Perceptual Color Experience on Conventional and AR Displays across Color Gamut and Ambient Luminance

YUNXIANG ZHANG, New York University, USA
YUTA ASANO, Reality Labs, Meta, USA
THOMAS WAN, Reality Labs, Meta, USA
QI SUN, New York University, USA
ALEXANDRE CHAPIRO, Reality Labs, Meta, USA

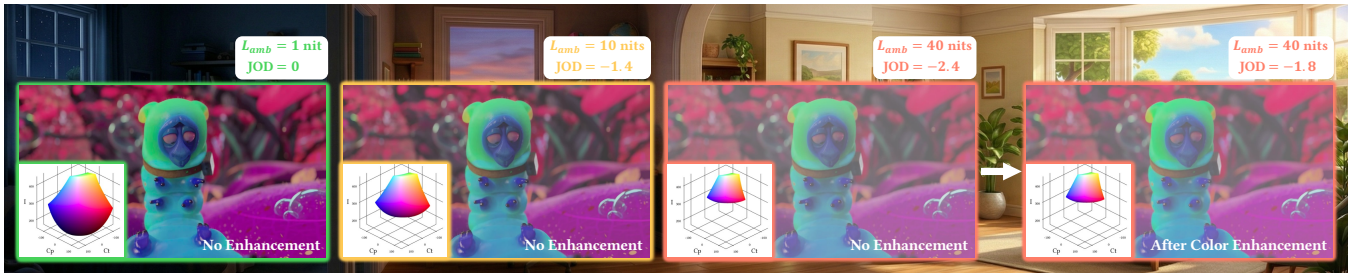

Fig. 1. Ambient light is ubiquitous in real-world viewing environments and can substantially alter display color appearance. Increasing ambient luminance shrinks the effective color gamut and degrades perceived color quality relative to a display's native capability. We systematically evaluate this effect across display characteristics and ambient conditions, and introduce color-enhancement algorithms that exploit unused display-gamut headroom for compensation. This figure shows a video presented on a display with Display P3 gamut and 280-nit peak luminance under three ambient luminance levels, L_{amb} = 1, 10, and 40 nits, together with a color-enhanced result at L_{amb} = 40 nits (rightmost). Insets visualize the effective color gamut in ICtCp color space [Dolby Laboratories 2016] after accounting for ambient light, with just-objectionable-difference (JOD) scores indicating perceived color quality. Because HDR/WCG imagery cannot be faithfully reproduced in PDF format, all images in this manuscript are tone- and gamut-mapped for illustration. Image credit: Netflix.

A display's color gamut defines the range of colors it can reproduce and is a key factor of perceptual color experience. Modern displays increasingly support high dynamic range (HDR) and wide color gamut (WCG), enabling brighter and more vivid color reproduction. In practical viewing environments, however, ambient illumination often shrinks the effective gamut and degrades perceived color quality relative to a display's native capability.

We study how color gamut, dynamic range, and ambient luminance jointly shape perceived color quality across conventional flat-panel and optical see-through augmented reality (OST-AR) displays. To this end, we gather a WCG HDR video dataset and develop a display-simulation pipeline that emulates diverse display capabilities and ambient conditions. Using this framework, we conduct a direct-view user study to collect preference data and derive a perceptual model that predicts perceived color quality from these factors. A second study evaluates model generalization to OST-AR viewing conditions and examines the effect of background pattern. To mitigate ambient-induced color degradation, we introduce three color-enhancement algorithms and validate them in a third study. Our findings provide practical guidance for display-system design aimed at improving real-world color experience.

Authors' Contact Information: Yunxiang Zhang, yunxiang.zhang@nyu.edu, New York University, Brooklyn, USA; Yuta Asano, yutaasano1986@gmail.com, Reality Labs, Meta, Redmond, USA; Thomas Wan, thomaswan@meta.com, Reality Labs, Meta, Sunnyvale, USA; Qi Sun, qisun@nyu.edu, New York University, Brooklyn, USA; Alexandre Chapiro, alex@chapiro.net, Reality Labs, Meta, Sunnyvale, USA.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
SA Conference Papers '26, Kuala Lumpur, Malaysia
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2842-6/2026/12
https://doi.org/10.1145/3829340.3842215

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

Additional Key Words and Phrases: Wide Color Gamut, High Dynamic Range, Color Perception, Perceptual Color Volume, Ambient Light

ACM Reference Format:
Yunxiang Zhang, Yuta Asano, Thomas Wan, Qi Sun, and Alexandre Chapiro. 2026. Perceptual Color Experience on Conventional and AR Displays across Color Gamut and Ambient Luminance. In *SIGGRAPH Asia 2026 Conference Papers (SA Conference Papers '26), December 01–04, 2026, Kuala Lumpur, Malaysia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3829340.3842215

1 Introduction

Wide-color-gamut (WCG) and high-dynamic-range (HDR) displays are increasingly deployed across televisions, monitors, mobile devices, and optical see-through augmented reality (OST-AR) systems. By expanding color-gamut coverage and dynamic range, these displays enable brighter and more vivid color reproduction. In practice, however, displays are commonly viewed outside ideal dark-room conditions. Ambient illumination introduces an additive light component, via panel reflection on conventional flat-panel displays or optical transmission in OST-AR displays, that reduces the effective color gamut and degrades perceived color quality relative to a display's native capability. As display technologies continue to evolve toward wider gamuts and higher brightness, understanding how display specifications and ambient conditions jointly shape perceptual color experience becomes increasingly important.

Prior work has examined visual quality using HDR/WCG images [Choudhury et al. 2021; Heckaman et al. 2007; Sakurai et al. 2008b] and videos [Athar et al. 2019; Mantiuk et al. 2024; Shang et al. 2023]. However, these studies generally consider limited color-gamut and dynamic-range configurations and do not characterize their interaction with ambient light. Recent work has studied various aspects of visual perception in OST-AR under ambient illumination, including brightness perception [Murdoch 2020], contrast perception [Kim et al. 2025], color appearance [Hassani and Murdoch 2019; Zhang et al. 2021a], and video artifact visibility [Chapiro et al. 2024]. Yet, the perceptual benefits of expanded display color capabilities across varying ambient conditions remain inadequately quantified.

In this research, we systematically study color experience across three interacting display and viewing factors: color gamut, dynamic range, and ambient luminance. To support this investigation, we collect a WCG HDR video dataset with broad color coverage and diverse content categories. We develop a display-simulation pipeline that combines tone mapping, gamut mapping, and an ambient-light model to emulate diverse display characteristics and ambient conditions on an HDR/WCG reference display. Using this framework, we conduct a direct-view study to measure perceptual preferences across display and ambient conditions and scale the responses into just-objectionable-differences (JODs) [Perez-Ortiz et al. 2019]. From these data, we establish a perceptual model that predicts perceived color quality from color gamut, dynamic range, and ambient luminance. In a follow-up user study, we evaluate how well the model generalizes from direct-view to OST-AR conditions and test how background structure affects perceived color quality.

To mitigate ambient-induced color-fidelity degradation, we introduce three color-enhancement algorithms that exploit unused display-gamut headroom for ambient compensation, and validate their perceptual benefits in OST-AR through a third user study. Our findings provide quantitative guidance for display-system design and practical methods for enhancing color experience in real-world viewing environments.[1] In summary, we contribute:

- A WCG HDR video dataset with broad color/luminance coverage;
- User studies quantifying how color gamut, dynamic range, and ambient luminance jointly shape perceived color quality across conventional flat-panel and OST-AR displays;
- A perceptual model predicting perceived color quality from display color gamut, dynamic range, and ambient luminance;
- Color-enhancement algorithms that exploit unused gamut headroom to compensate for ambient-induced color degradation.

2 Related Work

Color Representation and Color Gamut. Color appearance models, including CIECAM02 [Moroney et al. 2002], iCAM06 [Kuang et al. 2007], and appearance-based HDR encoding frameworks [Kunkel et al. 2013], model effects of viewing conditions, adaptation, and spatial context on perceived color attributes [Fairchild 2013]; ambient-adaptive systems further adjust rendering based on environmental illumination [Wu et al. 2019]. Display color reproduction is characterized through gamut and color-volume metrics, including discernible-color estimates and effects of flare, peak luminance, and diffuse white [Jiang et al. 2018, 2020; Masaoka et al. 2013; Wen 2006], as well as Gamut Rings for perceptual display comparisons [Masaoka et al. 2018, 2020, 2025; Smith et al. 2020]. Gamut-mapping methods transform colors between source and target gamuts while balancing hue preservation, lightness consistency, and contrast maintenance [Morovič 2008; Morovič and Luo 2001], spanning global compression/expansion [Xu et al. 2018a,b], spatially varying approaches [Bonnier et al. 2007; Farup et al. 2007; Kimmel et al. 2005], perceptually guided techniques [Zamir et al. 2014, 2019], and HDR/WCG content [Preiss et al. 2014; Sikudova et al. 2015]. Psychophysical evaluations reveal content- and method-dependent trade-offs [Dugay et al. 2008; Montag and Fairchild 1997]. These works provide foundations for representing, measuring, and transforming color.

Perceptual Color Quality Assessment. Prior perceptual studies have examined how color gamut and dynamic range affect image quality and preference. Paired-comparison studies showed that increasing display gamut volume generally improves preference, with diminishing returns and strong content dependence [Heckaman et al. 2007; Sakurai et al. 2008a,b]. Later work studied objective and subjective quality metrics for HDR/WCG images and videos [Choudhury and Daly 2020; Choudhury et al. 2021; Rousselot et al. 2019; Shang et al. 2023], and perceptual optimization of HDR rendering pipelines [Cao et al. 2024]. However, many studies focus on still images [Choudhury et al. 2021; Heckaman et al. 2007; Rousselot et al. 2019], while video evaluations often fix either gamut or dynamic range [Athar et al. 2019; Mantiuk et al. 2024]. Systematic evaluation is also limited by the lack of diverse, calibrated WCG HDR datasets: existing datasets are often domain-specific [Froehlich et al. 2014; Keller et al. 2024; Song et al. 2016] and sparsely cover high-chroma colors despite using WCG containers. Thus, perceived color quality across the joint space of color gamut, dynamic range, and ambient luminance remains under-characterized, particularly for dynamic content.

Visual Perception in OST-AR. OST-AR introduces additional complexity because virtual content is optically combined with real-world backgrounds. Prior work has examined color appearance shifts [Hassani and Murdoch 2019], brightness perception [Murdoch 2020], and contrast sensitivity in OST-AR systems [Livingston et al. 2009], highlighting the role of environmental light in color blending and contrast reduction. Recent studies further investigate background-dependent color appearance [Zhang et al. 2021a], color contrast enhancement [Zhang et al. 2021b], dichoptic and interocular effects [Erickson et al. 2022a; Wang et al. 2023], supra-threshold contrast perception [Kim et al. 2025], and transparency or color scission [Murdoch 2025]. Application-driven work has studied AR interface design and usability [Gabbard et al. 2020; Hou et al. 2025; Yamin et al. 2024], while reviews summarize broader perceptual challenges in AR [Cooper 2023; Erickson et al. 2022b]. Despite this progress, existing work largely focuses on isolated perceptual effects or application-specific scenarios. A unified empirical characterization of how display gamut, luminance, ambient illumination, and background structure jointly influence perceived color quality in OST-AR remains lacking, as does a clear connection between perceptual modeling for conventional and AR displays.

[1] Code, data, and model: yunxiangzhang.github.io/projects/color-perception/index.html

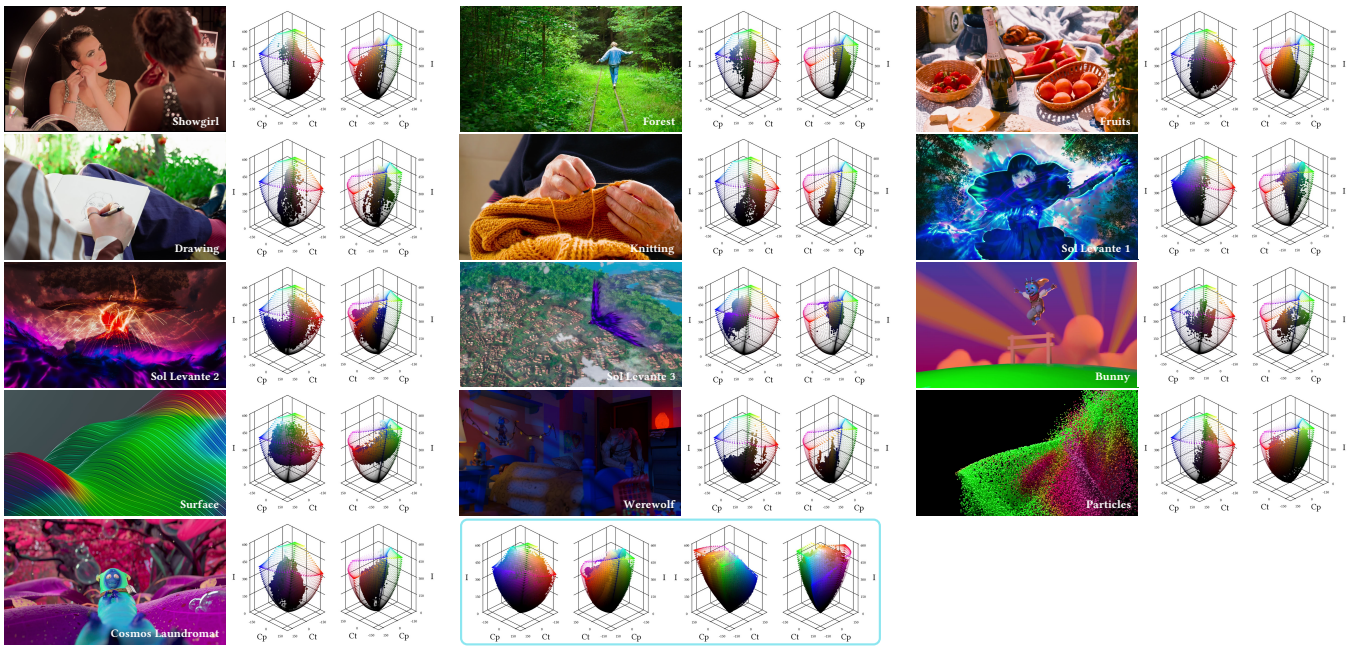

Fig. 2. *Our WCG HDR video dataset.* Each video clip is accompanied by its color distribution visualized in ICtCp color space [Dolby Laboratories 2016] from two azimuth angles, where I, Ct, and Cp denote luma, blue–yellow chroma, and red–green chroma, respectively. Video pixels are spatially subsampled with stride 64 and temporally subsampled at 2 FPS, then plotted as colored points. The dotted outer shell denotes the BT.2020 gamut boundary at 1,000-nit peak luminance and 0.001-nit black level, enclosing the reference color volume; our dataset covers 50% of this volume (see Appendix A.3 for details). The blue-boxed plots show the aggregated color distribution of the full dataset from four azimuth angles. Image credits: HdM Stuttgart, TU Ilmenau, Pexels, and Netflix.

3 Method

3.1 A Wide-Color-Gamut HDR Video Dataset

Recent advances in imaging and display technologies have accelerated the adoption of HDR/WCG media formats, driving a transition from legacy standards with limited color gamuts (e.g., BT.709 [ITU-R 2015b]) toward newer standards supporting substantially wider color reproduction (e.g., DCI-P3 [DCSS 2005] and BT.2020 [ITU-R 2015a]). However, the perceptual benefits of such expanded color capabilities remain inadequately quantified and understood.

Evaluating these benefits requires video content spanning a wide color gamut. Existing HDR video datasets, while often encoded in WCG containers (e.g., HDR10 uses BT.2020 color primaries), typically cover only part of the supported gamut. Many videos contain few high-chroma pixels, yielding effective coverage close to BT.709. Moreover, prior datasets frequently focus on homogeneous content domains, such as cinematic scenes [Froehlich et al. 2014] or casual footage [Keller et al. 2024; Song et al. 2016]. This content bias leads to uneven color distributions, with certain hues or luminance ranges being underrepresented or absent.

