https://doi.org/10.2352/EI.2024.36.9.IQSP-268

Evaluation of Subjective Video Quality of Television Displays

Eddie Pei[1], Hosub Lee[2], Elena Fedorovskaya[1], Susan Farnand[1]
[1] Rochester Institute of Technology; Rochester, NY, USA
[2] Samsung Research America; Irvine, CA, USA

Abstract

With the fast-evolving video and display technologies, there is an interest in better understanding user preferences for video quality and the factors that impact these preferences. This study focuses on the subjective video quality assessment (VQA) of TV displays, considering a range of factors that influence viewer experience. We conducted two psychophysical experiments to investigate the latent factors affecting human-perceived video quality. Our results offer insights into how different factors contribute to video quality perception. This research guides researchers and developers aiming to improve display and environmental settings to give end-users an optimal viewing experience.

Introduction

With the advancement in video and display technologies, the demand for high-quality video has increased markedly over recent years. In this context, assessing the video quality to ensure the optimal user viewing experience has become crucial. Video quality assessment (VQA) involves analyzing various factors that affect the overall human-perceived visual quality of video content, such as image resolution, content genre, viewer characteristics (e.g., age), and many others. Park et al. studied how people perceive image quality differently depending on the type of content. Through the experiments, they concluded that content genre-based image quality adjustment (e.g., increased saturation for sports scenes) is necessary [1]. Kufa et al. investigated the perceptual quality of video content presented in Full HD and Ultra HD resolutions at different viewing distances. This research highlighted the importance of considering diverse viewing conditions when performing VQA [2]. Wetzel et al. stated that different environmental factors could affect the legibility of the large screen display. Factors such as ambient lighting, viewing distance, and viewing angle significantly impacted the visual legibility on the screen [3]. Chubarau et al. also stated the importance of evaluating image quality under different viewing conditions and display systems [4]. Regarding this, Baek et al. studied how ambient lighting conditions impact viewers' visual perception of display devices. They found that the subjects preferred matching the TV display's color temperature with the surroundings [5]. Gofaizen et al. emphasized the importance of sharpness in perceived image quality on display devices. They explored image processing approaches to correctly control the sharpness level of images shown on TV [6]. Moorthy et al. assessed the video quality in mobile devices. They prepared the content with different levels of quality (e.g., via distortion) considering real-life situations (e.g., slow network) and collected the human opinions scores of its perceptual quality [7]. Winkler discussed the connections between the human visual system and video quality and highlighted the challenges in developing vision models for perceptual video quality assessment [8]. Additional studies discussed other factors, including spatiotemporal attribute [9], dynamic range (SDR and HDR) [10], and demography [11].

VQA involves the use of subjective and objective methods. Objective methods use mathematical or statistical tools to analyze the technical aspects of the image or video content itself, and it is highly efficient and easy to deploy. Over the years, more advanced objective methods have been developed, including spatial domain methods such as deep similarity index (DeepSim), based on a deep neural network [12]. Zhai and Min provided a thorough review of the objective VQA methods [13]. However, these methods are generally pixel-by-pixel based, not considering the actual user-perceived visual quality that environmental factors can influence. Conversely, subjective methods rely on human observers to perceive and evaluate the quality of video content based on their personal preferences and opinions. Pinson and Wolf summarized different subjective evaluation methods while providing valuable insights into the strengths and limitations of different methodologies [14]. Both objective and subjective methods can help evaluate video quality. However, the subjective VQA method is more suitable for measuring observers' image quality preferences while considering user-specific and situational factors.

In this study, we aim to empirically uncover and understand TV users' video quality preferences and impacting factors, and consequently, we adopt the subjective VQA method. With the factors chosen based on the literature and our experience, we designed and conducted two psychophysical experiments with human observers ($N = 37$). The rest of this paper introduces the experimental setup, analysis methodologies, and the results.

Factors

As discussed, subjective VQA can be influenced by many factors, including but not limited to, viewer characteristics such as age, culture [15], and experience. Factors like the observer's degree of chromatic adaptation as well as the correlated color temperatures (CCT) and level of the surrounding illumination can also play a significant role.

