Subjective Evaluation of HDR10 Rendering Consistency Across Illuminance Changes

Pooshpanjan Roy Biswas, Thibault Cabana*, Adrien Carmone**
***DXOMARK**

Abstract
Smartphones are exposed to a variety of illumination changes. For the best user experience, it is important that their rendering is consistent across illuminance changes. In this study the rendering consistency performance of three different smartphones is studied across three illuminance conditions (0, 320 and 9000 lux) using a haploscopic setup. It was found that the Samsung S24 Ultra performed the best in maintaining a consistent rendering across these illuminance conditions.

Author Keywords
HDR10; illuminance; true-tone; adaptive-color; ambient-lighting etc

1. Introduction
Modern smartphones have developed various algorithms to account for illumination changes. The auto-brightness feature changes the luminance of the devices to give a comfortable reading experience. Along with this, color adaptive features, such as the True-Tone feature in iPhone 16 Pro Max (1), the Adaptive Color Tone feature in Samsung Galaxy 24 Ultra and the AI Adaptive Colors feature in Xiaomi 14T Pro (2) implement various algorithms through which the color appearance of rendering across illumination changes is maintained. Apart from this, there are illumination dependent tone-mapping operators, that further affect rendering experience.

Maintaining color appearance across illumination changes (illuminance and CCT) is of utmost importance as they affect final user experience and for relevant cases, also maintain mastering intent across the entire video production pipeline (3). Additionally, the final user would want to have the same perception of a memory color, or of a known color (a tshirt's color from a photograph that he took, a friend's skin color etc.) to remain consistent when there is an illumination change. The CIE has put forward various Color Appearance Models (CAMs) with which the color appearance correlates between illumination changes could be maintained (4,5). The illumination changes that can be incorporated in the models can be the illuminance as well as Correlated Color Temperature changes. The major workflow of such CAMs is the inclusion of a luminance adaptation as well as chromatic adaptation, and often these calculations are computationally very expensive. Smartphones often incorporate consistency across viewing conditions in a similar manner as proposed by complex CAMs, as explained in the patent filed by Apple Inc. (6) . These algorithms also have a similar luminance adaptation as well as chromatic adaptation using the data captured by Ambient Light Sensors (ALS) of the ambient lighting environment.

Subjective evaluation of consistency of renderings between different illumination conditions is not a trivial task. Fairchild describes the various techniques for comparing color appearance between varying viewing conditions (7) for creating corresponding colors data under different viewing conditions. They are described as memory matching, short term memory matching, magnitude estimation, asymmetric matching and haploscopic matching. Each method has its own merit and demerit. For more details, please refer to (7). For memory matching methods, subjects go back and forth between reference and test conditions to either match stimuli or estimate their magnitude (magnitude estimation methods using set scales; brightness, colorfuless, hue etc.). Memory matching involves sufficient time required to adapt to the individual viewing conditions. For each condition, the subject first looks at the stimuli, memorizes the appearance, moves to the test conditions, adapts again and then perform the color matching or magnitude estimation task (7). This method often requires special and substantial training period, complicated procedures for data analysis, lower precision than that of haploscopic technique, limited capacity for retaining information, and memory distortion (8).

For haploscopic evaluation, each eye is exposed to a separate half of the experiment scene both maintained at a different viewing condition (9). This ensures that each eye is adapted to the respective viewing condition. This holds the assumption that each eye is chromatically adapted individually and independently but since cognitive mechanisms depend on the input signals from both eyes, there could be some extent of confusion in the response of these mechanisms that confounds the results of traditional haploscopic (simultaneous viewing) experiments (10). Nevertheless, as haploscopic experiments are much faster to conduct than memory matching and due to its high precision nature, haploscope experiments are an excellent way of comparing stimuli under two viewing conditions (8). This technique has been used since a long time in classical color science experiments (11,12). Different modifications of the methods have also been proposed to counter the associated drawbacks of the methods, for example, the Successive-Ganzfeld Haploscopic Viewing Technique, which relies on the use of a specific type of stimulus pattern known as a Ganzfeld (for more details refer to (10)). Very recently, (13) used a haploscopic setup to collect corresponding colors data across two viewing conditions, 314 lux and 9420 lux, which are similar to the viewing conditions considered for our current study. In the past, the CIE TC 1-34 has also suggested the usage of a halpsocopic setup for color appearance and corresponding colors research (14).