We gather a WCG HDR video dataset with diverse content and broad color coverage by combining selected videos from existing sources with newly created content. As illustrated in Figure 2, the dataset consists of 13 video clips spanning cinematic scenes, user-generated content, animated films, and simple graphics. It includes both real-world scenes and non-photorealistic renderings, with memory colors such as human faces, fruits, and greenery. We use the ICtCp color space [Dolby Laboratories 2016], recommended by the International Telecommunication Union (ITU) for HDR/WCG imagery [ITU-R 2025], to visualize color distributions.

Content sources: *Fruits* is from Pexels[2]; *Sol Levante* and *Cosmos Laundromat* are from Netflix Open Content[3]; *Showgirl* is from HdM-HDR-2014 [Froehlich et al. 2014]; *Forest*, *Knitting*, and *Drawing* are from AVT-VQDB-UHD-2-HDR [Keller et al. 2024]; and the remaining clips are created in Blender. All videos are manually mastered to the BT.2020 gamut using DaVinci Resolve[4] to enhance colorfulness while preserving a natural appearance. When necessary, videos are resampled/cropped to 1080p. All videos are encoded using HEVC with BT.2020 primaries, perceptual quantizer (PQ) electro-optical transfer function (EOTF), 10-bit quantization, and no chroma subsampling. Additional dataset details are provided in Appendix A.

3.2 Display Simulation

Studying user preferences across display and ambient conditions without building each hardware setup requires accurate emulation on a reference display. We introduce a display-simulation pipeline that integrates tone-mapping operators (TMOs), gamut-mapping algorithms (GMAs), and ambient-light modeling to convert source BT.2020 RGB videos to target conditions. Many TMOs [Eilertsen et al. 2017] and GMAs [Morovič and Luo 2001] have been proposed in the literature, with relative performance depending on content

[2]https://www.pexels.com/
[3]https://opencontent.netflix.com/
[4]https://www.blackmagicdesign.com/products/davinci-resolve/

and application-specific constraints. As our goal is not to compare or introduce new tone- or gamut-mapping methods, we adopt established methods that follow ITU recommendations and common display-pipeline practice. A comprehensive evaluation of alternative TMOs and GMAs is beyond the scope of this work. Figure 13 (Appendix) illustrates the processing stages.

Given a display-encoded frame I_{disp}, we first apply PQ decoding to recover linear BT.2020 RGB, denoted as I_{in}. This linearized frame then undergoes tone mapping, gamut mapping, and ambient-light simulation, producing a linear output I_{out} for the target condition.

3.2.1 Tone Mapping. We follow the ITU recommendation for HDR-to-SDR conversion [ITU-R 2021] and adopt the tone curve proposed by Chen et al. [2023]. As shown by Chen et al. [2025], using a fixed tone curve across peak luminance levels can introduce substantial contrast distortions, particularly under strong dynamic-range compression. We therefore use the optimization framework of Tariq et al. [2023] to perform scene-adaptive tone-curve adjustment.

Given the target display with black level L_{min} and peak luminance L_{max}, our TMO proceeds in four stages: tone extraction, tone mapping, color correction, and black-level simulation. Implementation details of our TMO are provided in Appendix B.

3.2.2 Gamut Mapping. Following ITU guidance on gamut conversion, we implement two GMAs: an RGB-space soft-clipping method (**RGB-SC**) [ITU-R 2017] broadly supported on hardware platforms, and an ICtCp-space constant-hue-luminance method (**ICtCp-CHL**) [Morovič 2008] that preserves hue and luminance.

RGB-SC. Given the target color gamut G_{target}, defined by its RGB primaries and white point, and the tone-mapped linear BT.2020 RGB frame I_{TM}, **RGB-SC** proceeds as follows:

(1) Apply the BT.2020-to-G_{target} transform to convert I_{TM} into linear RGB values expressed in the target color space.

$$I_{\text{RGB}} = \mathcal{M}_{\text{BT.2020}\to G_{\text{target}}}(I_{\text{TM}}), \quad \text{[Color-Space Transform]} \tag{1}$$

(2) Soft-clip I_{RGB} using the knee function $\mathcal{S}(\cdot\,;\theta_{\text{GM}})$ introduced by MacDonald et al. [2002]. The function is parameterized by $\theta_{\text{GM}} = (\kappa, C^{\text{s}}_{\text{max}}, C^{\text{t}}_{\text{max}})$, where $C^{\text{s}}_{\text{max}}$ and $C^{\text{t}}_{\text{max}}$ specify the maximum input and output values, and $\kappa \in [0,1]$ controls the knee position $C_{\text{start}} = \kappa\, C^{\text{t}}_{\text{max}}$ (see Figure 15 in the appendix). Compared to hard clipping, this approach provides a smooth transition near the target gamut boundary, maintaining most in-gamut colors unchanged while mitigating contrast compression in high-chroma areas [Braun and Fairchild 1999]. The knee-function formulation is provided in Appendix C.

$$\tilde{I}_{\text{RGB}} = \mathcal{S}(I_{\text{RGB}};\theta_{\text{GM}}), \quad \text{[RGB Soft-Clipping]} \tag{2}$$

(3) Transform $\tilde{I}_{\text{RGB}}$ back to linear BT.2020 RGB.

$$I_{\text{GM}} = \mathcal{M}_{G_{\text{target}}\to\text{BT.2020}}(\tilde{I}_{\text{RGB}}), \quad \text{[Color-Space Transform]} \tag{3}$$

Given the same inputs, **ICtCp-CHL** proceeds as follows:

(1) Transform I_{TM} into ICtCp cylindrical coordinates as I_{LCh}, where L, C, and h denote luminance, chroma, and hue.

$$I_{\text{LCh}} = \mathcal{M}_{\text{BT.2020}\to\text{ICtCp}}(I_{\text{TM}}), \quad \text{[Color-Space Transform]} \tag{4}$$

(2) Soft-clip the chroma component C along lines of constant hue and luminance using the same knee function $\mathcal{S}(\cdot\,;\theta_{\text{GM}})$. Unlike **RGB-SC**, the parameters $C^{\text{s}}_{\text{max}}(L,h)$ and $C^{\text{t}}_{\text{max}}(L,h)$ are hue- and luminance-dependent; they represent the maximum chroma for the source (BT.2020) and target (G_{target}) gamuts at each (L,h). These values are obtained from precomputed 2D lookup tables (LUTs) that map (L,h) to the corresponding gamut-boundary chroma.

$$\tilde{C} = \mathcal{S}(C;\theta_{\text{GM}}), \tilde{I}_{\text{LCh}} = [L,\tilde{C},h], \quad \text{[Chroma Soft-Clipping]} \tag{5}$$

(3) Transform $\tilde{I}_{\text{LCh}}$ back to linear BT.2020 RGB.

$$I_{\text{GM}} = \mathcal{M}_{\text{ICtCp}\to\text{BT.2020}}(\tilde{I}_{\text{LCh}}), \quad \text{[Color-Space Transform]} \tag{6}$$

3.2.3 Ambient-Light Modeling. Ambient-light influence on display viewing is commonly modeled in linear colorimetric spaces to preserve physical accuracy. For example, prior work models the optical blending of display emission and background illumination in CIE 1931 XYZ color space [Chapiro et al. 2024]. Our pipeline operates in linear BT.2020 RGB, enabling analogous simulation of ambient light reflected from flat-panel displays or transmitted through OST-AR optics. We assume diffuse ambient illumination and, for flat-panel displays, approximate the display as a diffuse surface with uniform reflectance. Under this assumption, the eye-reaching ambient contribution can be modeled as an additive term:

$$I_{\text{out}} = I_{\text{GM}} + I_{\text{amb}}, \quad \text{[Ambient-Light Simulation]} \tag{7}$$

where I_{amb} denotes the ambient light reaching the observer's eyes after accounting for panel reflectance (flat-panel) or optical transmittance (OST-AR), rather than the incident ambient luminance. This model excludes glossy or spatially varying reflections, such as those produced by localized light sources.

3.3 Direct-View User Study

Utilizing the display-simulation pipeline, we conduct a direct-view user study to measure perceived color quality across systematically varied color gamuts, dynamic ranges, and ambient luminance levels.

Participants. We recruited 41 participants (ages 22–48; 18 female), all of whom had normal or corrected-to-normal vision and passed a color-vision screening using the Ishihara test. The study protocol was approved by an Institutional Review Board (IRB), and all participants provided written informed consent prior to participation.

Hardware. For accurate wide-gamut color reproduction, we used a 31.1" EIZO CG3145 display with DCI-4K resolution, 60 Hz refresh rate, contrast ratio up to 1,000,000:1, 1,000-nit peak luminance, 10-bit color depth, and over 99% DCI-P3 coverage. It was calibrated to BT.2020 primaries and a D65 white point using a Konica Minolta CS-2000 spectroradiometer. A chin rest aligned observers with the display and fixed the viewing distance at 114.3 cm (0.875 diopters, D), yielding a $32.0^\circ \times 18.3^\circ$ field of view. Videos were upsampled from 1080p to 4K UHD via 2×2 pixel replication and center-aligned on the display, yielding an effective angular resolution of 59.0 pixels per degree (ppd), consistent with 20/20 visual acuity (60 ppd). Figure 4 (right) illustrates the hardware configuration.

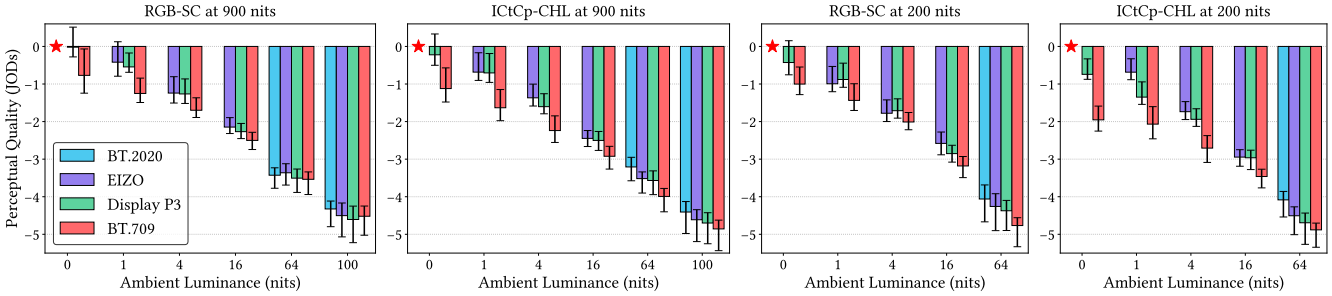

Fig. 3. *Direct-view user study results.* Perceptual quality plotted against ambient luminance for each color gamut. Panels correspond to peak-luminance × GMA conditions. Error bars indicate 95% confidence intervals. The red star marks the block-wise reference condition (EIZO, zero ambient, **RGB-SC**).

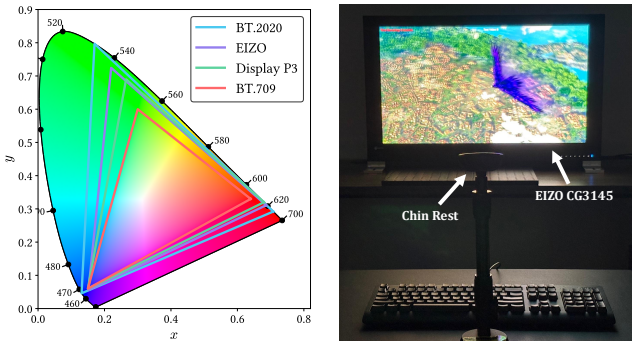

Fig. 4. *Direct-view user study setup.* EIZO native gamut shown in the CIE 1931 *xy* chromaticity diagram (left) and hardware configuration (right).

Stimuli. We selected 9 scenes: *Showgirl*, *Sol Levante 1–2*, *Bunny*, *Forest*, *Particles*, *Knitting*, *Werewolf*, and *Surface*. Each video was tone-mapped to two peak luminance levels, 200 and 900 nits, with a black level of 0.001 nit using the TMO in Section 3.2.1. The tone-mapped videos were mapped to four color gamuts of decreasing size, BT.2020, EIZO, Display P3[5], and BT.709 (Figure 4, left), using the GMAs in Section 3.2.2. As all gamuts share a D65 white point, chromatic adaptation was not required. Ambient light reaching the observer's eyes was modeled as D65 white with luminance L_{amb}. To guide the selection of ambient levels, we define additive contrast as:

$$c_{\text{add}} = (L_{\text{max}} + L_{\text{amb}})/(L_{\text{min}} + L_{\text{amb}}), \quad \text{[Additive Contrast]} \quad (8)$$

We estimated additive contrast for representative display models under typical indoor and outdoor lighting [Zhang et al. 2023], assuming 1–5% panel reflectance for flat panels and 20–90% optical transmittance for OST-AR (see Appendices D.1 and D.2 for details). Based on these estimates and a pilot study with two expert participants, we sampled six eye-reaching ambient luminance levels for L_{amb}: 0, 1, 4, 16, 64, and 100 nits. The 0-nit condition isolates intrinsic perceptual differences across color gamuts under near-dark viewing. Although the EIZO display cannot reproduce the full BT.2020 gamut, more than 93% of the ambient-blended BT.2020 gamut falls within the EIZO gamut for L_{amb} = 64–100 nits, enabling its simulation. In total, this yielded 648 stimuli: 9 scenes × 2 GMAs × 36 conditions. Appendix D.3 details the stimuli selection process.

[5]Display P3 shares the same RGB primaries as DCI-P3 but uses a D65 white point.

Procedures. We used a two-interval forced-choice (2IFC) protocol. Each trial included a reference video and two test videos, and participants were instructed to select the test video that better matched the reference in color fidelity. The reference was always presented first. Participants could toggle between the three videos in any order and as many times as needed, but were required to view all three at least once before responding. To prevent direct visual comparisons across conditions, we inserted a 500-ms gray screen at the mean luminance of the incoming stimulus between video switches, which also provided sufficient adaptation time prior to each stimulus onset [Hayhoe et al. 1987]. A 15-second trial timer was displayed to discourage excessively long trials but was not enforced. Participants completed an 18-trial training session prior to the main experiment.

We allowed comparisons across color gamuts, ambient luminance, and GMAs, but not across scenes or peak luminance, because color perception is luminance-dependent [Hunt 1952; Purdy 1931] and a single reference condition across multiple peak luminance levels would be ambiguous. For each scene–peak-luminance block, the reference was the stimulus with the largest effective gamut (EIZO, zero ambient, **RGB-SC**); its **ICtCp-CHL** counterpart was excluded to avoid ambiguity. Under this design, exhaustive pairwise comparison would require 10,854 trials. We instead used ASAP [Mikhailiuk et al. 2021] to adaptively schedule comparisons that maximize expected information gain, linking all conditions while reducing the study to 612 trials. To further reduce participant workload, the study was divided into three sessions of 204 trials, each covering three scenes. Participants were assigned to one session, rotated to balance scene coverage. The average completion time was 50.6 minutes.

Results. Pairwise comparison results were converted to a perceptual quality scale in just-objectionable-difference (JOD) units [Perez-Ortiz et al. 2019] using `pwcmp`.[6] Observers with an interquartile score above 1.5 were excluded as outliers (2 out of 41 participants). JOD differences map to observer preference probabilities (see Figure 17 in the appendix). The reference in each block was set to 0 JOD.