Experimental Methodology

The goal of the experiments is to determine how users perceive picture quality on TV displays and what latent factors affect this human visual perception. Toward this end, we designed two experiments to evaluate the impact of varied combinations of select factors on specific video quality, implemented by a TV display's picture setting configurations and an image processing algorithm. In the first experiment, we asked subjects' preferences regarding specific picture settings we developed based on TV usage log data [16, 17]. These settings differ from the factory default, mainly on brightness, contrast, and sharpness levels. The second experiment focused on the video's vividness (or colorfulness). Here, we used three different levels of vividness determined by a proprietary perceptual color enhancement algorithm [18] and collected subjects' preferred vividness levels accordingly. Both of these experiments involved the same participants, TV displays, lighting environment, and similar content.

Participants

Individuals with normal color vision were recruited as participants in the experiments. Each participant provided informed consent after being briefed on the study procedures. RIT's Human Subjects Research Office has approved this experiment (approval FWA #00000731).

Stimuli and Lighting Condition

The experiment comprised six distinct visual stimuli, as depicted in Figure 1. We selected the content to represent various video categories, including bright and dark settings, diverse skin tones, and animated and real-life scenes.

(a) Animation

(b) Game video

(c) Dark movie scene

(d) Nature scene

(e) Soccer scene

(f) Bright movie scene

Figure 1: Visual Stimuli (Content)

The experiments were conducted in the Munsell Color Science Laboratory Dynamic Visual Adaptation Lab, which features a ceiling-mounted, five-channel tunable LED system. Various desired lighting conditions can be achieved by adjusting the weights for these channels. This study employed four distinct lighting conditions: (1) Dark Warm, (2) Bright Warm, (3) Dark Cool, and (4) Bright Cool (see Table 1). We confirmed ambient lighting conditions by measuring lights via an MK350N spectroradiometer.

Experiment I: Procedure

We divided the experiment into four blocks, each representing one of the specific lighting condition explained above. A one-minute adaptation period between the blocks was given to the observers, allowing them to adapt to the ambient light changes adequately.

Table 1: Select Factors

CCT (2)	Illuminance (2)	Stimuli (6)	TV Settings (5)
2700K (Warm)	15L (Dark)	Animation	Dark Standard Mode*
5500K (Cool)	350L (Bright)	Game	Bright Movie Mode*
		Movie (Dark)	Low Vividness**
		Nature	Medium Vividness**
		Sports	High Vividness**
		Movie (Bright)	

Note: * and ** were used for Experiment I and II, respectively.

Two identical 65-inch Samsung TVs (Model: QN85C) are installed side-by-side and used as the displays. We instructed observers to sit nine feet from the center of two TV displays (see Figure 2). The viewing angles for both TVs are identical. A 15-minute warm-up period is allowed for the TVs and the LED ceiling lights.

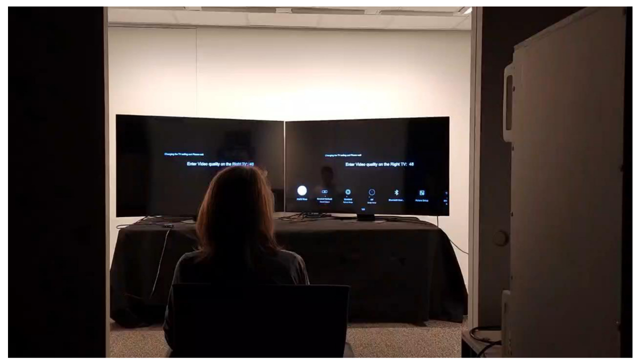


Figure 2: Experimental Setup

We prepared 8-second-long 4K SDR video clips per selected content genre (see Figure 1). The original video format was Apple's QuickTime (MOV), and we encoded it using HEVC for easy content playback on TVs. As discussed before, we prepared specific picture settings (also known as Picture Mode) derived from analyzing TV usage log data [16, 17]. Here, we focused on two popular Picture Modes, Standard and Movie. Our data analysis indicated that some users prefer to customize the default Standard and Movie Picture Mode. To explain, some Standard Mode users tend to lower the TV display's brightness level by 22%, compared to the default (namely, Dark Standard Mode). Different preferences exist for Movie Mode, like increasing the brightness setting by 37% more than the default (namely, Bright Movie Mode). Thus, we presented the observers with images having two different picture settings (i.e., default vs. user-preferred) under each Standard and Movie Picture Mode and asked their preferences.