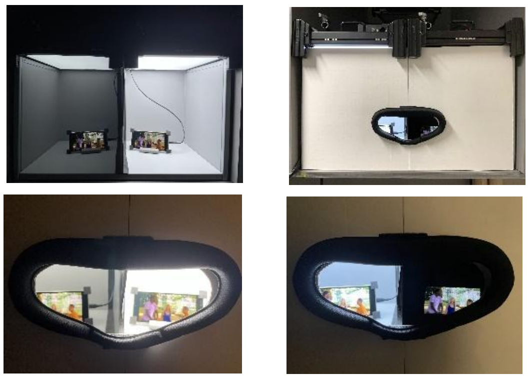
For the current study, a hapslocope was thus used to study the consistency of HDR10 video renderings across three viewing conditions for three smartphones pairs.

2. Experimental Apparatus
Haploscope: A modular and portable haplsocope was developed by DXOMARK to conduct the current study (see Figure 1). It was placed in a dark room. It comprised of two halves in the left and right side. The left side was always maintained at the reference illuminance level of 320 lux while the right side could be changed to either dark (~ 0 - 1 lux) or bright (~9000 lux) test situation. The haploscope was covered from the front as well, with a sufficient opening in the center to see the two halves of the setup with each eye respectively. The halves were divided using a gray PVC board and a soft foam was placed in the opening so

ISSN 0097-996X/25/5603-1511-$1.00 © 2025

that the area above the nose also is blocked, in order to limit illumination contamination from either halves to the other eye. This was done so that the illumination from the two halves have minimum interaction between them, which was especially important for the dark test situation.

a. **b.**



c. **d.**

Figure 1: The Haploscope developed by DXOMARK. a. setup to show cross section setup. b. Close actual experiment setup. c. Opening to compare reference (left half) to bright illuminance (right half). d. Opening to compare reference (left half) to dark situation (right half).

Each half was illuminated using a KinoFlo LED luminaire (Celeb 201 on the left and Celeb 250 on the right). The luminaires had a Color Rendering Index (CRI) of 95 and 96 respectively. The luminaires were controlled using a DMX controller by ENTTEC Pro via python.

Devices under test:

Three smartphones pairs were considered for this study:

1. Apple iPhone 16 Pro Max (x2)
2. Samsung Galaxy S24 Ultra (x2)
3. Xiaomi 14T Pro (x2)

All the illuminance and color adaptive features on these phones were turned on as follows:

1. Apple iPhone 15 Pro Max: Automatic Brightness ON (Accessibility Settings), True Tone ON
2. Samsung Galaxy S24 Ultra: Adaptive Brightness ON, Adaptive Color Tone ON, Screen Mode Vivid, White Balance Default
3. Xiaomi 14T Pro: Auto Brightness ON, Adaptive Colours ON (Original Colour PRO), Colour Temperature Default

These features ensured that the phones have their illumination and color adaptive features turned on so that changes in illumination could be accounted for and the color appearance of the presented stimuli are well adapted according to their respective algorithms in place.

Four HDR10 fixed frame video patterns from the DXOMARK HDR10 dataset were used for this experiment (see **Figure 2**), comprising of different scene compositions. They contained memory colors (sky, water bodies, plantations), preference colors (skintones), buildings and night scenes. The four patterns were named Daylight, Lake, Park and Night. More details about the recording of the video patterns and their mastering conditions could be found in a previous work by the authors ((15)).

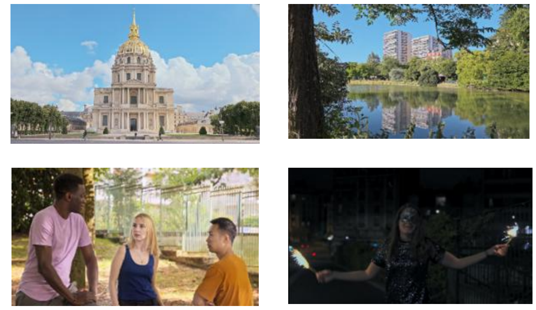


Figure 2: Fixed frame scenes used in the experiment, Daylight (top-left), Lake (top-right), Park (bottom-left) and Night (bottom-right).