The scene-aggregated JOD scores are summarized in Figure 3, grouped by peak luminance and GMA. Error bars show 95% confidence intervals estimated from 500 bootstrap samples. Qualitatively, perceived color quality decreases as gamut size shrinks and ambient luminance increases, consistent with the reduction in effective

[6]https://github.com/mantiuk/pwcmp/

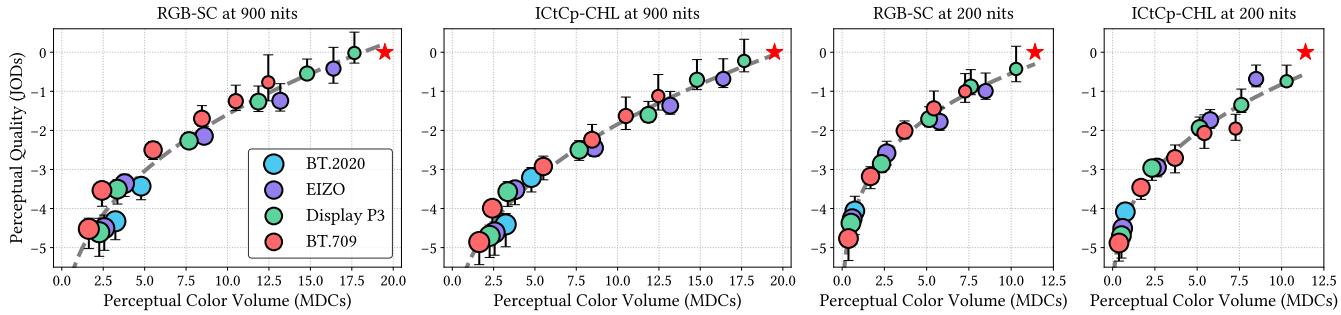

Fig. 5. *Analytical model relating perceptual color volume to perceived color quality.* JOD scores are plotted against the corresponding MDC values for all conditions studied in Section 3.3, separated by peak luminance and GMA. Dashed curves show the fitted models. Marker size increases with ambient luminance.

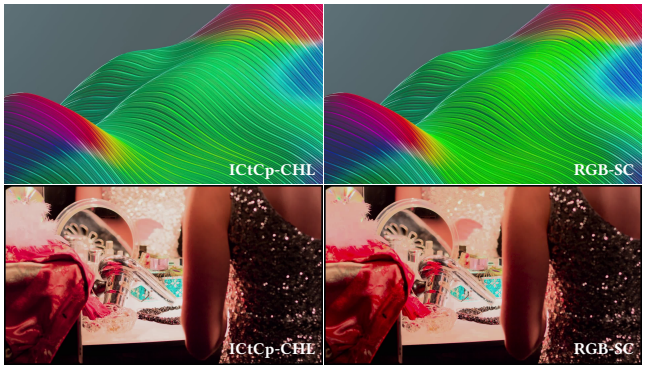

Fig. 6. *Comparison of GMAs.* When mapping to narrower gamuts, **ICtCp-CHL** strictly preserves hue but can substantially reduce chroma at certain hue angles, whereas **RGB-SC** allows hue shifts that retain higher chroma and was generally preferred in our study. The examples are BT.2020 content gamut-mapped down to BT.709. Image credit: HdM Stuttgart.

color gamut under both factors. In addition, perceptual differences between gamuts diminish with increasing ambient luminance.

An N-way analysis of variance (ANOVA) on JOD scores revealed significant main effects of ambient luminance, scene, color gamut, GMA, and peak luminance (all $p < .005$; full statistics in Table 5). Among these factors, ambient luminance accounted for the largest share of variance. A significant interaction between peak luminance and ambient luminance ($p < .001$) shows that the impact of ambient light depends on available display luminance, with the 900-nit condition partially mitigating ambient-induced degradation compared with the 200-nit condition. A significant color-gamut × ambient-luminance interaction ($p < .001$) indicates that the perceptual advantage of wider gamuts diminishes as ambient luminance increases. The color-gamut × GMA interaction ($p < .01$) suggests that GMA performance depends on the target gamut. While **RGB-SC** generally outperforms **ICtCp-CHL**, this difference increases for narrower gamuts, where stronger gamut compression is required. This behavior reflects an inherent trade-off between hue preservation and chroma retention in the two GMA designs. **ICtCp-CHL** maps out-of-gamut colors toward the target gamut boundary along constant-hue-luminance directions in ICtCp color space, strictly preserving hue but potentially causing substantial chroma reduction at certain hue angles. In contrast, **RGB-SC** projects out-of-gamut colors to the closest points on the target gamut boundary in RGB color space, permitting hue shifts that help retain higher chroma, an outcome that was perceptually preferred in our study. Figure 6 illustrates the trade-off between **ICtCp-CHL** and **RGB-SC**.

3.4 Perceptual Model

To quantify ambient-induced gamut loss, as illustrated in Figure 13 (Appendix), we adopt Dolby's perceptual color volume (PCV) formulation [2023]. PCV operates in ICtCp color space, which is designed to be approximately linear in just-noticeable-difference (JND) units and perceptually uniform for HDR/WCG imagery. Following this formulation, we convert ICtCp coordinates to JND units and compute color volume in millions of distinguishable colors (MDCs) for all conditions studied in Section 3.3. Appendix A.3 details this process.

Correlation analysis showed strong agreement between MDC and JOD. For **RGB-SC**, we found linear (Pearson's $r = 0.868$, $p < .001$), rank-order (Spearman's $r = 0.902$, $p < .001$), and distance correlations ($r = 0.869$, $p < .001$); for **ICtCp-CHL**, we found Pearson's $r = 0.884$ ($p < .001$), Spearman's $r = 0.920$ ($p < .001$), and distance correlation $r = 0.884$ ($p < .001$). These results indicate that PCV provides a robust predictor of perceived color quality and motivate an analytical model mapping MDC to JOD. As shown in Figure 5, perceptual quality increases with PCV and gradually saturates as PCV approaches the reference condition, exhibiting a quasi power-law relationship. To model this behavior, we evaluated three candidate functional forms—power, logarithmic, and rational—fitted using nonlinear least squares. We compared the resulting models using Akaike Information Criterion (AIC) and adopted the power model.

Although peak luminance is incorporated into the MDC computation, Figure 5 suggests that the JOD–MDC relationship remains luminance-dependent. Because two peak-luminance settings are insufficient to robustly fit a unified model, we first fit separate models at each peak luminance and then interpolate their parameters in log-luminance space. The resulting unified model is defined as:

$$\begin{aligned} Q(V,L) &= k_1(L)\cdot V^{\,k_2(L)} + k_3(L), \quad L \in [L_{\max}^{\mathrm{SDR}}, L_{\max}^{\mathrm{HDR}}] \\ k_i(L) &= (1-\alpha)k_i^{\mathrm{SDR}} + \alpha k_i^{\mathrm{HDR}}, \qquad i \in \{1,2,3\} \\ \alpha &= \log(L/L_{\max}^{\mathrm{SDR}}) \,/\, \log(L_{\max}^{\mathrm{HDR}}/L_{\max}^{\mathrm{SDR}}) \end{aligned} \tag{9}$$

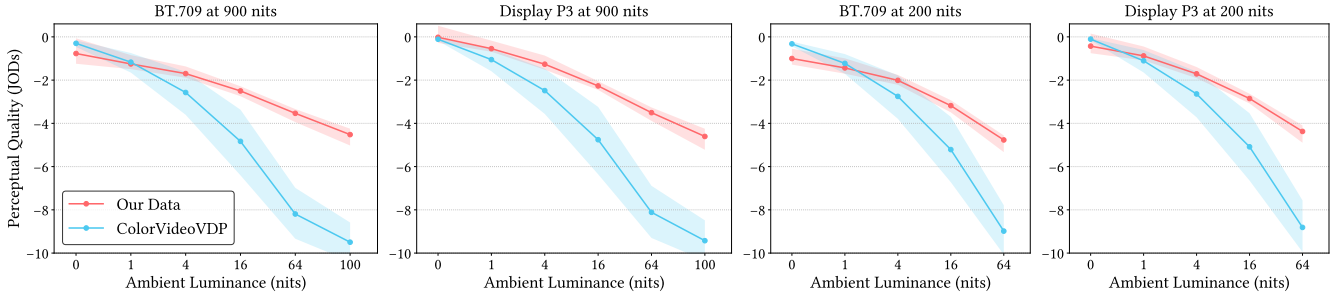

Fig. 7. *Evaluation of ColorVideoVDP on direct-view study stimuli.* Comparison of our user-study measurements with ColorVideoVDP [Mantiuk et al. 2024] predictions, both aggregated across scenes. Panels correspond to peak-luminance × gamut conditions. Shaded bands indicate 95% confidence intervals.

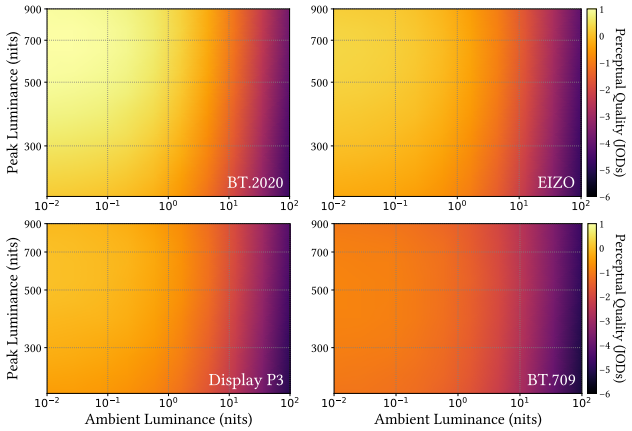

Fig. 8. *Model-predicted color quality across display and ambient conditions.* The fitted model for **RGB-SC** is evaluated across ambient luminance and peak luminance and visualized as JOD heatmaps for four color gamuts.

where Q and V denote perceived color quality (JOD) and perceptual color volume (MDC). $L_{\max}^{\mathrm{SDR}}$ and $L_{\max}^{\mathrm{HDR}}$ denote the two peak-luminance levels: 200 nits for SDR and 900 nits for HDR.

Separate models were fit for the two GMAs. The model achieved a root-mean-square error (RMSE) of 0.237 and a coefficient of determination $R^2 = 0.973$ for **RGB-SC**, and an RMSE of 0.216 and an R^2 of 0.976 for **ICtCp-CHL**. The fitted JOD–MDC relationships are shown as dashed curves in Figure 5. The unified model is visualized as JOD heatmaps across ambient luminance and peak luminance in Figure 8. Details on model selection and fitted model parameters (k_i^{SDR} and k_i^{HDR}, $i \in \{1, 2, 3\}$) are provided in Appendix E.

To position our model relative to prior metrics, we evaluate ColorVideoVDP [Mantiuk et al. 2024], a state-of-the-art visual difference predictor for HDR/WCG videos, on our direct-view study stimuli. ColorVideoVDP accounts for the physical display and viewing configuration, including resolution, dynamic range, viewing distance, and ambient illumination, and reports predictions in JOD units. We evaluate both peak-luminance levels for the BT.709 and Display P3 gamuts using **RGB-SC**-mapped stimuli. Compared to our model, ColorVideoVDP exhibits substantially larger errors relative to the user-study measurements: RMSE is 2.64 vs. 0.24.

As shown in Figure 7, ColorVideoVDP underestimates perceptual differences between BT.709 and Display P3 under near-dark conditions, while substantially overestimating ambient-induced quality degradation as ambient luminance increases. Its predictions also exhibit considerably larger variation across scenes. One possible explanation is that ColorVideoVDP was primarily calibrated on distortions such as light-source nonuniformity, reduced clarity and contrast, and chroma subsampling, rather than perceptual differences induced by systematically varying display gamut and ambient illumination. This comparison is nonetheless limited to two gamuts, two peak-luminance levels, and the stimuli from our direct-view study. A broader evaluation across video content, display characteristics, and viewing conditions is therefore needed to determine whether these failure modes generalize.

3.5 Color Enhancement

The preceding sections show that increasing ambient luminance rapidly shrinks the effective color gamut and degrades perceived color quality relative to a display's native capability. To mitigate this ambient-induced degradation, we introduce three color-enhancement algorithms: an RGB-space color expansion method (**RGB-CE**), a Schlick-style color correction method (**Schlick-CC**) [Schlick 1995], and an ICtCp-space chroma expansion method (**ICtCp-CE**). Rather than attempting to recover the original content colors, these methods aim to better utilize available display gamut to improve color quality while maintaining a natural appearance.

RGB-CE and **Schlick-CC** operate in RGB space and require minimal computation, making them suitable for resource-constrained display pipelines. **RGB-CE** is the most lightweight, as it can be implemented as a simple linear transformation, whereas **Schlick-CC** additionally requires a cross-channel maximum operation. **ICtCp-CE** works in the more perceptually uniform ICtCp color space and preserves hue, but requires a computationally intensive LUT-based implementation. In addition, **Schlick-CC** and **ICtCp-CE** incorporate an ambient term, which enables controllable ambient-aware enhancement. When additive contrast c_{add} is very low, fully compensating for the ambient contribution may introduce chromatic artifacts and distort natural color appearance. We therefore compensate only partially by defining the compensated ambient level L_{comp}. Together, these methods span lightweight RGB manipulations to



Fig. 9. *Comparison of color-enhancement algorithms.* The original BT.709 content (a) is shown at 280-nit peak luminance. For (b)–(e), 40-nit ambient luminance is simulated. For the enhancement algorithms (c)–(e), the output display is assumed to have a Display P3 gamut. Image credits: Pexels and Netflix.

perceptually constrained chroma expansion, trading computational costs against visual quality and controllability.

RGB-CE. Given the target black level $L_{\min}$, peak luminance $L_{\max}$, and linear RGB input I_{RGB} in the target color space G_{target}:

(1) Compute the luminance component Y from I_{RGB} using the RGB-to-XYZ transform of G_{target}.

$$I_{\mathrm{XYZ}} = [X, Y, Z] = \mathcal{M}_{G_{\mathrm{target}}\to\mathrm{XYZ}}(I_{\mathrm{RGB}}), \quad \text{[Luminance]} \tag{10}$$

(2) Perform color expansion with scaling factor $s \in [1, 2]$.

$$\tilde{I}_{\mathrm{RGB}} = s\, I_{\mathrm{RGB}} + (1 - s)Y, \quad \text{[Color Expansion]} \tag{11}$$

(3) Clip the expanded RGB values to the valid dynamic range.

$$I_{\mathrm{CE}} = \left[\tilde{I}_{\mathrm{RGB}}\right]_{L_{\min}}^{L_{\max}}, \quad \text{[RGB Clipping]} \tag{12}$$

where $[\cdot]_{L_{\min}}^{L_{\max}}$ denotes element-wise clipping.

Schlick-CC. Given the same inputs as **RGB-CE** and the compensated ambient term L_{comp}, **Schlick-CC** proceeds as follows:

(1) Similar to the tone-extraction step in tone mapping, extract the input tone as the channel-wise maximum of I_{RGB}. While different luminance proxies can be used, recent work shows that the maximum RGB channel helps avoid out-of-gamut clipping and associated color distortions [Burke et al. 2020].

$$T_{\mathrm{RGB}} = \max(I^R_{\mathrm{RGB}}, I^G_{\mathrm{RGB}}, I^B_{\mathrm{RGB}}), \quad \text{[Tone Extraction]} \tag{13}$$

(2) Apply Schlick-style color correction [1995] to scale the RGB channels while preserving their relative ratios.

$$\tilde{I}_{\mathrm{RGB}} = ((T_{\mathrm{RGB}} + L_{\mathrm{comp}})/T_{\mathrm{RGB}})\, I_{\mathrm{RGB}}, \quad \text{[Color Correction]} \tag{14}$$

(3) Remove L_{comp} and clip to the valid dynamic range.

$$I_{\mathrm{CE}} = \left[\tilde{I}_{\mathrm{RGB}} - L_{\mathrm{comp}}\right]_{L_{\min}}^{L_{\max}}, \quad \text{[RGB Clipping]} \tag{15}$$

ICtCp-CE. Given the same inputs as **Schlick-CC**:

(1) Transform I_{RGB} and $I_{\mathrm{RGB}} + L_{\mathrm{comp}}$ into ICtCp color space and express them in cylindrical coordinates.

$$\begin{aligned} I_{\mathrm{LCh}} &= [L, C, h] = \mathcal{M}_{G_{\mathrm{target}}\to\mathrm{ICtCp}}(I_{\mathrm{RGB}}), \\ \tilde{I}_{\mathrm{LCh}} &= [\tilde{L}, \tilde{C}, \tilde{h}] = \mathcal{M}_{G_{\mathrm{target}}\to\mathrm{ICtCp}}(I_{\mathrm{RGB}} + L_{\mathrm{comp}}), \end{aligned} \tag{16}$$

(2) Query the maximum allowable chroma $C^{\mathrm{t}}_{\max}$ at $(\tilde{L}, h)$ within the effective gamut under ambient level L_{comp} from precomputed 2D LUTs, and expand chroma toward this boundary.

$$C_{\mathrm{out}} = \min(C, C^{\mathrm{t}}_{\max}(\tilde{L}, h)), \quad \text{[Chroma Expansion]} \tag{17}$$

(3) Transform $[\tilde{L}, C_{\mathrm{out}}, h]$ back to the target RGB space G_{target} and remove the compensated ambient term L_{comp}.

$$I_{\mathrm{CE}} = \mathcal{M}_{\mathrm{ICtCp}\to G_{\mathrm{target}}}([\tilde{L}, C_{\mathrm{out}}, h]) - L_{\mathrm{comp}}, \tag{18}$$

Figure 9 visually compares the enhancement results. Additional details on the parameter settings are provided in Appendix F.