We adopted the double stimulus continuous quality scale (DSCQS) method regarding the evaluation methodology. The reference stimulus (default Picture Mode on the left TV) was assigned a score of 50, and observers rated the video quality of the control stimulus (user-preferred Picture Mode on the right TV) on a scale as low as 0 but without upper limits compared to the reference. Observers responded to the same content twice (two repetitions), and we fully randomized the order of the video content to prevent memory effects.

Considering four ambient lighting conditions, two Picture Mode categories (Standard and Movie), five visual stimuli, and two repetitions, each observer performed 80 video quality assessments. This yields approximately 25 minutes of experiment time per subject. We used a MATLAB program to control the experimental procedures and a Bluetooth keypad as the input device for observers to submit their responses.

Experiment II: Procedure

All the experimental settings (e.g., viewing distance) are the same as the first experiment, except that we focused on the vividness of video content displayed on TV and utilized an additional stimulus (a bright movie scene; see Figure 1f). Both TVs are configured with identical picture settings (Standard Mode) but differ in their vividness level. The reference TV employs a baseline (default) vividness, while the comparison TV is set to (1) low, (2) medium, and (3) high levels of vividness for separate trials.

This experiment consists of four blocks of different lighting conditions, three distinct vividness levels on TV displays, and six stimuli with two repetitions. This section is divided into eight segments, encompassing 120 trials. The duration of this experiment is approximately 40 minutes per subject. The MATLAB program controls the entire experimental procedure as it did for the first experiment.

After completing both experiments, we surveyed observer characteristics such as age, gender, and television viewing habits and briefly interviewed each subject about their video quality preferences.

Results and Discussion

A total of 37 observers participated. There are 2,960 observations for Experiment I and 5,328 observations for Experiment II.

Analysis of Factors

Experiment I

An ANOVA test was applied to the data. The results are shown in Figure 3. Drawing conclusions from the P_{value}, the significant factors include TV's picture settings (Display Setting), luminous intensity (Intensity), and the type of visual stimuli (Genre). The interactions between Display Setting & Intensity and Display Setting & Genre are significant in Experiment I.

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Display Setting	80444.5	1	80444.5	451.76	0
CCT	329.1	1	329.1	1.85	0.1741
Intensity	3830	1	3830	21.51	0
GENRE	6002.6	4	1500.7	8.43	0
Display Setting:CCT	97.4	1	97.4	0.55	0.4596
Display Setting:Intensity	1309.8	1	1309.8	7.36	0.0067
Display Setting:GENRE	1663	4	415.8	2.33	0.0534
CCT:Intensity	54.9	1	54.9	0.31	0.5789
CCT:GENRE	198	4	49.5	0.28	0.8924
Intensity:GENRE	236.9	4	59.2	0.33	0.8562
Error	522991.5	2937	178.1		
Total	617157.7	2959			

Figure 3: ANOVA (Experiment I)

The observer was asked to compare the default picture settings to user-preferred ones under two presets (i.e., Standard and Movie Picture Mode). Here, we aimed to understand how study subjects perceived the user-preferred Picture Mode compared to the default. To determine the overall preferences of observers towards the given Picture Mode, we analyzed the collected data as follows. Data from Experiment I was segmented by Picture Mode: Standard and Movie. A response under 50 indicates a preference for the reference (default) Picture Mode. A response equal to 50 indicates no preference between the two Picture Modes. A response exceeding 50 signifies a preference for the user-preferred Picture Mode (e.g., Bright Movie). We then performed two t-tests separately for each Standard and Movie Mode data. The null hypothesis was that the user-preferred Picture Mode is equivalent to or worse than the default. The t-test results indicated we could not reject the null hypothesis at the significance level of .05 for Standard Mode but could do so for Movie Mode data. We therefore concluded that observers preferred the default picture settings under Standard Mode but preferred the user-preferred settings (increased brightness) for Movie Mode.