3. Experiment Task:

An observer was seated on a chair in front of the haploscopic setup. He/she was asked to adjust the level of the chair so that the eyes are at the opening of the haploscope in a way that the left eye is looking at the left half and the right eye was looking at the right half. The smartphones were placed inside the haploscope in advance. The observer was then presented with the following instructions:

"In front of you, you see two phones placed in the left and right halves of a haploscope. Each half is illuminated differently. After spending 1-2 minutes in adapting the eyes to the respective illumination, the experiment will commence. Identical fixed frame video will be played on the two smartphones for 30 seconds. The task is to score them on a scale of 0-10 where 0 means that the video renderings on the two phones are completely dissimilar while 10 means that the video rendering between the two phones are exactly the same. You can consider color, contrast and brightness perception as important criteria in giving the score. Whenever the illumination on the right will be changed, 1-2 minutes of adaptation would need to be done before the next videos are played. There is no correct or wrong answer and there is no time limit for the entire experiment. If you feel exhausted, a break from the session could be taken."

As explained before, the reference illumination condition on the left was compared to either dark or bright illumination condition on the right. Thus, each subject did (4 videos x 2 illumination pair (ref-dark or ref-bright comparison) x 3 smartphone pairs) = 24 comparisons. The total time taken for these 24 comparisons was approximately 30 minutes with sufficient adaptation time in between illumination changes. It should be noted the four videos were played in random order for all observers, as well as the illumination change order was random (Phones 1 Ref-Dark followed by Ref-Bright for observer 1 could be Phones 1 Ref-Bright followed by Ref-Dark for observer 2). These randomization efforts were meant to avoid order bias associated to the devices. The devices were also camouflaged by placing black paper squares on their corners so that their brand identity is not associated with their scores.

The devices were controlled via **Android Debug Bridge** (adb) for the Android devices and with JellyFin (16) for the iPhones, all controlled via a python framework running on a laptop. The score were uttered by the observers to the experiment conductor without moving their head or eyes. The experiment conductor entered the score on the observer's behalf. This was done so that the observers' eyes adaptations states are not lost by moving their eyes away from the illumination.

4. Results and Data Analysis:

Objective Differences Among Rendering of the Devices: The four videos on the three phones under the three environments were captured using a Radiant Prometric imaging colorimeter. Seventeen Regions of Interest (ROI) comprising of skin tones, shadows, clothes, memory colors, etc. were selected on the acquisitions and CAM16-UCS color appearance correlates (J', a' and b') were calculated for each ROI using the viewing conditions under which the each of the videos were evaluated. Finally, color difference in CAM16-UCS (17) was calculated between each ROI under the reference condition with dark conditions, and under reference condition with the bright condition. This process was done to that objective differences between the rendering modes could be documented.

Table 1: Mean color difference expressed in CAM16-UCS space for 17 ROIs between reference and dark situation and reference and bright situation.

	iPhone 16 Pro Max	Samsung S24U	Xiaomi 14T Pro
Mean Ref to Bright	17.58	20.65	21.18
Mean Ref to Dark	35.47	32.37	20.6

14 observers (12 males, 2 females) with an average age of 34 were considered. Out of the 14 observers, 10 were experts and 4 were naïve in doing psychophysical experiments. The scores provided by the observers on a scale of 0 to 10 were averaged for the three phones. The distribution of scores for the three phones from all observers for the three phones were also compared with each other using for statistical significance of differences. ANOVA (Analysis of Variance) was employed to determine if there were statistically significant differences in **Similarity Scores** among the three devices—**Samsung S24 Ultra**, **Xiaomi 14T Pro**, and **iPhone 16 Pro Max**. This test assessed whether the variation in scores between the devices exceeded the variation within each device's scores, providing an overall significance (p-value < 0.05). When ANOVA indicated significant differences, Tukey's HSD (Honestly Significant Difference) post hoc test was applied to perform pairwise comparisons, identifying which specific devices differed significantly while controlling for Type I error. This two-step approach allowed for a comprehensive analysis of device performance across observer groups and categories. 95% confidence intervals (C.I.) were also calculated for the means of each group.