4 Evaluation

Visual perception in OST-AR differs from conventional direct-view displays because virtual content can appear at different focal depths from the real-world background and is optically blended with environmental light, producing transparency-like appearances [Chapiro et al. 2024]. Prior work suggests that observers may partially discount either display-emitted or environmental light in brightness [Murdoch 2020] and color perception [Zhang 2022], consistent with scission effects [Metelli 1974]. We therefore construct an OST-AR viewing setup to evaluate how well the model derived from direct-view data (Section 3.4) generalizes to OST-AR conditions, and to test whether background structure modulates perceived color quality. We also quantify the perceptual benefits of the color-enhancement algorithms introduced in Section 3.5 under OST-AR viewing.

Hardware. We used two 31.1" EIZO CG3146 displays, which share the specifications of the EIZO CG3145, to present virtual content, and a 76" DynaScan DK751DH5 display to present the background. The EIZO displays were placed to the left and right of the observer. Two neutral-response beam splitters (Edmund Optics) were placed in front of the observer and oriented at 45° toward the EIZO displays, forming an optical see-through setup. The beam splitters reflected 80% of the EIZO-emitted light toward the observer and transmitted 20% of the DynaScan-emitted light. Because the DynaScan has a black level of nearly 0.5 nits, we placed a neutral-density (ND) filter between it and the beam splitters to enable near-dark backgrounds. The system produced effective viewing distances of 83 cm (1.20 D) for the virtual content and 210 cm (0.476 D) for the background, i.e., a 0.724 D separation between the virtual and background planes. Prior work has shown this separation to be sufficient for clear depth

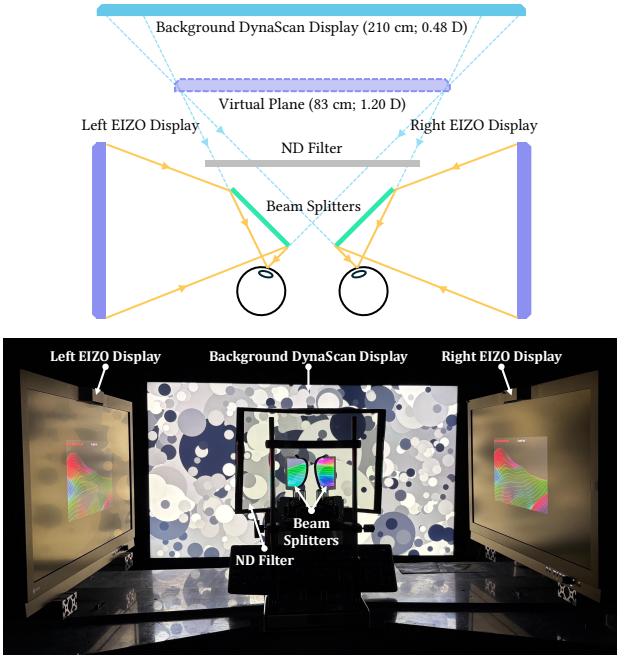

Fig. 10. *OST-AR system configuration and viewing setup.* Top: schematic of our OST-AR viewing system. Bottom: photograph of the implemented setup.

separation [Chapiro et al. 2024]. The setup created a field of view of 46° × 26° and an effective angular resolution of 44.2 ppd, consistent with commercial AR devices. All three displays were calibrated using a CS-2000 spectroradiometer through the complete optical path, including the beam splitters and ND filter, such that the calibration accounted for the optical effects introduced by the hardware setup. Figure 10 illustrates the resulting OST-AR system configuration.

4.1 Perceived Color Quality under OST-AR Viewing

A second user study with 18 participants was conducted to evaluate model generalization to OST-AR and examine background effects.

Stimuli. We selected 4 scenes from Section 3.3: *Showgirl*, *Surface*, *Forest*, and *Sol Levante 1*. Each video was tone-mapped to two peak luminance levels, 200 and 800 nits, with a 0.001-nit black level. The tone-mapped videos were mapped to two color gamuts, Display P3 and BT.709, using **RGB-SC**. Before presentation, video luminance values were scaled by 1.25 to compensate for the 80% beam-splitter reflectance. We sampled four ambient levels for L_{amb}: 0.01, 5, 25, and 100 nits. Two backgrounds were used (Figure 16, Appendix): *flat* (uniform D65 white) and *leaves* (a high-contrast pattern following Chapiro et al. [2024]). This yielded 128 stimuli in total: 4 scenes × 32 conditions. Appendix G describes the stimuli selection process. The same 2IFC and ASAP study procedure as explained in Section 3.3 were adopted to schedule pairwise comparisons. Comparisons were allowed across color gamuts, ambient luminance, and backgrounds, but not across scenes or peak luminance. The block-wise reference was set to the Display P3, 0.01-nit ambient, *flat* condition.

Results. We applied the same JOD conversion and outlier-filtering procedures as in Section 3.3. The scene-aggregated JOD scores are

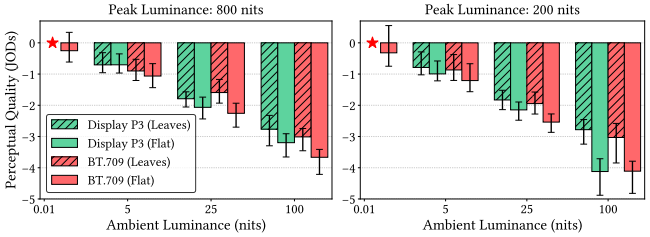

Fig. 11. *OST-AR color-quality study results.* Perceptual quality plotted against ambient luminance. Error bars show 95% confidence intervals. The red star marks the block-wise reference (Display P3, 0.01-nit ambient, *flat*).

shown in Figure 11. Consistent with the direct-view study, perceived color quality decreases with shrinking gamut size and increasing ambient luminance. We applied the model derived from direct-view data to predict results from this OST-AR study. Correlation analysis between model predictions and measured JOD scores for the *flat* background showed strong agreement: Pearson's linear correlation ($r = 0.990$, $p < .001$), Spearman's rank-order correlation ($r = 0.993$, $p < .001$), and distance correlation ($r = 0.992$, $p < .001$). These results suggest that the model derived from direct-view data transfers well to the controlled OST-AR conditions evaluated here.

An N-way ANOVA revealed significant main effects of ambient luminance, background pattern, color gamut, peak luminance, and scene (all $p < .05$; see full statistics in Table 8). Ambient luminance again accounted for the largest share of variance, indicating its dominant influence on perceived color quality under OST-AR viewing. The significant main effect of background type ($p < .001$) further suggests that background structure can affect the perceived color quality of virtual content: the *leaves* background produced substantially higher JOD scores than the *flat* background. We hypothesize that the high-contrast structure of the *leaves* background facilitates perceptual separation of the background from the virtual content, allowing observers to partially discount the ambient contribution. Previous work has observed similar effects in color-matching tasks [Hassani 2019]. A significant ambient-luminance and background-type interaction ($p < .001$) indicates that this effect becomes more pronounced with increasing ambient luminance.

These findings, however, should be interpreted within the scope of our controlled OST-AR setup. The beam-splitter system enables calibrated control of display gamut and background luminance, but does not reproduce the full optical characteristics of commercial AR headsets, such as spatial and spectral nonuniformity introduced by waveguides. Moreover, the two static backgrounds represent only a small subset of the complex and dynamic environments encountered in real-world AR. A broader evaluation across AR devices and more diverse, dynamic backgrounds is therefore needed to establish how well these findings generalize to real-world AR environments.

4.2 Color Enhancement under OST-AR Viewing

A third user study with 16 participants was conducted to evaluate the color-enhancement algorithms (Section 3.5) under OST-AR viewing.

Stimuli. We used the remaining 4 scenes: *Fruits*, *Drawing*, *Cosmos Laundromat*, and *Sol Levante 3*. We sampled 3 contrast levels

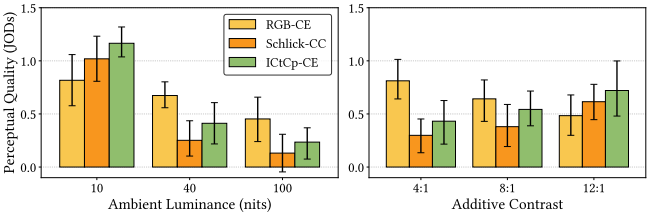

Fig. 12. *OST-AR color-enhancement study results grouped by L_{amb} and c_{add}.* In each block, the no-enhancement baseline was set to 0 JOD.

representative of OST-AR displays, 12:1, 8:1, and 4:1 [Chapiro et al. 2024], and 3 ambient levels, 10, 40, and 100 nits. Two combinations were excluded since they required impractical or infeasible peak luminance. Each video was tone-mapped to the specified peak luminance with a 0.001-nit black level, then mapped to the BT.709 gamut using **RGB-SC**. As in Section 4.1, video luminance was scaled by 1.25 before presentation, and the *flat* background was used.

Besides the three color-enhancement algorithms, we included a no-enhancement baseline in which no ambient-light compensation was applied. For the enhancement algorithms, the target display was assumed to have a Display P3 gamut, offering additional gamut headroom to compensate for ambient-induced color-fidelity losses. This setup reflects practical scenarios where visual content is mastered for a standard gamut such as BT.709, while display devices use custom color primaries that exceed these standards (e.g., the EIZO gamut shown in Figure 4, left). In total, this yielded 112 stimuli: 4 scenes × 28 conditions. Appendix H describes the stimuli selection process. We employed the same procedure as in Section 4.1, allowing comparisons only across enhancement methods. The block-wise reference was set to the original video without ambient light.

Results. The JOD scores are presented in Figure 12. In each block, the no-enhancement baseline was set to 0 JOD. All three methods consistently improved perceived color quality relative to the baseline. **ICtCp-CE** and **Schlick-CC** performed better at lower ambient luminance and higher additive contrast, whereas **RGB-CE** showed the opposite trend. **RGB-CE** and **Schlick-CC** operate directly in RGB color space and are more lightweight, making them suitable for resource-constrained display pipelines. **ICtCp-CE** requires a more compute-heavy LUT-based implementation but preserves hue and luminance, offering a more perceptually constrained solution when higher computational cost is acceptable. These results highlight a practical trade-off between visual quality and system efficiency.

An N-way ANOVA on JOD scores revealed significant main effects of all three factors ($p < .001$; full statistics in Table 10). Enhancement method accounted for the largest share of variance, confirming measurable perceptual improvements over the no-enhancement baseline. Significant scene × enhancement-method ($p = .029$) and viewing-condition × enhancement-method ($p = .031$) interactions further indicate that relative effectiveness depends on both scene content and the severity of ambient-induced color degradation.

These enhancement methods also have limitations. **Schlick-CC** uses a simple Schlick-style color correction, while more sophisticated methods may better preserve hue and produce more natural appearances [Artusi et al. 2018; Mantiuk et al. 2009]. **ICtCp-CE** explicitly preserves hue and luminance but relies on scene-dependent parameters, whose robustness across broader content remains to be evaluated. Future work could incorporate more advanced color correction and automatic, content-adaptive parameter selection.

5 Discussion and Future Work

This work has several practical applications. The model established in Section 3.4 enables color-quality comparisons across display specifications and viewing conditions; the color-enhancement algorithms introduced in Section 3.5 demonstrate how unused display-gamut headroom can mitigate ambient-induced color degradation; and our findings can inform display-system design by jointly considering peak luminance, color gamut, and expected ambient illumination.

Our work also has several limitations. While our studies provide a systematic characterization of the interactions between display and ambient conditions on perceived color quality, only limited sampling of this high-dimensional space was feasible. Future work may further refine this sampling. We assume a D65 ambient illuminant; while the model can be adapted to other illuminants, further validation is needed under chromatic illumination. Our controlled beam-splitter setup enables calibrated measurements but does not reproduce the full optical characteristics of commercial AR headsets, such as spatial and spectral nonuniformity introduced by waveguides. Moreover, the two static backgrounds represent only a small subset of complex, dynamic real-world environments. Finally, our stimuli are generated using specific tone- and gamut-mapping methods; alternative mappings may affect perceived color quality and warrant further evaluation.

Overall, we characterize how color gamut, dynamic range, and ambient luminance jointly shape perceived color quality across conventional and controlled OST-AR viewing conditions, and introduce a perceptual model and color-enhancement methods informed by these measurements. Our findings provide a foundation for perceptually adaptive display systems that jointly account for display capabilities and real-world viewing conditions.

Acknowledgments

We thank Maurizio Nitti for creating four HDR/WCG videos in our dataset (*Bunny*, *Surface*, *Werewolf*, and *Particles*). We are grateful to Fartun Sheygo and Rachana Thorsten for administering the user studies, to all participants for their time, and to Ali Yousefi, Katelyn Troastle, John Hill, Chelsea Penter, and Romain Bachy for assistance with user-study logistics. Fruitful discussions with Krzysztof Wolski, Rafał Mantiuk, Ajit Ninan, Patrick Griffis, Jon McElvain, Lili Zhang, and Volga Aksoy helped shape this work. We also appreciate Reza Saeedpour and Ki Han for their support with the CS-2000 spectroradiometer. This research was partially supported by National Science Foundation grants #2232817 and #2225861.

References

Alessandro Artusi, Tania Pouli, Francesco Banterle, and Ahmet Oğuz Akyüz. 2018. Automatic saturation correction for dynamic range management algorithms. *Signal Processing: Image Communication* 63 (2018), 100–112.

Shahrukh Athar, Thilan Costa, Kai Zeng, and Zhou Wang. 2019. Perceptual Quality Assessment of UHD-HDR-WCG Videos. In *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 1740–1744.

Tunç Ozan Aydin, Nikolce Stefanoski, Simone Croci, Markus Gross, and Aljoscha Smolic. 2014. Temporally coherent local tone mapping of HDR video. *ACM Transactions on Graphics (TOG)* 33, 6 (2014), 1–13.

Ronan Boitard, Kadi Bouatouch, Remi Cozot, Dominique Thoreau, and Adrien Gruson. 2012. Temporal coherency for video tone mapping. In *Applications of Digital Image Processing XXXV*, Vol. 8499. SPIE, 113–122.

Nicolas Bonnier, Francis Schmitt, Michael Hull, and Christophe Leynadier. 2007. Spatial and Color Adaptive Gamut Mapping: A Mathematical Framework and Two New Algorithms. In *Color and Imaging Conference*, Vol. 15. Society of Imaging Science and Technology, 267–272.

Gustav J Braun and Mark D Fairchild. 1999. General-Purpose Gamut-Mapping Algorithms: Evaluation of Contrast-Preserving Rescaling Functions for Color Gamut Mapping. In *Color and Imaging Conference*, Vol. 7. Society of Imaging Science and Technology, 167–172.

Adam Burke, Michael D Smith, and Michael Zink. 2020. Color Volume and Hue-Preservation in HDR Tone Mapping. *SMPTE Motion Imaging Journal* 129, 4 (2020), 45–50.

Peibei Cao, Rafal K Mantiuk, and Kede Ma. 2024. Perceptual Assessment and Optimization of HDR Image Rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 22433–22443.

Alexandre Chapiro, Dongyeon Kim, Yuta Asano, and Rafał K Mantiuk. 2024. AR-DAVID: Augmented Reality Display Artifact Video Dataset. *ACM Transactions on Graphics (TOG)* 43, 6 (2024), 1–11.

