffusing angles (5, and 10 degrees, respectively). The alignment of the overall display layers is performed prior to the experiment, and the 3D geometric calibration is performed per individual to compensate the layer image alignment due to the individual difference of the interpupilary distance.

S8.2 Stimuli used in Experiment 3

Fig. S7 presents four images used for the Experiment 3. These are the images that with most uniform histograms in terms of mapped luminance values in the DIV2K dataset [Agustsson and Timofte 2017].

S8.3 Additional results - Experiment 3

Per-image results of Experiment 3. We present additional results of contrast matching with complex images in Fig. S8, processed on a per-image basis. Although there are slight variations in measurements between different images, our image-independent contrast matching model predicts the individual measured values with good accuracy.

Validation with different distance types and contrast scaling operators. Equation 12 uses logarithmic scaling to reflect the non-linear perception of contrast, and applies the L1 distance to compute the difference between the test and reference displays. Table S7 presents the validation RMSE for different combinations of distance types (L1, Lp ($p = 1.5$), and L2) and contrast scaling methods (log, linear, and sqrt). Lp ($p = 1.5$) refers to the Minkowski distance of order 1.5.

Table S7. **Validation RMSE with different distance types and scaling operations.** The lowest RMSE is bolded.

Scaling Type / Distance	L1	Lp ($p = 1.5$)	L2
log	**0.142**	0.208	0.252
linear	0.606	0.622	0.632
sqrt	0.468	0.482	0.491

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.

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.

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.

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.

S. J. Daly. Visible differences predictor: an algorithm for the assessment of image fidelity. In *Human Vision, Visual Processing, and Digital Display III*, volume 1666, pages 2–15. SPIE, 1992.

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.

B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder. Foveated 3D graphics. *ACM Transactions on Graphics (TOG)*, 31(6):1–10, 2012.

R. Hess. Rod-mediated vision: Role of post-receptoral filters. *Night vision*, pages 3–48, 1990.

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.

Magic Leap, Inc. Magic leap. https://developer-docs.magicleap.cloud/docs/guides/features/dimmer-feature/. Accessed: 2025-05-21.

R. K. Mantiuk, P. Hanji, M. Ashraf, Y. Asano, and A. Chapiro. Colorvideovdp: A visual difference predictor for image, video and display distortions. 43(4), 2024. ISSN 0730-0301. doi: 10.1145/3658144. URL https://doi.org/10.1145/3658144.

Microsoft Corporation. Microsoft HoloLens documentation. https://learn.microsoft.com/en-us/hololens/. Accessed: 2024-12-23.