Results revealed that Samsung S24 has better performance than Xiaomi 14T Pro and iPhone 16 Pro Max for both bright to reference and dark to reference illumination condition comparisons. For bright to reference conditions comparison, the average mean score for Samsung S24 Ultra was 6.95 (95% C.I: 6.43 - 7.46) which was statistically significantly better than the mean score of 5.23 for iPhone 16 Pro Max (95% C.I: 4.68 – 5.79) but not significantly better than the mean score of 6.38 for Xiaomi 14T Pro (95% C.I: 5.83 – 6.92). For dark to reference conditions comparison, the average mean score for Samsung S24 Ultra was 5.77 (95% C.I: 5.23 - 6.31), was significantly better than the mean score of 4.36 for Xiaomi 14T Pro (95% C.I: 3.85 – 4.87) as well as the mean score of 4.84 for iPhone 16 Pro Max (95% C.I: 4.28 – 5.4). It is important to note that for bright to reference comparison, iPhone 16 Pro Max performed the worst compared to others (statistically significant), and for darkto reference comparison, Xiaomi 14T Pro had the worst performance which was statistically significant when compared to Samsung S24 Ultra but not when compared to iPhone 16 Pro Max.

Table 2: Overall results for bright and dark situations compared to the reference lighting situation with 95% confidence intervals. Statistically significant cases are marked by exponents i, S or X for iPhone 16 Pro Max, S24 Ultra or Xiaomi 14T Pro respectively

		Mean	95% CI Lower	95% CI Upper
Bright to Reference Comparison	Samsung S24 Ultra	**6.95**^{i}	6.43	7.46
	Xiaomi 14T Pro	6.38^{S}	5.83	6.92
	iPhone 16 Pro Max	**5.23** ^{S,X}	4.68	5.79
Dark to Reference Comparison	Samsung S24 Ultra	**5.77**$^{i,\,X}$	5.23	6.31
	Xiaomi 14T Pro	**4.36**^{S}	3.85	4.87
	iPhone 16 Pro Max	4.84^{S}	4.28	5.40

The data was also evaluated scene-wise for the four types of scene contents, daylight, park, lake and night (see **Table 3** below). It was found that Samsung S24 Ultra (S24U) had a higher mean score for all scene types for both bright and dark to reference comparisons (except lake scene in bright to reference comparison). For bright to reference comparison, iPhone 16 PM (16PM) had the worst performance for all scene types. The most common reason cited by observers was color shift, loss in contrast and overall reflection. For dark to reference comparison, Xiaomi 14T Pro (X14TP) had the worst performance for all scenes except "Night" and the most cited reason for this was the worse brightness adaptation of the device in the dark and loss in contrast. Because of the reduction in dataset size for such scene type comparison, not all these conclusions were found statistically significant, details of which can be found in the **Table 3** below. Thus, the inferences are based only on absolute average scores.

For comparison between the 10 experts and 4 naïve observers, it was found that for bright to reference comparison, iPhone 16 Pro Max performed the worst with an average score of 4.83 (statistically significantly), while S24 Ultra performed the best (6.72) with Xiaomi 14T Pro in between (6.1). For the casual observers, no statistical significance was found, primarily due to low dataset size, although Samsung S24 Ultra had the highest average score (7.5 compared to 7 and 6.25 for X14P and 16PM respectively). For the dark to reference comparison, S24U had statistically significantly best performance in terms of average score (6.09 versus 4.38 and 4.88 for X14P and 16PM respectively). A statistically significant worst performance was not identified for this case in between 16PM and X14P. Casual observers scored the16PM (4.75) a bit higher than S24U (4.58) or X14P (4.31), nevertheless, the findings were not statistically significant. It is also important to note that the overall mean scores provided by the observers did not agree well with the trends predicted by objective metrics, such as CAM16-UCS. This further proves the significance of such psychophysical studies in understanding user preference towards overall video rendering.

5. Conclusions

A subjective evaluation was performed using a haplsocope where the appearance agreement between renderings under different illuminance for the same smartphone displays was studied. It was found that Samsung S24 Ultra performed the best in implementing color appearance uniformity across illuminance changes. This was found among the overall observer pool but also across sub-groups and for different scene contents. iPhone 16 Pro Max performed the worst when renderings under bright situations was compared to the reference situation, while it was Xiaomi 14T Pro that performed the worst when comparing renderings under dark situation to a reference situation. This experiment provided the first evidence that all smartphones manufacturers handle cross illuminance color appearance changes differently, and there is a substantial room for improvement in the implementation of algorithms in this domain.