The Experiment I data was further divided by ambient lighting conditions and content. There are four lighting conditions: (1) Dark Warm, (2) Bright Warm, (3) Dark Cool, and (4) Bright Cool, and five different visual stimuli: (1) Animation, (2) Game, (3) Movie (Dark), (4) Nature, and (5) Sports. In both Standard and Movie Mode, the observers' responses for each video content under specific lighting conditions are presented in Figure 4. Regarding Movie Picture Mode, mean responses consistently exceed 50 regardless of ambient lighting and content (see Figure 4b). This suggests that observers generally prefer the user-preferred (bright) Movie Mode. In contrast, most of the means fall below 50 for the Standard Mode, indicating a preference for the default setting. However, in the case of warm and low-intensity lighting conditions, observers tend to favor the user-preferred setting (darker than the default Standard) except for Animation content.

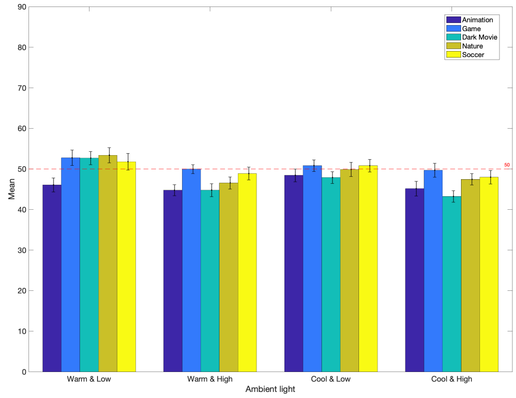
We performed multiple comparison tests on the entire Experiment I data to better understand the inter-relationships between ambient lighting and content genre. Table 2 demonstrates the results of the multiple comparison t-tests across different lighting conditions with varying content pairs. Since each lighting condition consists of ten comparison content groups, we performed the Bonferroni correction ($\alpha_{corrected} = \frac{0.05}{10} = 0.005$).

Table 2: Multiple t-tests on Content/Lighting (Experiment I)

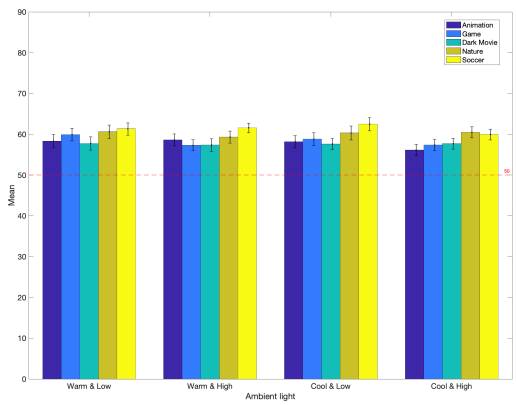
Content\Lighting	Dark Warm	Bright Warm	Dark Cool	Bright Cool
Animation vs. Game	0.0097	**0.0039**	0.2665	0.0721
Animation vs. Movie	0.0057	0.9898	0.7966	0.4061
Animation vs. Nature	**0.0045**	0.3747	0.5387	0.3190
Animation vs. Sports	0.0336	0.0518	0.2821	0.2499
Game vs. Movie	0.9784	0.0089	0.1558	**0.0045**
Game vs. Nature	0.8116	0.0668	0.6810	0.3127
Game vs. Sports	0.7313	0.5710	0.9949	0.4755
Movie vs. Nature	0.7771	0.4175	0.3811	0.0364
Movie vs. Sports	0.7336	0.0716	0.1698	0.0306
Nature vs. Sports	0.5665	0.2852	0.6924	0.8021

Note: We used dark scenes under Movie genre for Experiment I.
Note: Significant p-values instances are emphasized.

As can be seen, we confirmed several circumstances that led users to prefer customized picture settings other than the default. Specifically, our study subjects are more likely to prefer the dark version of Standard Mode when watching Nature content compared to Animation content under the room lighting condition, which was dark and warm (see Animation vs. Nature and Dark Warm in Table 2; $p < .005$). The Nature content was sunset scenes with warm colors, so we suspect this makes subjects perceive darkness, favoring a darker image tone. The opposite goes for the Movie Mode. The subjects tend to prefer the bright Movie



(a) Mean responses for user-preferred Dark Standard Mode



(b) Mean responses for user-preferred Bright Movie Mode

Figure 4: Overall Preferences of TV's Picture Settings

Mode when watching cinematic content over game content under the bright and bluish room lighting (see Game vs. Movie and Bright Cool in Table 2; $p < .005$). This finding sounds reasonable based on the literature: people prefer brighter images on a display device (e.g., TV) in a bright environment and vice versa [16, 17].