E. Peli. Contrast in complex images. *JOSA A*, 7(10):2032–2040, 1990.

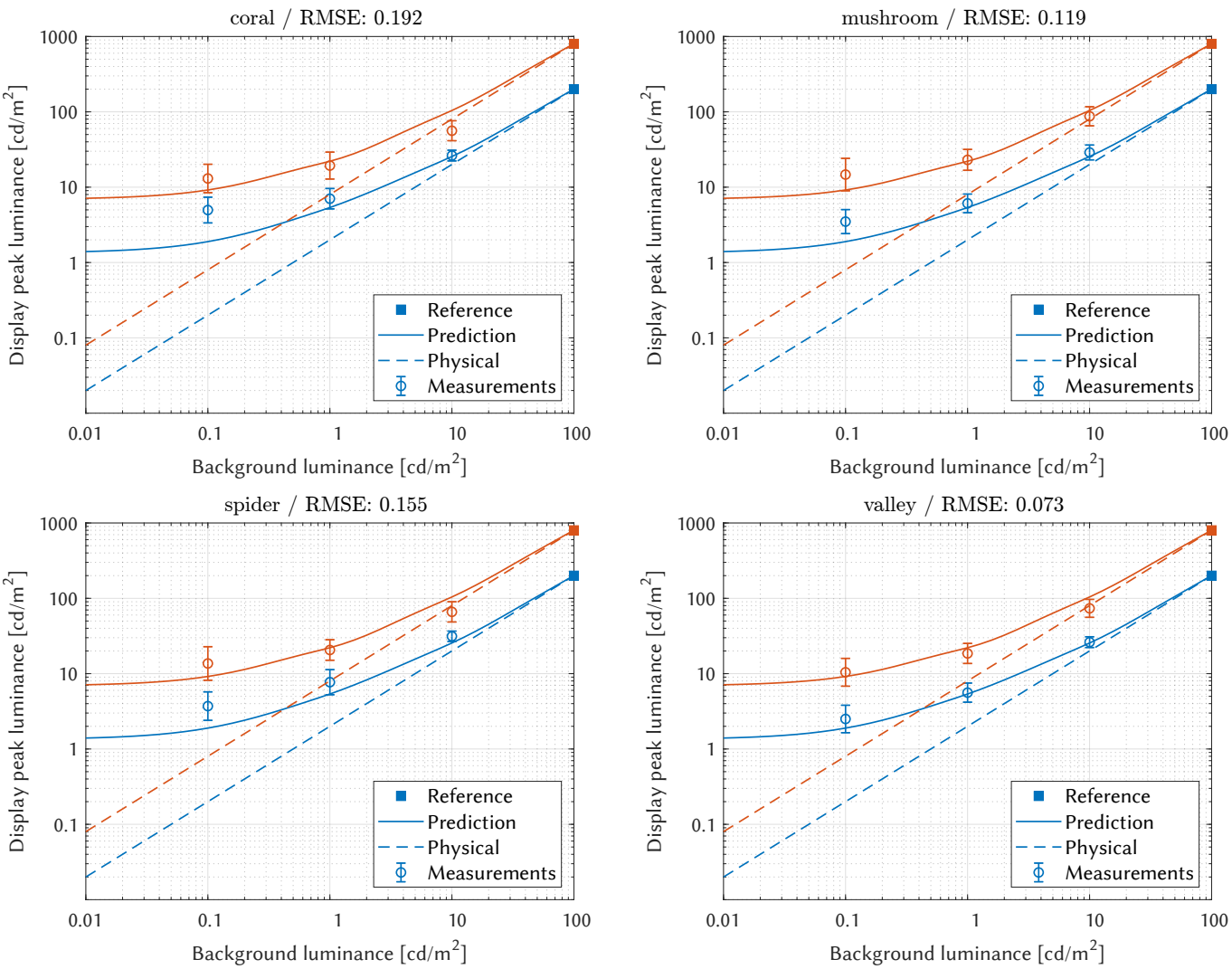

Fig. S8. **Results of Experiment 3 per image.** Each plot is additionally presented with RMSE showing on the top.

E. Reinhard. High dynamic range imaging. In *Computer Vision: A Reference Guide*, pages 1–6. Springer, 2020.

J. Rovamo, O. Luntinen, and R. Näsänen. Modelling the dependence of contrast sensitivity on grating area and spatial frequency. *Vision Research*, 33(18):2773–2788, Dec. 1993. ISSN 00426989. doi: 10.1016/0042-6989(93)90235-O.

J. Tumblin and H. Rushmeier. Tone reproduction for realistic images. *IEEE Computer graphics and Applications*, 13(6):42–48, 1993.

O. T. Tursun, E. Arabadzhiyska-Koleva, M. Wernikowski, R. Mantiuk, H.-P. Seidel, K. Myszkowski, and P. Didyk. Luminance-contrast-aware foveated rendering. *ACM Transactions on Graphics (TOG)*, 38(4):1–14, 2019.

A. B. Watson and J. I. Yellott. A unified formula for light-adapted pupil size. *Journal of vision*, 12(10):12–12, 2012.

H. R. Wilson. A transducer function for threshold and suprathreshold human vision. *Biological cybernetics*, 38(3):171–178, 1980.

Received 23 Jan 2025; Accepted 17 September 2025