Bin Chen, Akshay Jindal, Michal Piovarči, Chao Wang, Hans-Peter Seidel, Piotr Didyk, Karol Myszkowski, Ana Serrano, and Rafał K Mantiuk. 2023. The Effect of Display Capabilities on the Gloss Consistency between Real and Virtual Objects. In *SIGGRAPH Asia 2023 Conference Papers*. 1–11.

Kenneth Chen, Nathan Matsuda, Jon McElvain, Yang Zhao, Thomas Wan, Qi Sun, and Alexandre Chapiro. 2025. What Is HDR? Perceptual Impact of Luminance and Contrast in Immersive Displays. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*. 1–11.

Anustup Choudhury and Scott Daly. 2020. Comparing Common Still Image Quality Metrics in Recent High Dynamic Range (HDR) and Wide Color Gamut (WCG) Representations. *Electronic Imaging* 32 (2020), 1–9.

Anustup Choudhury, Robert Wanat, Jaclyn Pytlarz, and Scott Daly. 2021. Image Quality Evaluation for High Dynamic Range and Wide Color Gamut Applications using Visual Spatial Processing of Color Differences. *Color Research & Application* 46, 1 (2021), 46–64.

Emily A Cooper. 2023. The Perceptual Science of Augmented Reality. *Annual Review of Vision Science* 9, 1 (2023), 455–478.

DCSS. 2005. *Digital Cinema System Specification.* DCI Specification Version 1.0. Digital Cinema Initiatives, LLC. https://www.dcimovies.com/dci-specification

Dolby Laboratories. 2016. *What Is ICtCp?* White Paper. Dolby. https://professional.dolby.com/siteassets/pdfs/ictcp_dolbywhitepaper_v071.pdf

Dolby Laboratories. 2023. *Perceptual Color Volume: Measuring the Distinguishable Colors of HDR and WCG Displays.* White Paper. Dolby. https://professional.dolby.com/siteassets/pdfs/measuringperceptualcolorvolume_v07.253.pdf

Fabienne Dugay, Ivar Farup, and Jon Y Hardeberg. 2008. Perceptual Evaluation of Color Gamut Mapping Algorithms. *Color Research & Application* 33, 6 (2008), 470–476.

Gabriel Eilertsen, Rafal Konrad Mantiuk, and Jonas Unger. 2017. A Comparative Review of Tone-Mapping Algorithms for High Dynamic Range Video. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 565–592.

Austin Erickson, Gerd Bruder, and Gregory F Welch. 2022a. Analysis of the Saliency of Color-Based Dichoptic Cues in Optical See-Through Augmented Reality. *IEEE transactions on visualization and computer graphics* 29, 12 (2022), 4936–4950.

Austin Erickson, Kangsoo Kim, Gerd Bruder, and Greg Welch. 2022b. A Review of Visual Perception Research in Optical See-Through Augmented Reality. In *Proceedings of ICAT-EGVE 2020-International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*.

Mark D Fairchild. 2013. *Color Appearance Models.* John Wiley & Sons.

Ivar Farup, Carlo Gatta, and Alessandro Rizzi. 2007. A Multiscale Framework for Spatial Gamut Mapping. *IEEE Transactions on Image Processing* 16, 10 (2007), 2423–2435.

Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, and Harald Brendel. 2014. Creating Cinematic Wide Gamut HDR-Video for The Evaluation of Tone Mapping Operators and HDR-Displays. In *Digital Photography X*, Vol. 9023. SPIE, 279–288.

Joseph L Gabbard, Missie Smith, Coleman Merenda, Gary Burnett, and David R Large. 2020. A Perceptual Color-Matching Method for Examining Color Blending in Augmented Reality Head-Up Display Graphics. *IEEE Transactions on Visualization and Computer Graphics* 28, 8 (2020), 2834–2851.

Nargess Hassani. 2019. *Modeling Color Appearance in Augmented Reality.* Rochester Institute of Technology.

Nargess Hassani and Michael J Murdoch. 2019. Investigating Color Appearance in Optical See-Through Augmented Reality. *Color Research & Application* 44, 4 (2019), 492–507.

Mary M Hayhoe, Norma I Benimoff, and DC Hood. 1987. The Time-Course of Multiplicative and Subtractive Adaptation Process. *Vision Research* 27, 11 (1987), 1981–1996.

Rodney L Heckaman, Masato Sakurai, Mark D Fairchild, Takehiro Nakatsue, and Yoshihide Shimpuku. 2007. The Effect of Display Gamut Volume on Image Preference. In *Color and Imaging Conference*, Vol. 15. Society of Imaging Science and Technology, 201–206.

Guanhua Hou, Qi Dong, and Huiwen Wang. 2025. The Effect of Dynamic Effects and Color Transparency of AR-HUD Navigation Graphics on Driving Behavior Regarding Inattentional Blindness. *International Journal of Human–Computer Interaction* 41, 12 (2025), 7581–7592.

Robert William Gainer Hunt. 1952. Light and Dark Adaptation and the Perception of Color. *Journal of the Optical Society of America* 42, 3 (1952), 190–199.

ITU-R. 2015a. *ITU-R BT.2020: Parameter Values for Ultra-High Definition Television Systems for Production and International Programme Exchange.* ITU-R Recommendation BT.2020-2. International Telecommunication Union, Radiocommunication Sector. https://www.itu.int/rec/R-REC-BT.2020/

ITU-R. 2015b. *ITU-R BT.709: Parameter Values for The HDTV Standards for Production and International Programme Exchange.* ITU-R Recommendation BT.709-6. International Telecommunication Union, Radiocommunication Sector. https://www.itu.int/rec/R-REC-BT.709/

ITU-R. 2017. *ITU-R BT.2407: Colour Gamut Conversion from Recommendation ITU-R BT.2020 to Recommendation ITU-R BT.709.* ITU-R Report BT.2407-0. International Telecommunication Union, Radiocommunication Sector. https://www.itu.int/pub/R-REP-BT.2407/

ITU-R. 2021. *ITU-R BT.2446: Methods for Conversion of High Dynamic Range Content to Standard Dynamic Range Content and Vice-Versa.* ITU-R Report BT.2446-1. International Telecommunication Union, Radiocommunication Sector. https://www.itu.int/pub/R-REP-BT.2446/

ITU-R. 2025. *ITU-R BT.2100: Image Parameter Values for High Dynamic Range Television for Use in Production and International Programme Exchange.* ITU-R Recommendation BT.2100-3. International Telecommunication Union, Radiocommunication Sector. https://www.itu.int/rec/R-REC-BT.2100/

Fu Jiang, Mark D Fairchild, and Kenichiro Masaoka. 2018. Estimation of HDR WCG Display Color Gamut Volume. In *Color and Imaging Conference*, Vol. 26. Society for Imaging Science and Technology, 344–349.

Fu Jiang, Mark D Fairchild, and Kenichiro Masaoka. 2020. Effect of Peak Luminance on Perceptual Color Gamut Volume. In *Color and Imaging Conference*, Vol. 28. Society for Imaging Science and Technology, 13–18.

Dominik Keller, Thomas Goebel, Valentin Siebenkees, Julius Prenzel, and Alexander Raake. 2024. AVT-VQDB-UHD-2-HDR: An Open 8K HDR Source Dataset for Video Quality Research. In *Proceedings of International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 186–192.

Dongyeon Kim, Maliha Ashraf, Alexandre Chapiro, and Rafał K Mantiuk. 2025. Supra-Threshold Contrast Perception in Augmented Reality. In *Proceedings of the SIGGRAPH Asia 2025 Conference Papers*. 1–11.

Ron Kimmel, Doron Shaked, Michael Elad, and Irwin Sobel. 2005. Space-Dependent Color Gamut Mapping: A Variational Approach. *IEEE Transactions on Image Processing* 14, 6 (2005), 796–803.

Jiangtao Kuang, Garrett M Johnson, and Mark D Fairchild. 2007. iCAM06: A Refined Image Appearance Model for HDR Image Rendering. *Journal of Visual Communication and Image Representation* 18, 5 (2007), 406–414.

Timo Kunkel, Greg Ward, Bongsun Lee, and Scott Daly. 2013. HDR and Wide Gamut Appearance-based Color Encoding and Its Quantification. In *2013 Picture Coding Symposium (PCS)*. IEEE, 357–360.

Mark A Livingston, Jane H Barrow, and Ciara M Sibley. 2009. Quantification of Contrast Sensitivity and Color Perception using Head-Worn Augmented Reality Displays. In *2009 IEEE Virtual Reality Conference*. IEEE, 115–122.

Lindsay MacDonald, Ján Morovic, and Kaida Xiao. 2002. A Topographic Gamut Compression Algorithm. *Journal of Imaging Science and Technology* 46, 3 (2002), 228–236.

Radoslaw Mantiuk, Rafal Mantiuk, Anna Tomaszewska, and Wolfgang Heidrich. 2009. Color correction for tone mapping. In *Computer graphics forum*, Vol. 28. Wiley Online Library, 193–202.

Rafal K Mantiuk, Param Hanji, Maliha Ashraf, Yuta Asano, and Alexandre Chapiro. 2024. ColorVideoVDP: A Visual Difference Predictor for Image, Video and Display Distortions. *ACM Transactions on Graphics (TOG)* 43, 4 (2024), 1–20.

Kenichiro Masaoka, Roy S Berns, Mark D Fairchild, and Farhad Moghareh Abed. 2013. Number of Discernible Object Colors is A Conundrum. *Journal of the optical society of America A* 30, 2 (2013), 264–277.

Kenichiro Masaoka, Fu Jiang, Mark D Fairchild, and Rodney L Heckaman. 2018. 2D Representation of Display Color Gamut. In *SID Symposium Digest of Technical Papers*, Vol. 49. Wiley Online Library, 1048–1051.

Kenichiro Masaoka, Fu Jiang, Mark D Fairchild, and Rodney L Heckaman. 2020. Analysis of Color Volume of Multi-Chromatic Displays using Gamut Rings. *Journal of the Society for Information Display* 28, 3 (2020), 273–286.

Kenichiro Masaoka, Euan Smith, Karl Lang, Brian Berkeley, Johan Bergquist, and John Penczek. 2025. Visualization of Reproducible Object Colors in Standard Color Spaces

using The Gamut Ring Intersection. *Journal of the Society for Information Display* 33, 4 (2025), 231–245.

Fabio Metelli. 1974. The perception of transparency. *Scientific American* 230, 4 (1974), 90–99.

Aliaksei Mikhailiuk, Clifford Wilmot, Maria Perez-Ortiz, Dingcheng Yue, and Rafał K Mantiuk. 2021. Active Sampling for Pairwise Comparisons via Approximate Message Passing and Information Gain Maximization. In *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2559–2566.

Ethan D Montag and Mark D Fairchild. 1997. Psychophysical Evaluation of Gamut Mapping Techniques using Simple Rendered Images and Artificial Gamut Boundaries. *IEEE Transactions on Image Processing* 6, 7 (1997), 977–989.

Nathan Moroney, Mark Fairchild, Robert Hunt, and Changjun Li. 2002. The CIECAM02 Color Appearance Model. In *IS&T/SID 10th Color Imaging Conference, Scottsdale, 2002*. 23–27.

Ján Morovič. 2008. *Color Gamut Mapping*. John Wiley & Sons.

Ján Morovič and M Ronnier Luo. 2001. The Fundamentals of Gamut Mapping: A Survey. *Journal of Imaging Science and Technology* 45, 3 (2001), 283–290.

Michael J Murdoch. 2020. Brightness Matching in Optical See-Through Augmented Reality. *Journal of the Optical Society of America A* 37, 12 (2020), 1927–1936.

Michael J Murdoch. 2025. Transparency and Scission in Augmented Reality. *Electronic Imaging* 37 (2025), 1–3.

Maria Perez-Ortiz, Aliaksei Mikhailiuk, Emin Zerman, Vedad Hulusic, Giuseppe Valenzise, and Rafał K Mantiuk. 2019. From Pairwise Comparisons and Rating to A Unified Quality Scale. *IEEE Transactions on Image Processing* 29 (2019), 1139–1151.

Jens Preiss, Mark D Fairchild, James A Ferwerda, and Philipp Urban. 2014. Gamut Mapping in A High-Dynamic-Range Color Space. In *Color Imaging XIX: Displaying, Processing, Hardcopy, and Applications*, Vol. 9015. SPIE, 79–85.

D McL Purdy. 1931. Spectral Hue as a Function of Intensity. *The American Journal of Psychology* (1931).

Maxime Rousselot, Olivier Le Meur, Rémi Cozot, and Xavier Ducloux. 2019. Quality Assessment of HDR/WCG Images Using HDR Uniform Color Spaces. *Journal of Imaging* 5, 1 (2019), 18.

Masato Sakurai, Rodney L Heckaman, Stacey E Casella, Mark D Fairchild, Takehiro Nakatsue, and Yoshihide Shimpuku. 2008a. Effects of Display Properties on Perceived Color-Gamut Volume and Preference. *Journal of the Society for Information Display* 16, 12 (2008), 1203–1211.

Masato Sakurai, Takehiro Nakatsue, Yoshihide Shimpuku, Rodney L Heckaman, Stacey E Casella, and Mark D Fairchild. 2008b. Effect of Color Gamut Volume in Display on Image Preference. In *Sid symposium digest of technical papers*, Vol. 39. Wiley Online Library, 795–798.

Christophe Schlick. 1995. Quantization Techniques for Visualization of High Dynamic Range Pictures. In *Photorealistic Rendering Techniques*. Springer, 7–20.

Zaixi Shang, Joshua P Ebenezer, Abhinau K Venkataramanan, Yongjun Wu, Hai Wei, Sriram Sethuraman, and Alan C Bovik. 2023. A Study of Subjective and Objective Quality Assessment of HDR Videos. *IEEE Transactions on Image Processing* 33 (2023), 42–57.

Elena Sikudova, Tania Pouli, Alessandro Artusi, Ahmet Oguz Akyuz, Francesco Banterle, Zeynep Miray Mazlumoglu, and Erik Reinhard. 2015. A Gamut-Mapping Framework for Color-Accurate Reproduction of HDR Images. *IEEE Computer Graphics and Applications* 36, 4 (2015), 78–90.

Euan Smith, Rodney L Heckaman, Karl Lang, John Penczek, and Johan Bergquist. 2020. Measuring the Color Capability of Modern Display Systems. *Journal of the Society for Information Display* 28, 6 (2020), 548–556.

Li Song, Yankai Liu, Xiaokang Yang, Guangtao Zhai, Rong Xie, and Wenjun Zhang. 2016. The SJTU HDR Video Sequence Dataset. In *Proceedings of International Conference on Quality of Multimedia Experience (QoMEX)*. 100.

Taimoor Tariq, Nathan Matsuda, Eric Penner, Jerry Jia, Douglas Lanman, Ajit Ninan, and Alexandre Chapiro. 2023. Perceptually Adaptive Real-Time Tone Mapping. In *SIGGRAPH Asia 2023 Conference Papers*. 1–10.

Minqi Wang, Jian Ding, Dennis M Levi, and Emily A Cooper. 2023. The Effect of Interocular Contrast Differences on the Appearance of Augmented Reality Imagery. *ACM Transactions on Applied Perception* 21, 1 (2023), 1–23.

Senfar Wen. 2006. Display Gamut Comparison with Number of Discernible Colors. *Journal of Electronic Imaging* 15, 4 (2006), 043001–043001.

Jiaying Wu, Lu Zhang, Serhan Isikman, and Cheng Chen. 2019. Enhanced Viewing Experience Considering Chromatic Adaptation. In *SID Symposium Digest of Technical Papers*, Vol. 50. Wiley Online Library, 857–860.

Lihao Xu, Baiyue Zhao, and Ming Ronnier Luo. 2018a. Color Gamut Mapping between Small and Large Colour Gamuts: part II. Gamut Extension. *Optics Express* 26, 13 (2018), 17335–17349.

Lihao Xu, Baiyue Zhao, and Ming Ronnier Luo. 2018b. Colour Gamut Mapping between Small and Large Colour Gamuts: Part I. Gamut Compression. *Optics Express* 26, 9 (2018), 11481–11495.