The same analysis was applied between four different lighting conditions. The results are listed in Table 3. There are statistically significant differences between Dark and Warm and Bright Warm, Dark Warm and Bright Cool, and Bright Warm and Dark Cool ($p < .0083$, Bonferroni-corrected for six comparisons). The results indicate that the room illuminance level has a more substantial effect than the CCT.

Experiment II

An ANOVA test was applied to the data collected for Experiment II. The test results indicated the Genre was the significant factor (see Figure 5). The interaction of Display Setting & CCT was significant, too. This experiment's display (picture) settings did not drastically change the video in terms of color appearance. Instead, it shows different levels of vividness (low, medium, high) according to the experiment protocols. During the experiment,

Table 3: Multiple t-tests on Lighting (Experiment I)

Lighting Condition Pairs	p_{value}
Dark Warm vs. Bright Warm	**0.00**
Dark Warm vs. Dark Cool	0.3577
Dark Warm vs. Bright Cool	**0.0058**
Bright Warm vs. Dark Cool	**0.0011**
Bright Warm vs. Bright Cool	0.1923
Dark Cool vs. Bright Cool	0.0582

Note: Significant p-values instances are emphasized.

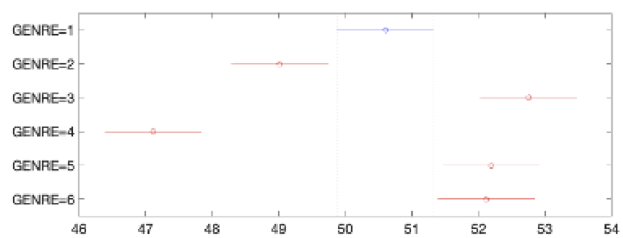
the observers could barely identify the difference between baseline and control displays. This might be one of the reasons why we obtained different results about impacting factors compared to Experiment I, which showed quite noticeable differences in picture quality between the two TVs.

The observers' responses grouped by the Genre factor are shown in Figure 5b. Animation (Genre 1) and Game (Genre 2) content are significantly different from other content; they are both animated scenes, which may make them stand out. It also indicates that Movie (Dark; Genre 3), Sports (Genre 5), and Movie (Bright; Genre 6) content are similar to each other and distinct from others. The commonality between these videos is that they all include human skin tones.

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Display Setting	291.9	2	145.94	1.26	0.2835
CCT	278	1	277.99	2.4	0.1213
Intensity	288.1	1	288.06	2.49	0.1147
GENRE	21428.2	5	4285.65	37.03	0
Display Setting:CCT	923.5	2	461.74	3.99	0.0186
Display Setting:Intensity	382.5	2	191.26	1.65	0.1917
Display Setting:GENRE	1416.9	10	141.69	1.22	0.2695
CCT:Intensity	86.5	1	86.54	0.75	0.3872
CCT:GENRE	315.7	5	63.14	0.55	0.7419
Intensity:GENRE	686.2	5	137.24	1.19	0.3134
Error	612613.3	5293	115.74		
Total	638725.2	5327			

(a) ANOVA results



(b) Responses per content genre

Figure 5: ANOVA (Experiment II)

We also performed a two-sample t-test to investigate differences in viewer preferences towards Movie (Dark) and Movie (Bright) content under a bright TV viewing environment. A null hypothesis was that there exist no differences in user preferences towards dark and bright scenes if they are in the same movie genre (see Figures 1c and 1f). The test rejected the null hypothesis at the 5% significance level. Therefore, we state that people perceive picture quality differently, even in the same content, and the current genre categorization may not effectively capture people's actual perception and preferences.

Appearance Based Analysis

The study subjects' feedback was collected to get ideas of how they evaluated the videos. According to the feedback, the descriptors commonly used were brightness, chroma, naturalness, skin tone, and grass color. These words suggested that some observers judged by overall scene color attributes, and others were more focused on object colors, such as human skin tone and grass colors. To get more insight into the correlation between preferences and these factors, we carried out two color appearance-based analyses. The color attributes and the representative colors of the objects in the videos were analyzed.