Table 3: Scene wise results for bright and dark situations compared to the reference lighting situation. Statistically significant cases are marked by exponents i, S or X for iPhone 16 Pro Max, S24 Ultra or Xiaomi 14T Pro respectively

		Bright to Reference	**Dark to Reference**
Category	**Device**	**Mean**	**Mean**
Daylight	S24U	7.07	5.5
Daylight	X14TP	6.36	3.79
Daylight	16PM	5.64	4.71
Park	S24U	6.64	6.07
Park	X14TP	6.14	4.36
Park	16PM	5.29	4.93
Lake	S24U	$7.00^{S,i}$	5.57
Lake	X14TP	$7.14^{X,i}$	4.43
Lake	16PM	$5.14^{X,i}$	5
Night	S24U	$7.07^{S,i}$	5.93
Night	X14TP	5.86	4.86
Night	16PM	$4.86^{S,i}$	4.71

6. References

1. Apple. Adjust the brightness and color temperature on your iPhone or iPad - Apple Support (BH) [Internet]. 2024 [cited 2024 Nov 28]. Available from: https://support.apple.com/en-bh/109351
2. Xiaomi. Xiaomi 14T Pro - Xiaomi Global. 2024 [cited 2024 Nov 29]; Available from: https://www.mi.com/global/product/xiaomi-14t-pro/
3. Poynton C. Colour Appearance Issues in Digital Video, HD/UHD, and D-cinema Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy Under Special Arrangements with Graduate and Postdoctoral Studies and Faculty of Applied Sciences. 2018.
4. CIE. CIE 248:2022 The CIE 2016 Colour Appearance Model for Colour Management Systems: CIECAM16. 2022 Mar 1 [cited 2024 Dec 6]; Available from: https://cie.co.at/publications/cie-2016-colour-appearance-model-colour-management-systems-ciecam16
5. CIE. A colour appearance model for colour management systems: CIECAM02 | CIE [Internet]. 2004 [cited 2024 Dec 6]. Available from: https://cie.co.at/publications/colour-appearance-model-colour-management-systems-ciecam02
6. Chen. United States US 20170 103728A1 (12) Patent Application Publication (10) Pub Related U.S. Application Data (63) Continuation of application No. 14/673,667, filed on. 530(9).
7. Fairchild MD. Color Appearance Models. Color Appearance Models. 2013. 199–212 p.
8. Luo MR. A review of chromatic adaptation transforms. Vol. 30. 2000.
9. Fairchild MD. Color Appearance Models. Color Appearance Models. 2013. 199–212 p.
10. Fairchild. Successive-Ganzfeld Haploscopic Viewing Technique for Color-Appearance Research. 1994.
11. Hunt RWG. Light and Dark Adaptation and the Perception of Color*.
12. Bartleson CJ, Bseneman EJ. Brightness Perception in Complex Fields. Vol. 57, JOURNAL OF THE OPTICAL SOCIETY OF AMERICA. 1967.
13. Bao W, Wei M. Change of color appearance due to extremely high light level: Corresponding colors under 100 and 3000 cd/m2. In: Final Program and Proceedings - IS and T/SID Color Imaging Conference. Society for Imaging Science and Technology; 2019. p. 320–5.
14. Fairchild MD. CIE TC1-34 Testing Colour-Appearance Models Testing Colour-Appearance Models: Guidelines for Coordinated Research.
15. Bernardi P, Cabana T, Guichard F. P-44: Real-Content Based Method for HDR Tone Curve Characterization. In: SID Symposium Digest of Technical Papers. Wiley Online Library; 2023. p. 1678–81.
16. Jellyfin. The Free Software Media System | Jellyfin [Internet]. 2023 [cited 2024 Jan 22]. Available from: https://jellyfin.org/
17. Li C, Li Z, Wang Z, Xu Y, Luo MR, Cui G, et al. Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS. Color Res Appl. 2017;42(6):703–18.