Putra AR Yamin, Jaehyun Park, Hyun K Kim, and Muhammad Hussain. 2024. Effects of Button Colour and Background on Augmented Reality Interfaces. *Behaviour & Information Technology* 43, 4 (2024), 663–676.

Syed Waqas Zamir, Javier Vazquez-Corral, Marcelo Bertalmı, et al. 2014. Gamut Mapping in Cinematography Through Perceptually-Based Contrast Modification. *IEEE Journal of Selected Topics in Signal Processing* 8, 3 (2014), 490–503.

Syed Waqas Zamir, Javier Vazquez-Corral, and Marcelo Bertalmio. 2019. Vision Models for Wide Color Gamut Imaging in Cinema. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43, 5 (2019), 1777–1790.

Lili Zhang. 2022. *Lightness, Brightness, and Transparency in Optical See-Through Augmented Reality*. Rochester Institute of Technology.

Lili Zhang, Romain Bachy, and James Hillis. 2023. *Egocentric Luminance Statistics*. Poster. Reality Labs, Meta Platforms Inc. https://sites.google.com/view/lili-zhang-us/neti-poster

Lili Zhang, Michael J Murdoch, and Romain Bachy. 2021a. Color Appearance Shift in Augmented Reality Metameric Matching. *Journal of the Optical Society of America A* 38, 5 (2021), 701–710.

Yunjin Zhang, Rui Wang, Yifan Peng, Wei Hua, and Hujun Bao. 2021b. Color Contrast Enhanced Rendering for Optical See-Through Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (2021), 4490–4502.

This appendix provides additional implementation details to support the reproducibility of our WCG HDR video dataset, display-simulation pipeline, color-enhancement algorithms, perceptual model derivation, and user study designs.

A WCG HDR Video Dataset

This section provides additional details on the WCG HDR video dataset introduced in Section 3.1, including preprocessing steps, scene characteristics, output video metadata, and the computation of color-volume coverage.

A.1 Video Preprocessing

The source videos are available at different resolutions: *Fruits* from Pexels[7] is 4K UHD (3840 × 2160); *Sol Levante* from Netflix Open Content[8] is 1080p (1920 × 1080); *Cosmos Laundromat*, also from Netflix Open Content, is DCI 2K Scope (2048 × 858); *Showgirl* from HdM-HDR-2014 [Froehlich et al. 2014] is 1080p (1920 × 1080); *Forest*, *Drawing*, and *Knitting* from AVT-VQDB-UHD-2-HDR [Keller et al. 2024] are DCI-4K (4096×2160); and the newly created videos (*Bunny*, *Surface*, *Werewolf*, and *Particles*) are rendered and exported at 1080p (1920 × 1080). Source videos not already at 1080p are standardized to 1080p through spatial resampling with the Spline36 filter implemented in `zimg`[9], followed by center cropping when necessary. From each source video, we sample one or more representative clips (3–7 seconds each; see below for details) without scene cuts, ensuring that each clip contains a single continuous shot. This avoids abrupt content changes during stimulus presentation, which could confound observers' color judgments. The original temporal resolution of each clip is preserved (24–60 frames per second, FPS). These processing choices are aligned with the user-study setups and requirements described in Sections 3.3 and 4. Because the source videos collected from existing datasets are not available in stereoscopic format, identical content is presented to the left and right eyes in the stereoscopic OST-AR studies (Section 4).

A.2 Scene Characteristics

Descriptions of the 13 video clips in our dataset are provided below. These clips span real-world footage, cinematic scenes, animated films, and synthetic graphics, with varied color distributions, dynamic ranges, and scene dynamics.

- *Showgirl* (5 seconds, 60 FPS): a cinematic dressing-room shot of a female performer applying makeup, with warm skin tones, saturated reds and cyans, localized highlights, and slow camera/subject motion.
- *Forest* (5 seconds, 60 FPS): an outdoor nature scene of a person in a denim jacket and jeans walking along a forest path, dominated by foliage greens with broad low-to-high luminance variation.
- *Fruits* (6 seconds, 24 FPS): a slow camera pan over an outdoor picnic scene with fruits and food items under bright daylight illumination, featuring high-chroma reds, oranges, and yellows at high luminance levels.
- *Drawing* (5 seconds, 60 FPS): an over-the-shoulder shot of a person sketching on white paper, combining high local contrast around the sketch, dark blue clothing, saturated red socks, and saturated green foliage in the background.
- *Knitting* (5 seconds, 60 FPS): a close-up shot of an elderly woman's hands knitting yellow-orange fabric while wearing a dark blue sweater, emphasizing warm hues and skin tones.
- *Sol Levante 1* (5.7 seconds, 24 FPS): an animated action shot with saturated blue, cyan, and purple energy effects, strong local contrast, rapid subject motion, and broad low-to-high luminance coverage.
- *Sol Levante 2* (3.4 seconds, 24 FPS): an animated volcanic scene with saturated magenta, purple, blue, and orange-red illumination, spanning a wide dynamic range.
- *Sol Levante 3* (3.3 seconds, 24 FPS): an animated aerial panning shot of a bright blue sea, green landscape, and a saturated magenta-purple foreground character at high luminance levels, with broad camera and subject motion.
- *Bunny* (4.2 seconds, 30 FPS): a stylized animation of a bunny running through a grassy field with sky, clouds, and a horizon line in the background, featuring saturated greens, yellow-oranges, and blue-purples, smooth color gradients, and large subject motion.
- *Surface* (5 seconds, 30 FPS): an abstract synthetic scene of multicolored surfaces waving and morphing over time, providing smooth color gradients and broad coverage of highly saturated colors across many hues.
- *Werewolf* (7 seconds, 30 FPS): an animated scene transitioning from dark saturated blues and reds to a moderately bright scene with localized saturated yellow and red highlights, accompanied by large character and camera motion.
- *Particles* (5 seconds, 30 FPS): an abstract particle animation with dense green, magenta, red, and yellow distributions against a dark background, providing saturated color samples and dynamic structure.
- *Cosmos Laundromat* (4 seconds, 24 FPS): an animated fantasy scene with saturated red, magenta, purple, cyan, and green elements, combining stylized rendering with high chroma diversity.

A.3 Perceptual Color Volume

To quantify the color coverage of our dataset, we follow the perceptual color volume (PCV) formulation introduced by Dolby [2023]. PCV operates in ICtCp color space, which is designed to be approximately linear in just-noticeable-difference (JND) units and perceptually uniform for HDR/WCG imagery. ICtCp is defined with respect to the perceptual quantizer (PQ) electro-optical transfer function (EOTF), which maps normalized signal values in $[0, 1]$ to absolute luminance in $[, 10{,}000]$ nits. The component ranges of ICtCp are defined such that the BT.2020 gamut at 10,000-nit peak luminance and 0-nit black level fits within $I \in [0, 1]$ and $C_t, C_p \in [-0.5, 0.5]$. Following the PCV formulation, we convert ICtCp coordinates into JND units by scaling each axis as:

$$\tilde{I} = 720\, I, \quad \tilde{C}_t = 360\, C_t, \quad \tilde{C}_p = 720\, C_p, \tag{19}$$

[7] https://www.pexels.com/
[8] https://opencontent.netflix.com/
[9] https://github.com/sekrit-twc/zimg

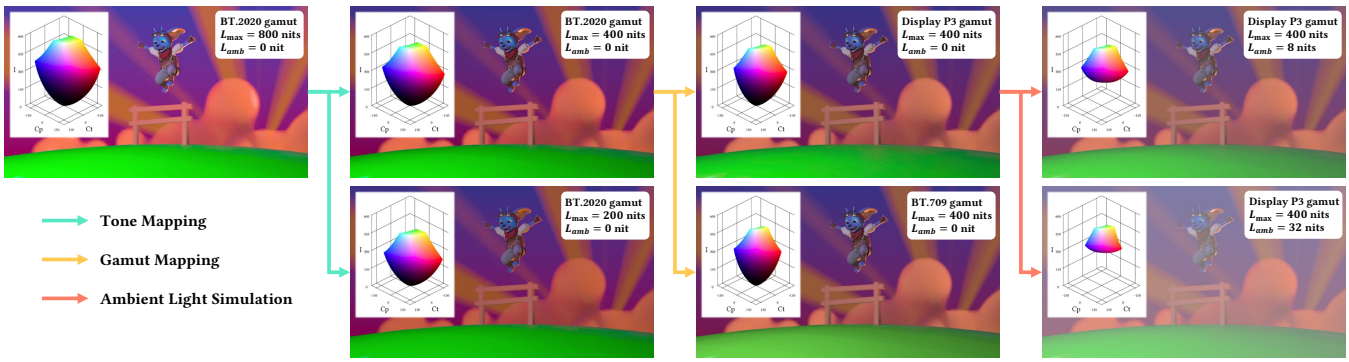

Fig. 13. *Display-simulation pipeline.* Our pipeline combines tone mapping, gamut mapping, and ambient-light modeling to emulate target viewing conditions with varying color gamut, dynamic range, and ambient luminance on a WCG/HDR reference display, avoiding the need to instantiate multiple physical setups. Insets visualize the corresponding effective color gamut in ICtCp color space after accounting for all three factors.

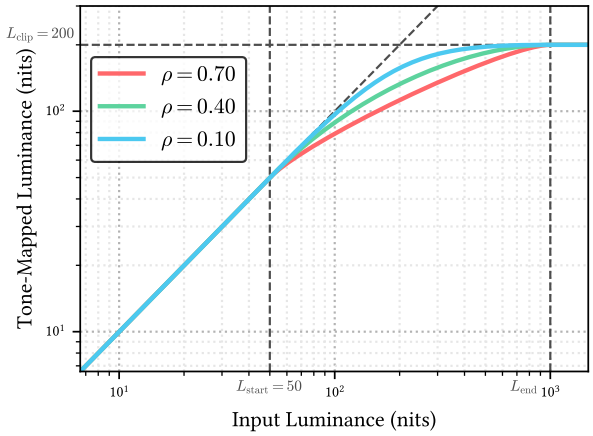

Fig. 14. *Bézier-based tone-mapping curve [Chen et al. 2023].* The curve follows a Bézier formulation and is parameterized by $\theta_{TM} = (\rho, L_{start}, L_{end}, L_{clip})$, where ρ controls the curve slope, and L_{start}, L_{end}, and L_{clip} define the starting, ending, and clipping luminance levels, respectively. Input luminance values below L_{start} are preserved, values between L_{start} and L_{end} are smoothly compressed, and values above L_{end} are clipped to L_{clip}. A larger ρ yields a smoother, more log-linear compression curve that better preserves highlight contrast, whereas a smaller ρ yields a curve closer to clipping. The figure shows example curves for three values of ρ.

This JND-scaled ICtCp space discretizes color volume into unit voxels, and the number of occupied voxels is reported in millions of distinguishable colors (MDCs).

The reference color volume used in Figure 2 is the BT.2020 gamut at 1,000-nit peak luminance and 0.001-nit black level, shown as the dotted outer shell. We first enumerate all integer-bounded voxels in JND-scaled ICtCp space that fall inside this reference color volume. We then aggregate all pixels from all 13 video clips, convert each pixel from BT.2020 RGB to ICtCp, apply the JND-unit scaling, and assign the resulting transformed pixel to its corresponding voxel. A voxel is considered covered if at least one dataset pixel falls inside it. The spatial and temporal subsampling described in Figure 2 is used only for visualization and not in this computation. The color-volume coverage of our dataset relative to the reference color volume is therefore computed as:

$$\text{Coverage} = \frac{|V_{\text{dataset}} \cap V_{\text{BT.2020}}|}{|V_{\text{BT.2020}}|}, \tag{20}$$

where $V_{\text{BT.2020}}$ denotes the set of voxels inside the reference BT.2020 color volume, and V_{dataset} denotes the set of voxels occupied by pixels from our dataset. Based on this procedure, our dataset covers 50% of the reference color volume.

B Tone Mapping

This section details the tone-mapping operator introduced in Section 3.2.1, including its four processing stages, tone-curve parameterization (visualized in Figure 14), and key parameter selection.

Given the target display with black level L_{min} and peak luminance L_{max}, the tone-mapping operator proceeds as follows:

(1) Extract the input tone as the channel-wise maximum of I_{in}. While tone mapping can operate on different luminance proxies, recent work shows that the maximum RGB channel helps avoid out-of-gamut clipping and associated color distortions [Burke et al. 2020; Chen et al. 2025].

$$T_{\text{in}} = \max(I_{\text{in}}^R, I_{\text{in}}^G, I_{\text{in}}^B), \qquad \text{[Tone Extraction]} \tag{21}$$

(2) Apply the tone curve proposed by Chen et al. [2023], $\mathcal{T}(\cdot\,; \theta_{\text{TM}})$, to the input tone T_{in}. The curve follows a Bézier formulation and is parameterized by $\theta_{\text{TM}} = (\rho, L_{\text{start}}, L_{\text{end}}, L_{\text{clip}})$, where ρ controls the curve slope, and L_{start}, L_{end}, and $L_{\text{clip}} = L_{\text{max}}$ define the starting, ending, and clipping luminance levels, respectively (see Figure 14).

$$T_{\text{out}} = \mathcal{T}(T_{\text{in}}; \theta_{\text{TM}}), \qquad \text{[Tone Mapping]} \tag{22}$$

(3) Perform Schlick color correction [1995] to preserve hue by maintaining relative RGB channel ratios.

$$\tilde{I}_{\text{TM}} = (T_{\text{out}}/T_{\text{in}})\, I_{\text{in}}, \qquad \text{[Color Correction]} \tag{23}$$

(4) Simulate the target black level L_{min} through affine luminance rescaling, following Mantiuk et al. [2024].

$$I_{\text{TM}} = L_{\text{min}} + (1 - L_{\text{min}}/L_{\text{max}})\, \tilde{I}_{\text{TM}}, \quad \text{[Black-Level Simulation]} \tag{24}$$

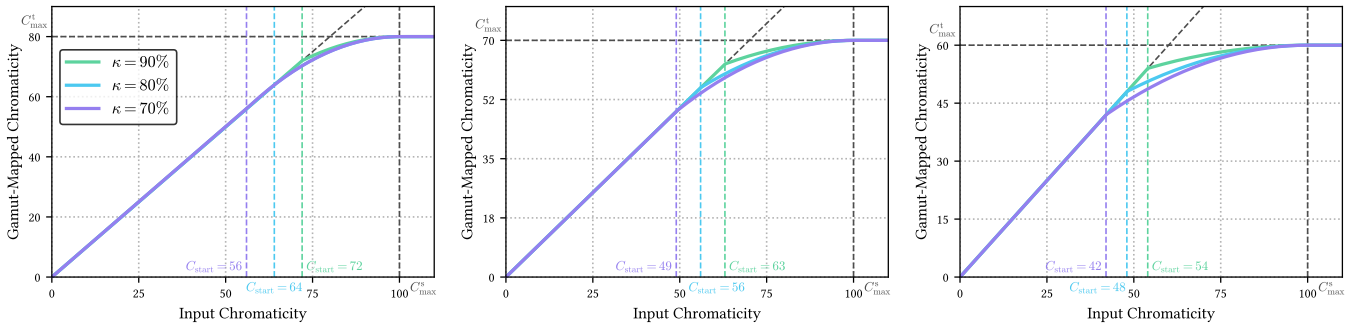

Fig. 15. *Knee functions for smooth gamut mapping [MacDonald et al. 2002].* The function is parameterized by $\theta_{\mathrm{GM}} = (\kappa, C^{\mathrm{s}}_{\mathrm{max}}, C^{\mathrm{t}}_{\mathrm{max}})$, where $C^{\mathrm{s}}_{\mathrm{max}}$ and $C^{\mathrm{t}}_{\mathrm{max}}$ denote the maximum input and output values, respectively, and $\kappa \in [0, 1]$ controls the knee position $C_{\mathrm{start}} = \kappa\, C^{\mathrm{t}}_{\mathrm{max}}$. The function maps an input scalar value C_{in} to an output value C_{out}: values below C_{start} are preserved, values between C_{start} and $C^{\mathrm{s}}_{\mathrm{max}}$ are smoothly compressed toward $C^{\mathrm{t}}_{\mathrm{max}}$, and values above $C^{\mathrm{s}}_{\mathrm{max}}$ are clipped to $C^{\mathrm{t}}_{\mathrm{max}}$. The three panels illustrate different target maxima $C^{\mathrm{t}}_{\mathrm{max}}$, and the curves within each panel show different knee positions controlled by κ.