Color Attributes Analysis

Display Characterization Models

This project examines TV displays with eight distinct settings along an ambient environment variation containing four unique lighting conditions. This results in a total of 32 combinations for the analysis of TV settings and ambient conditions. For the display characterization, the PR655 spectroradiometer was used to measure the radiance of different color ramps, afterward converting these measurements into absolute XYZ values. Consequently, 32 distinct gamma-offset-gain (GOG) models were created for different conditions. The transformation sequence can be mathematically formulated as follows:

$$RGB \longrightarrow SPD \longrightarrow XYZ \tag{1}$$

Each display model corresponds to a specific combination of display setting and ambient light, ensuring the capture of the display's color rendering capabilities under varied environmental lighting. These models aid in conducting the color attributes and appearance analysis.

Image Appearance Attributes

In the experiment, five videos were used for analysis (see Figure 1). Keyframes from these videos were extracted and saved as RGB png files. The RGB values were then converted to XYZ values using the GOG models. The XYZ values obtained are absolute and were the basis for further transformations or normalization.

To standardize the data, a chromatic adaptation model was used to convert these XYZ values under a XYZ_{D65} white point. Given that the experimental setup involved two TVs side-by-side with ambient lighting, three distinct white points were present: one for each TV and one for the ambient light. The white point for adaptation was chosen based on the brightest one for each combination. The chromatic adaptation is as follows:

$$XYZ_{D65} = M_{16}^{-1} \times M_{adp} \times M_{16} \times XYZ_{im} \tag{2}$$

where M_{16} is from CIECAM16, $M_{adp} = \text{diag}(D65./wXYZ_{scene})$; represents the adaptation matrix to convert scene white point to D65.

Image Appearance Analysis

The image appearance parameters were assessed using an image appearance model in IPT space, which includes lightness (I), chroma (C), and hue (h). The contrast sensitivity function (CSF) was utilized to filter the images. Three filters were applied to the luminance and two chromatic channels. The filters are calculated as follows:

$$csf_{lum} = a \times f^c \times e^{-b*f} \tag{3}$$

$$csf_{chroma} = a_1 \times e^{-b_1 * f^{c_1}} + a_2 \times e^{-b_2 * f^{c_2}} \tag{4}$$

The filters are operated in IPT color space. Fourier transformations are used. csf_{lum} is applied to I channel, and two csf_{chroma} are applied to P and T channels. The parameters' value can be found in [19]. And f is defined in terms of cycles per degree of visual angle (cpd), which is

$$f = \frac{ppi}{\frac{190}{\pi} \times tan^{-1}(\frac{1}{dis})} \tag{5}$$

where ppi is pixels per inch of the display, dis is the distance between observer and display.

During the keyframe analysis, two TVs were assessed: one as a reference and the other as the test display. Differences in appearance attributes were quantified, including lightness difference (ΔI) and chroma difference (ΔC).

Results

Figures 6 and 7 present the results obtained from Experiment I. Each data point represents a specific video content assessed under a particular lighting condition (see Table 5 in Appendix). In these plots, the y-axis represents the mean preference response of the observer, while the x-axis indicates the Δ values for various appearance attributes. The data points in these plots tend to be clustered into two distinct groups: Standard and Movie Picture Mode. Notably, in the Movie Mode, subjects generally show a preference for the test configuration (i.e., Movie Mode with increased brightness), whereas in the Standard Mode, the default one tends to be favored.

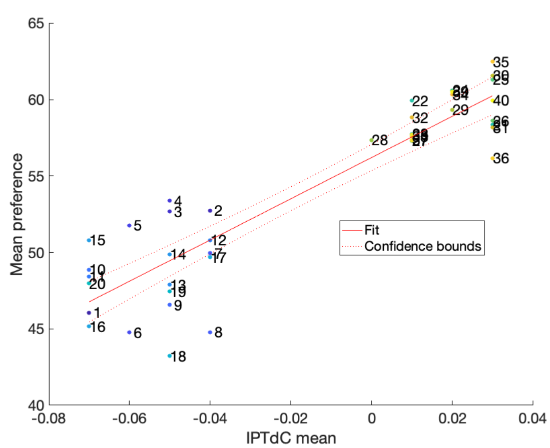


Figure 6: Mean IPTdC and Mean Response (R^2 =0.86)

The regression analysis is applied to determine how lightness and chroma influence preference. The regression models and their confidence bounds are plotted in Figures 6 and 7. The models have a reasonable performance with R^2 values of 0.86 and 0.71, respectively. Importantly, the coefficients for both models are positive, implying a linear relationship between lightness and

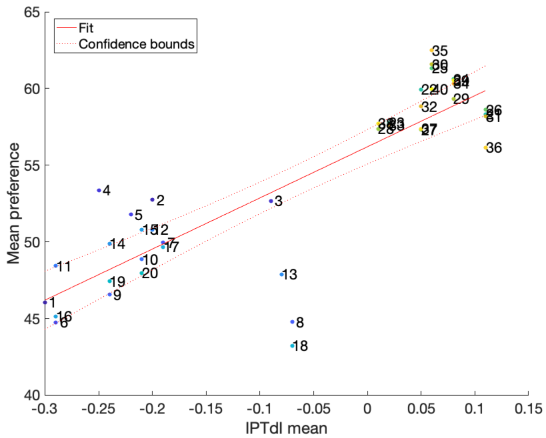


Figure 7: Mean IPTdI and Mean Response (R^2 =0.71)

chroma and preference so that as the lightness and chroma increase, preference appears to increase correspondingly.

The analysis based on the image appearance model shows a clear trend. The conclusion drawn from this analysis is that observers tend to prefer images that are brighter and more chromatic. This preference pattern is consistent across the different display settings in Experiment 1. To explain, a given video under four unique lighting conditions follows a similar order. As an example, data points 3, 13, 8, and 18, which are the Movie (Dark) content under four lighting conditions. Their order is followed by Dark Warm, Dark Cool, Bright Warm, and Bright Cool conditions. The results are aligned with the Bonferroni test in terms of lighting conditions, in which observers prefer the dimmer ambient lights.

Representative Color Analysis

In this section, image segmentation is used as a method to analyze visual appearances of the test content. Initially, each image is transformed into the CIELAB color space. Following this, a K-means clustering algorithm is applied to segment the images in CIELAB space, setting K to four. This segmentation aims to isolate four principal colors in each image. Figures 8 through 14 (Figures 11 to 14 are listed in the Appendix) show the segmented colors and their corresponding areas within each image. The presentation sequence starts with the original image, followed by the four representative colors, which are computed by averaging the colors within their respective areas. The subsequent four images display the individual areas associated with each of these representative colors.

Figure 8 shows four segmented areas representing the grass, skin tones, and two uniforms. They are aligned with the objects' colors that the observers used for evaluating their preferences.

For this study, a total of 32 display models were generated, each corresponding to a unique combination of TV's picture setting and ambient lighting condition. These display models convert the representative colors from RGB to XYZ. Subsequently, CIECAM16 is applied to convert these XYZ values to the D65 white point. The white luminance is the absolute white in cd/m^2, and adapting luminance is set to 20% white luminance and the average condition with full adaptation. The average condition was used because the luminance of the white in the scene was always