Following Tariq et al. [2023], we perform scene-adaptive tone-curve adjustment by optimizing the starting luminance L_{start} of the nonlinear compression region to minimize contrast distortions. To improve temporal stability, we optimize L_{start} only at keyframes sampled at 2 FPS and interpolate the optimized values in the log-luminance domain, yielding temporally smooth tone-mapping parameters for intermediate frames. This simple temporal smoothing strategy produced no visible temporal artifacts in our processed videos while maintaining high computational efficiency. For videos containing more complex and dynamic content, more sophisticated temporally coherent tone-mapping methods [Aydin et al. 2014; Boitard et al. 2012] may be required. Following Chen et al. [2025], we compute the slope parameter ρ using a heuristic based on the input tone distribution:

$$\rho = \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} \mathbb{1}\left(T_{\mathrm{in}}(\mathbf{x}) > L_{\mathrm{start}}\right), \tag{25}$$

where Ω denotes the set of pixels in the input frame, $T_{\mathrm{in}}(\mathbf{x})$ denotes the input tone at pixel $\mathbf{x}$, and $\mathbb{1}(\cdot)$ denotes the indicator function. Intuitively, ρ measures the fraction of pixels whose input tone exceeds the optimized starting luminance L_{start}. A larger value indicates that highlights occupy a larger portion of the frame, yielding a smoother, more log-linear compression curve that better preserves highlight contrast. Conversely, a smaller value indicates sparse highlights and yields a curve closer to clipping.

C Gamut Mapping

This section describes the knee-function formulation used by the gamut-mapping algorithms introduced in Section 3.2.2.

We adopt the knee function $\mathcal{S}(\cdot\,; \theta_{\mathrm{GM}})$ introduced by MacDonald et al. [2002]. The function is parameterized by $\theta_{\mathrm{GM}} = (\kappa, C^{\mathrm{s}}_{\mathrm{max}}, C^{\mathrm{t}}_{\mathrm{max}})$, where $C^{\mathrm{s}}_{\mathrm{max}}$ and $C^{\mathrm{t}}_{\mathrm{max}}$ denote the maximum input and output values, respectively, and $\kappa \in [0, 1]$ controls the knee position $C_{\mathrm{start}} = \kappa\, C^{\mathrm{t}}_{\mathrm{max}}$ (refer to Figure 15). The function maps an input scalar value C_{in} to an output value C_{out}, preserving values below C_{start} and smoothly compressing values above C_{start} toward $C^{\mathrm{t}}_{\mathrm{max}}$. Let

$$\zeta = \frac{1}{2} \frac{C_{\mathrm{in}} - C_{\mathrm{start}}}{C^{\mathrm{s}}_{\mathrm{max}} - C_{\mathrm{start}}}, \tag{26}$$

The knee function $C_{\mathrm{out}} = \mathcal{S}(C_{\mathrm{in}}; \theta_{\mathrm{GM}})$ is defined as:

$$C_{\mathrm{out}} = \begin{cases} C_{\mathrm{in}}, & C_{\mathrm{in}} \le C_{\mathrm{start}}, \\ C_{\mathrm{start}} + 4(C^{\mathrm{t}}_{\mathrm{max}} - C_{\mathrm{start}})(\zeta - \zeta^2), & C_{\mathrm{start}} < C_{\mathrm{in}} < C^{\mathrm{s}}_{\mathrm{max}}, \\ C^{\mathrm{t}}_{\mathrm{max}}, & C_{\mathrm{in}} \ge C^{\mathrm{s}}_{\mathrm{max}}. \end{cases} \tag{27}$$

Thus, C_{start} controls the extent of the unchanged region: values below C_{start} are preserved, values between C_{start} and $C^{\mathrm{s}}_{\mathrm{max}}$ are smoothly compressed, and values above $C^{\mathrm{s}}_{\mathrm{max}}$ are clipped to $C^{\mathrm{t}}_{\mathrm{max}}$.

In our implementation of **RGB-SC**, we define $\kappa = 0.8$, $C^{\mathrm{s}}_{\mathrm{max}} = 1.25\, L_{\mathrm{max}}$, and $C^{\mathrm{t}}_{\mathrm{max}} = L_{\mathrm{max}}$. For **ICtCp-CHL**, we also set $\kappa = 0.8$, while $C^{\mathrm{s}}_{\mathrm{max}}(L, h)$ and $C^{\mathrm{t}}_{\mathrm{max}}(L, h)$ denote the hue- and luminance-dependent maximum chroma values of the source (BT.2020) and target (G_{target}) gamuts, respectively. These values are obtained from the precomputed 2D lookup tables (LUTs) described in Section 3.2.2.

D Direct-View User Study

This section provides additional details on the stimuli selection process for the direct-view user study (Section 3.3) and reports the full ANOVA statistics (Table 5). We first summarize the display types considered in our study (Appendix D.1), then analyze ambient-luminance distributions in indoor and outdoor viewing environments (Appendix D.2). Finally, we describe how peak-luminance and ambient-luminance levels were sampled to define the study stimuli (Appendix D.3).

D.1 Display Types

We considered display types commonly used in practical applications, including laptops, phones/tablets, desktop monitors, televisions, cinema displays, VR headsets, and OST-AR glasses. Table 1 summarizes representative models in each category and their key color-related characteristics.

D.2 Ambient-Luminance Distributions

We used the indoor and outdoor ambient-luminance distributions reported by Zhang et al. [2023] to estimate representative ambient-luminance quantiles. Rather than using incident ambient luminance

Table 1. *Representative models of common display types and their color-related characteristics.* Peak luminance is reported as sustained full-screen luminance rather than transient or local peak luminance. Contrast is defined as the ratio between peak luminance and minimum luminance, i.e., black level. For all models except OST-AR glasses, contrast is reported under ideal near-dark viewing conditions. For OST-AR glasses, contrast is typically close to or below 10:1 [Chapiro et al. 2024], because these devices are driven as a function of ambient luminance rather than continuously at peak luminance to improve power efficiency. Unavailable exact values are denoted as N/A.

Category	Model	Peak Luminance (nits)	Color Gamut	Contrast	Technology
Laptop	MacBook Pro 2024	1,000	Display P3	1,000,000:1	Mini-LED
Phone & Tablet	iPhone 17 Pro	1,000	Display P3	2,000,000:1	OLED
	iPad Pro 2022	1,000	Display P3	1,000,000:1	LCD (IPS)
Desktop Monitor	Dell U2723QE	400	Custom (98% of DCI-P3)	2,000:1	LCD (IPS)
	ASUS PG27AQDM	400	Custom (99% of DCI-P3)	1,500,000:1	W-OLED
	Apple Pro Display XDR	1,000	Display P3	1,000,000:1	LCD (IPS)
	Apple Studio Display	600	Display P3	1,200:1	LCD (IPS)
Television	TCL S325	200	BT.709	3,500:1	LCD (VA)
	Samsung S95C	250	Custom (99% of DCI-P3)	1,000,000:1	QD-OLED
Cinema Display	Dolby Cinema	106	Custom (122% of DCI-P3)	7,500:1	Laser Projectors
VR Headset	Meta Quest 3	100	Custom (99% of BT.709)	64:1	LCD
	Meta Quest Pro	100	Custom (99% of BT.709)	64:1	Mini-LED
	Apple Vision Pro	100	Custom (92% of DCI-P3)	100:1	Micro-OLED
OST-AR Glasses	Meta Orion AI Glasses	N/A	N/A	N/A	Micro-LED
	Meta Ray-Ban Display	5,000	N/A	N/A	LCoS

Table 2. *Estimated ambient-luminance quantiles for indoor and outdoor viewing environments.* Incident ambient luminance is estimated from the distributions reported by Zhang et al. [2023] using log-domain interpolation within each luminance bin. For conventional flat-panel displays, eye-reaching ambient luminance is estimated using 1–5% panel reflectance; for OST-AR displays, it is estimated using 20–90% optical transmittance.

Quantile	Indoor Ambient Luminance (nits)					Outdoor Ambient Luminance (nits)				
	Incident	Flat 1%	Flat 5%	AR 20%	AR 90%	Incident	Flat 1%	Flat 5%	AR 20%	AR 90%
10%	1.67	0.017	0.084	0.33	1.51	2.15	0.022	0.11	0.43	1.94
20%	2.72	0.027	0.14	0.54	2.45	13.6	0.14	0.68	2.71	12.2
30%	4.23	0.042	0.21	0.85	3.80	56.2	0.56	2.81	11.2	50.6
40%	5.83	0.058	0.29	1.17	5.25	158.5	1.59	7.93	31.7	143
50%	8.41	0.084	0.42	1.68	7.57	299	2.99	14.9	59.7	269
60%	12.3	0.12	0.61	2.46	11.1	531	5.31	26.5	106	478
70%	18.2	0.18	0.91	3.63	16.3	1,000	10.0	50.0	200	900
80%	28.1	0.28	1.40	5.62	25.3	1,995	20.0	99.8	399	1,796
90%	50.5	0.51	2.53	10.1	45.5	4,786	47.9	239	957	4,308
100%	251	2.51	12.6	50.2	226	10,000	100	500	2,000	9,000

Table 3. *Estimated additive contrast for flat-panel displays in indoor and outdoor viewing environments.* Eye-reaching ambient luminance is estimated from the incident ambient luminance in Table 2 by assuming 3% panel reflectance. Additive contrast is computed using Equation (8) with $L_{\text{min}} = 0$ nit for representative peak-luminance settings: 200 nits (SDR) and 1,000 nits (HDR). Cell colors indicate additive-contrast ranges: green > 1,000:1, yellow 100:1–1,000:1, orange 10:1–100:1, and red < 10:1.

Environment	Peak Luminance	Ambient-Luminance Quantile									
		10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Indoor	200 nits (SDR)	3,993:1	2,452:1	1,577:1	1,145:1	794:1	543:1	367:1	238:1	133:1	28:1
	1,000 nits (HDR)	19,961:1	12,256:1	7,881:1	5,719:1	3,965:1	2,711:1	1,833:1	1,187:1	661:1	134:1
Outdoor	200 nits (SDR)	3,102:1	491:1	120:1	43:1	23:1	14:1	8:1	4:1	2:1	2:1
	1,000 nits (HDR)	15,505:1	2,452:1	594:1	211:1	112:1	64:1	34:1	18:1	8:1	4:1

Table 4. *Additive contrast of sampled direct-view user-study conditions.* Additive contrast is computed using Equation (8) with $L_{\text{min}} = 0.001$ nit for two peak-luminance settings, 200 nits (SDR) and 900 nits (HDR), across the sampled eye-reaching ambient luminance levels. Cell colors follow the convention in Table 3: green > 1,000:1, yellow 100:1–1,000:1, orange 10:1–100:1, and red < 10:1. The gray cell denotes the excluded SDR condition at $L_{\text{amb}} = 100$ nits.

Peak Luminance	Eye-Reaching Ambient Luminance L_{amb} (nits)					
	0	**1**	**4**	**16**	**64**	**100**
200 nits (SDR)	200,001:1	201:1	51:1	13.5:1	4.125:1	3:1
900 nits (HDR)	900,001:1	901:1	226:1	57.25:1	15.06:1	10:1

directly, we estimated the ambient light reaching the observer's eyes by accounting for display-dependent reflection or transmission: 1–5% panel reflectance for conventional flat-panel displays and 20–90% optical transmittance for OST-AR displays with transition lenses. The 1% lower bound for panel reflectance is motivated by the on-axis reflectivity of flat-panel displays with anti-reflective coatings, such as 1.65% for Apple Pro Display XDR, 1.5% for Samsung Galaxy S24 Ultra, and 1.1% for LG C3 OLED. For flat-panel displays without anti-reflective coatings, on-axis reflectivity can approach 5%, such as 4.8% for iPhone 15 Pro Max and 4.8% for Google Pixel 8 Pro. A 20–90% optical-transmittance range is assumed for OST-AR displays with transition lenses. Table 2 summarizes the quantiles for indoor and outdoor viewing environments and for both display classes.

D.3 Direct-View User-Study Conditions

With the ambient-luminance quantiles summarized in Table 2, we computed the corresponding additive contrast for each display class using Equation (8). For flat-panel displays commonly used under non-negligible ambient illumination, including laptops, phones/tablets, desktop monitors, and televisions, we considered two representative peak-luminance settings, 200 nits (SDR) and 1,000 nits (HDR), based upon the display characteristics summarized in Table 1. Table 3 reports the computed additive contrast for these settings. For display classes typically used under near-dark viewing conditions, such as cinema displays and VR headsets, additive contrast is expected to remain close to the native contrast reported in Table 1. In contrast, OST-AR displays often operate at much lower additive contrast (close to or below 10:1 [Chapiro et al. 2024]), because these devices are driven as a function of ambient luminance rather than continuously at peak luminance to improve power efficiency.

Based on these additive-contrast estimates and a pilot study with two expert participants, we sampled six eye-reaching ambient luminance levels for L_{amb}: 0, 1, 4, 16, 64, and 100 nits. These conditions cover more than the 90th percentile of eye-reaching ambient luminance for both flat-panel and OST-AR displays in indoor viewing environments, and approximately the 80th percentile for flat-panel displays and the 60th percentile for OST-AR displays in outdoor viewing environments (see Table 2). We used two peak-luminance settings: 200 nits (SDR) and 900 nits (HDR). The HDR peak luminance was limited to 900 nits because both the target display emission and the eye-reaching ambient component were simulated on the same EIZO display; therefore, the sum of the simulated display peak luminance L_{max} and eye-reaching ambient luminance L_{amb} must not exceed the EIZO peak luminance, 1,000 nits. The target black level L_{min} was set to the EIZO black level, 0.001 nit.

Table 5. *N-way ANOVA results for the direct-view user study.* The dependent variable is perceptual quality, measured in just-objectionable-difference (JOD) units [Perez-Ortiz et al. 2019]. Factors include scene, peak luminance, color gamut, ambient luminance, and gamut-mapping algorithm (GMA). The table reports main effects and two-way interaction effects. SS denotes sum of squares, and df denotes degrees of freedom.

Effect	**SS**	**df**	***F***	***p***
Scene	2289.8	8	555.09	< .001
Peak Luminance	4.91	1	9.52	.002
Color Gamut	92.73	2	89.92	< .001
Ambient Luminance	1999.4	3	1292.6	< .001
GMA	27.14	1	52.64	< .001
Scene × Peak Luminance	758.16	8	183.80	< .001
Scene × Color Gamut	44.38	16	5.38	< .001
Scene × Ambient Luminance	118.55	24	9.58	< .001
Scene × GMA	14.41	8	3.49	< .001
Peak Luminance × Color Gamut	0.15	2	0.14	.867
Peak Luminance × Ambient Luminance	18.71	3	12.09	< .001
Peak Luminance × GMA	0.20	1	0.38	.537
Color Gamut × Ambient Luminance	14.88	6	4.81	< .001
Color Gamut × GMA	5.39	2	5.23	.006
Ambient Luminance × GMA	1.78	3	1.15	.329

Table 4 summarizes the additive contrast for all sampled conditions. These conditions span an additive-contrast range of 4.125:1–900,001:1, covering the various displays discussed in Table 1. The 100-nit ambient condition was excluded for the 200-nit peak setting because its additive contrast is similar to that of the 64-nit condition. The 0-nit condition isolates intrinsic perceptual differences across color gamuts under near-dark viewing. Although the EIZO display cannot reproduce the full BT.2020 gamut, more than 93% of the ambient-blended BT.2020 gamut falls within the EIZO gamut under 64–100 nits of ambient luminance, enabling its simulation. This resulted in 5 SDR and 6 HDR ambient conditions for each of the EIZO, Display P3, and BT.709 gamuts, and 1 SDR and 2 HDR ambient conditions for the BT.2020 gamut, yielding 36 conditions per scene and GMA, and 648 stimuli in total (9 scenes × 2 GMAs × 36 conditions).

E Perceptual Model

This section provides additional details on the perceptual-model formulation introduced in Section 3.4 and reports the fitted model parameters (Table 6).

Motivated by the strong agreement between perceptual color volume (MDC) and perceived color quality (JOD), as indicated by consistently high linear, rank-based, and distribution-sensitive correlations (refer to Section 3.4 for details), we develop an analytical model that predicts perceived color quality from perceptual color volume. As illustrated in Figure 5, perceived color quality increases with perceptual color volume and gradually saturates as the test color volume approaches the reference color volume, suggesting a quasi power-law relationship. To model this behavior, we evaluate

Table 6. *Fitted parameters for the selected power-model formulation.*

GMA	Peak Luminance	k_1	k_2	k_3
RGB-SC	200 nits (SDR)	4.7263	0.2259	−8.4946
	900 nits (HDR)	2.5812	0.3819	−7.8089
ICtCp-CHL	200 nits (SDR)	2.8183	0.3304	−6.8539
	900 nits (HDR)	2.4101	0.3923	−7.7905

three candidate functional forms—power, logarithmic, and rational—and fit each of them using nonlinear least squares.

Although peak luminance is incorporated into the MDC computation, Figure 5 shows that the JOD–MDC relationship remains luminance-dependent. Because the two sampled peak-luminance settings are insufficient to directly fit a robust unified luminance-dependent model, we first build separate models for each peak-luminance setting and then merge the fitted parameters through log-luminance interpolation. The resulting unified formulations for the three candidate models are:

$$Q_{\text{power}}(V,L) = k_1(L)\, V^{\,k_2(L)} + k_3(L), \tag{28}$$

$$Q_{\log}(V,L) = k_1(L)\, \log(V+\epsilon) + k_2(L), \tag{29}$$

$$Q_{\text{rational}}(V,L) = \frac{k_1(L)\, V}{V + k_2(L)} + k_3(L), \tag{30}$$

with luminance-dependent parameters defined by

$$k_i(L) = (1-\alpha)k_i^{\text{SDR}} + \alpha k_i^{\text{HDR}}, \tag{31}$$

$$\alpha = \frac{\log\left(L/L_{\max}^{\text{SDR}}\right)}{\log\left(L_{\max}^{\text{HDR}}/L_{\max}^{\text{SDR}}\right)}, \quad L \in [L_{\max}^{\text{SDR}}, L_{\max}^{\text{HDR}}]. \tag{32}$$

Here, Q denotes perceived color quality in JOD units, and V denotes perceptual color volume in MDCs. $L_{\max}^{\text{SDR}}$ and $L_{\max}^{\text{HDR}}$ denote the two peak-luminance levels used for fitting: 200 nits for SDR and 900 nits for HDR. For the power and rational models, $i \in \{1,2,3\}$; for the logarithmic model, $i \in \{1,2\}$. We set $\epsilon = 10^{-6}$ in the logarithmic model for numerical stability.

We fit separate models for the two GMAs. For **ICtCp-CHL**, the power model achieved an RMSE of 0.2157 and $R^2 = 0.9762$, outperforming the logarithmic model (RMSE 0.2703, $R^2 = 0.9627$) and the rational model (RMSE 0.2311, $R^2 = 0.9727$). For **RGB-SC**, the power model also achieved the best fit, with an RMSE of 0.2374 and $R^2 = 0.9729$, outperforming the logarithmic model (RMSE 0.2798, $R^2 = 0.9624$) and slightly improving over the rational model (RMSE 0.2402, $R^2 = 0.9723$).

Model selection was performed using the Akaike Information Criterion (AIC), a metric used to evaluate and compare statistical models by balancing goodness of fit against model complexity. For **RGB-SC**, the power, logarithmic, and rational models achieved AIC values of −85.77, −78.62, and −84.98, respectively. For **ICtCp-CHL**, the corresponding AIC values were −92.29, −80.96, and −87.61. Although the rational model achieved comparable RMSE and R^2, its higher AIC suggests mild over-parameterization without a commensurate gain in explanatory power. The logarithmic model produced consistently higher error, indicating insufficient flexibility to capture the observed saturation behavior. Based on these results, we adopt the power model throughout this work.

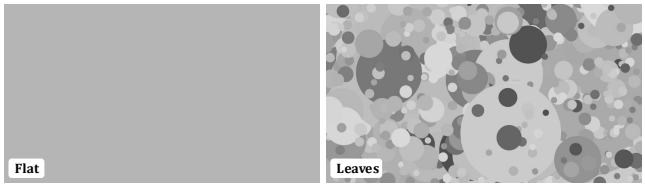

Fig. 16. *Backgrounds used in the OST-AR color-quality study (Section 4.1).* We used two backgrounds: *flat*, a uniform D65 white background, and *leaves*, a high-contrast leaves pattern following Chapiro et al. [2024]. Both backgrounds were presented with mean luminance calibrated such that, after beam-splitter and ND-filter transmission, the light reaching the observer produced the specified ambient luminance.

Table 6 reports the fitted parameters of the power model for each GMA. These parameters define the SDR and HDR endpoints used in the log-luminance interpolation in Equations (31) and (32). The exponent parameter k_2 is below 1 in all cases, consistent with the observed saturating relationship between perceptual color volume and perceived color quality: increasing perceptual color volume yields progressively smaller quality gains as the test color volume approaches the reference color volume.

F Color Enhancement

This section provides additional implementation details and parameter settings for the three color-enhancement algorithms introduced in Section 3.5.

For **RGB-CE**, we set the scaling factor $s \in [1,2]$ as a function of additive contrast $c_{\text{add}} = (L_{\max} + L_{\text{amb}})/(L_{\min} + L_{\text{amb}})$:

$$s = 1.45 - 0.025\, c_{\text{add}}, \tag{33}$$

For **Schlick-CC**, we define the compensated ambient term L_{comp} as a fraction of the display peak luminance $L_{\max}$, with the fraction decreasing log-linearly with $L_{\max}$:

$$L_{\text{comp}} = \min\left(L_{\text{amb}},\ L_{\max}\left(0.05 - 0.03\log\left(\frac{L_{\max}}{70}\right)\right)\right), \tag{34}$$

This makes the compensated ambient term display-luminance dependent while gradually reducing the compensated fraction at higher peak luminance, mitigating over-enhancement and chromatic artifacts.

For **ICtCp-CE**, we similarly define L_{comp} as a fraction of $L_{\max}$, but include a scene-dependent scaling factor $p \in [0,1]$ to adapt to the color characteristics of individual scenes:

$$L_{\text{comp}} = \min\left(L_{\text{amb}},\ p\, L_{\max}\left(1.0 - 0.4\log\left(\frac{L_{\max}}{70}\right)\right)\right). \tag{35}$$

For the OST-AR color-enhancement study in Section 4.2, we set p to 0.03, 0.04, 0.10, and 0.12 for *Fruits*, *Drawing*, *Cosmos Laundromat*, and *Sol Levante 3*, respectively.

Unlike **RGB-CE** and **Schlick-CC**, **ICtCp-CE** does not require an additional dynamic-range clipping step after chroma expansion. The chroma-expansion operation is bounded by the effective color gamut after accounting for the compensated ambient term; therefore, the transformed pixels remain within the target display range by construction.

Table 7. *Additive contrast of sampled OST-AR color-quality study conditions.* Additive contrast is computed using Equation (8) with $L_{\text{min}} = 0.0008$ nit (the EIZO black level after accounting for the 80% beam-splitter reflectance) for two peak-luminance settings, 200 nits (SDR) and 800 nits (HDR), across the sampled eye-reaching ambient luminance levels. Cell colors follow the convention in Table 3: green > 1,000:1, yellow 100:1–1,000:1, orange 10:1–100:1, and red < 10:1.

Peak Luminance	Ambient Luminance L_{amb} (nits)			
	0.01	**5**	**25**	**100**
200 nits (SDR)	18,520:1	41:1	9:1	3:1
800 nits (HDR)	74,075:1	161:1	33:1	9:1

Table 8. *N-way ANOVA results for the OST-AR color-quality study.* The dependent variable is perceptual quality, measured in just-objectionable-difference (JOD) units [Perez-Ortiz et al. 2019]. Factors include scene, peak luminance, color gamut, ambient luminance, and background type. The table reports main effects and two-way interaction effects. SS denotes sum of squares, and df denotes degrees of freedom.

Effect	**SS**	**df**	***F***	***p***
Scene	1.65	3	4.37	.007
Peak Luminance	0.66	1	5.29	.025
Color Gamut	0.93	1	7.44	.008
Ambient Luminance	109.01	2	433.71	< .001
Background Type	7.30	1	58.07	< .001
Scene × Peak Luminance	0.89	3	2.35	.080
Scene × Color Gamut	0.75	3	2.00	.123
Scene × Ambient Luminance	3.86	6	5.12	< .001
Scene × Background Type	0.51	3	1.35	.266
Peak Luminance × Color Gamut	0.01	1	0.06	.801
Peak Luminance × Ambient Luminance	0.17	2	0.68	.509
Peak Luminance × Background Type	0.54	1	4.31	.042
Color Gamut × Ambient Luminance	0.06	2	0.24	.786
Color Gamut × Background Type	0.13	1	1.05	.310
Ambient Luminance × Background Type	2.26	2	9.00	< .001

G User Study: Perceived Color Quality under OST-AR Viewing

This section provides additional details on the stimuli selection process for the OST-AR color-quality study (Section 4.1) and reports the full ANOVA statistics (Table 8).

Based on the additive-contrast estimates for representative display types in indoor and outdoor viewing environments (Appendix D.3), we sampled four eye-reaching ambient luminance levels for L_{amb}: 0.01, 5, 25, and 100 nits. The 0.01-nit level was the lowest luminance achievable with the DynaScan display after accounting for the beam-splitter and ND-filter transmittance. These levels cover more than the 90th percentile of eye-reaching ambient luminance for OST-AR displays in indoor viewing environments, and approximately the 60th percentile in outdoor viewing environments (see Table 2). We used two peak-luminance settings: 200 nits (SDR) and 800 nits (HDR). The HDR peak luminance was limited to 800 nits because video luminance values were scaled by 1.25 before presentation to compensate for the 80% beam-splitter reflectance; after this scaling, the displayed peak luminance must not exceed the 1,000-nit peak luminance of the EIZO displays. The target black level L_{min} was set to the EIZO black level after accounting for the beam-splitter reflectance, 0.0008 nit. Table 7 summarizes the additive contrast for all sampled conditions, which span 3:1–74,075:1.

Table 9. *Peak luminance required for sampled OST-AR color-enhancement viewing conditions.* Peak luminance values are computed from the sampled additive-contrast and ambient-luminance levels using Equation (8). Gray cells denote excluded conditions: $c_{\text{add}} = 4{:}1$ with $L_{\text{amb}} = 10$ nits requires an impractically low peak luminance, whereas $c_{\text{add}} = 12{:}1$ with $L_{\text{amb}} = 100$ nits requires an infeasibly high peak luminance that exceeds the EIZO display capability after accounting for the 80% beam-splitter reflectance.

Additive Contrast	Ambient Luminance (nits)		
	10	**40**	**100**
4:1	30	120	300
8:1	70	280	700
12:1	110	440	1,100

Table 10. *N-way ANOVA results for the OST-AR color-enhancement study.* The dependent variable is perceptual quality, measured in just-objectionable-difference (JOD) units [Perez-Ortiz et al. 2019]. Factors include scene, viewing condition, and enhancement method. Viewing condition denotes the sampled additive-contrast × ambient-luminance combinations. Enhancement method includes the no-enhancement baseline and the three color-enhancement algorithms. The table reports main effects and two-way interaction effects. SS denotes sum of squares, and df denotes degrees of freedom.

Effect	**SS**	**df**	***F***	***p***
Scene	4.75	3	10.49	< .001
Viewing Condition	7.17	6	7.91	< .001
Enhancement Method	7.84	3	17.31	< .001
Scene × Viewing Condition	6.98	18	2.57	.004
Scene × Enhancement Method	3.13	9	2.30	.029
Viewing Condition × Enhancement Method	5.28	18	1.94	.031

Two backgrounds were used (see Figure 16): *flat*, a uniform D65 white background, and *leaves*, a high-contrast leaves pattern following Chapiro et al. [2024]. Both backgrounds were presented with mean luminance calibrated such that, after beam-splitter and ND-filter transmission, the light reaching the observer produced the specified ambient luminance. This resulted in 4 SDR and 4 HDR ambient conditions for each of the Display P3 and BT.709 gamuts, yielding 16 conditions per scene and background, and 128 stimuli in total (4 scenes × 2 backgrounds × 16 conditions).

We used the same 2IFC protocol and ASAP framework described in Section 3.3 to schedule pairwise comparisons. Comparisons were allowed across color gamuts, ambient luminance, and backgrounds, but not across scenes or peak luminance. For each scene–peak-luminance block, the reference was set to the Display P3, zero-ambient, *flat* condition. Its *leaves* counterpart, which produced the same appearance under zero ambient luminance, was excluded to reduce experiment duration.

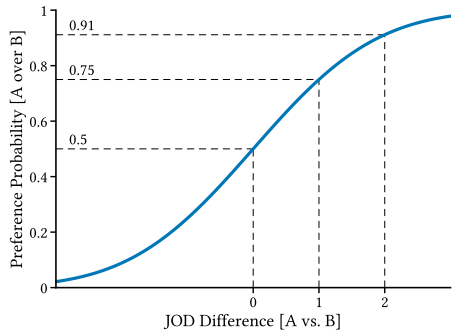

Fig. 17. *JOD difference vs. observer preference probability.* Just-objectionable-difference (JOD) [Perez-Ortiz et al. 2019] defines a relative quality scale in which score differences correspond to observer preference probabilities. A difference of 1 JOD between two conditions indicates that 75% of observers prefer the higher-scoring condition, 2 JODs correspond to 91% preference, and so on.

H User Study: Color Enhancement under OST-AR Viewing

This section provides additional details on the stimuli selection process for the OST-AR color-enhancement study (Section 4.2) and reports the full ANOVA statistics (Table 10).

OST-AR displays often operate at relatively low additive contrast (close to or below 10:1 [Chapiro et al. 2024]), because these devices are commonly used in bright indoor or outdoor environments and are driven as a function of ambient luminance rather than continuously at their peak luminance to improve power efficiency. Based on these considerations, we sampled three additive-contrast levels, 12:1, 8:1, and 4:1, and three eye-reaching ambient luminance levels, 10, 40, and 100 nits. The corresponding peak luminance was computed from the sampled additive-contrast and ambient-luminance levels using Equation (8). Two additive-contrast × ambient-luminance combinations were excluded: c_{add} = 4:1 with L_{amb} = 10 nits required an impractically low peak luminance, whereas c_{add} = 12:1 with L_{amb} = 100 nits required an infeasibly high peak luminance exceeding the EIZO display capability after accounting for the 80% beam-splitter reflectance. The target black level L_{min} was set to the EIZO black level after accounting for the beam-splitter reflectance, 0.0008 nit. Table 9 summarizes the peak luminance values for all sampled viewing conditions.

In addition to the three color-enhancement algorithms, we included a no-enhancement baseline in which no ambient-light compensation was applied. For the enhancement algorithms, the target display was assumed to have a Display P3 gamut, providing additional gamut headroom to compensate for ambient-induced color-fidelity losses. This setup reflects practical scenarios where visual content is mastered for a standard gamut such as BT.709, while display devices use custom color primaries that exceed these standards (e.g., the EIZO gamut shown in Figure 4, left). As in Section 4.1, video luminance values were scaled by 1.25 before presentation. The *flat* background, a uniform D65 white background, was used for all conditions. Its luminance was calibrated such that, after beam-splitter and ND-filter transmission, the light reaching the observer produced the specified ambient luminance. In total, this yielded 112 stimuli: 4 scenes × 4 enhancement methods (including the no-enhancement baseline) × 7 viewing conditions (3 additive contrast levels × 3 ambient luminance levels - 2 excluded combinations).