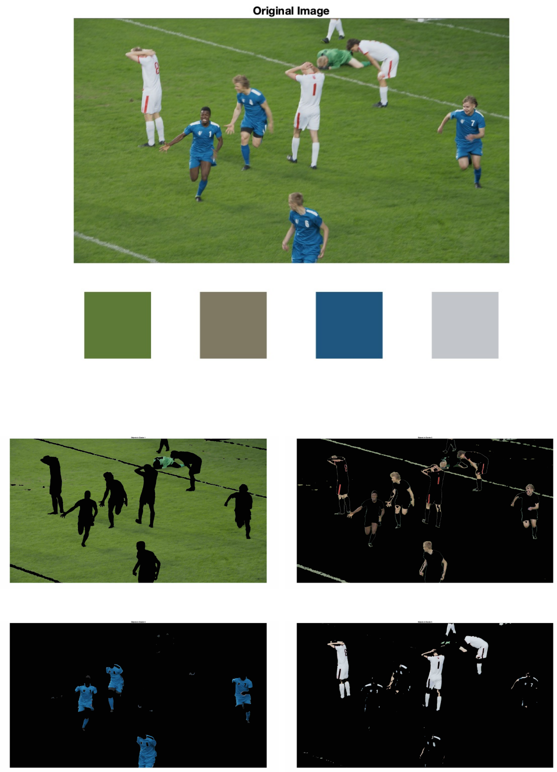


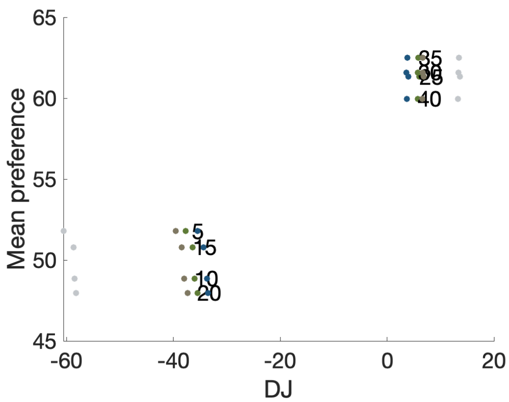
Figure 8: Color Segmentation Results (Sports Content)

above 200 cd/m^2. The color attributes were then calculated, and the differences in the attributes were determined. The analysis then focuses on correlating these appearance attributes with the viewers' preference responses for each unique experimental condition.

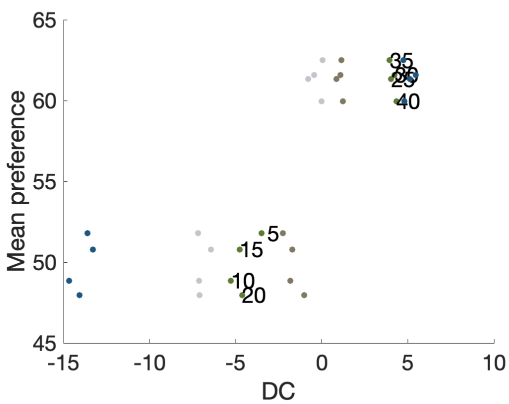
The attributes of the represented colors were calculated using CIECAM16 and used for analysis. Since each video features distinct representative colors, the data are categorized and analyzed according to the individual videos. Figure 9 depicts how the preference rating changes according to the lightness and chroma of four representative colors in Sports content (soccer scene). The index numbers remain consistent with those used in the image appearance attribute analysis (see Table 5). To enhance clarity and facilitate interpretation, the color of each dot is matched with the actual color of the represented area within the video. As we confirmed in the appearance-based analysis, for a given representative color, the more chromatic and brighter, the more observers preferred the video.

Common Representative Colors

There are several similar representative colors among the content used in our experiments. Specifically, Animation, Movie (Dark), and Sports content all have green as one of the representative colors (see Figures 1a, 1c, 1e, respectively). The hue angles for these greens were calculated in the CIECAM16. The hue angle ranges between 113° and 131°. The hue differences are within 20°, as shown in Figure 10. The green hues in different items, an animated green tree, grass in the soccer field, and a green block, and their color attributes difference are not too far



(a) Lightness plot of Sports content



(b) Chroma plot of Sports content

Figure 9: Mean Preference per Lightness and Chroma of Representative Colors

away from each other (see Table 4), but the preferences of these videos are very different. An interesting aspect to consider is the context in which the green color appears. In Movie (Dark) and Sports content, green is a part of real-life objects, while in Animation content, it is part of animated scene. Even though similar shades of green were presented to the observers, their preferences significantly varied between real-life and animated footage. Similarly, Animation, Game, and Nature content all contain blue sky (see Figures 1a, 1b, 1d), but Animation and Game content are both animated, and Nature content is a real scene. As a result, the preference for Nature content is significantly different than for Animation and Game content. Even when blue is ranked as the second prominent representative color for Game and Nature content, their preferences are significantly different from each other.

Based on all the comparisons above, we state that the substantial differences in content lie in whether the scene was animated or captured in real life. Therefore, the naturalness of the videos and observers' memory color of specific objects (e.g., grass) affect people's video quality preferences.

Table 4: Differences in Attributes of Green Representative Color

Video	$\Delta Lightness$	$\Delta Chroma$	$\Delta Saturation$	Δhue
Animation	5.20	1.94	0.43	-2.82
Movie (Dark) 1	5.55	3.92	0.37	-0.38
Movie (Dark) 2	2.15	2.87	0.64	-0.24
Sports	3.02	0.08	-0.73	-2.42

Note: The calculation is based on the CIECAM16.

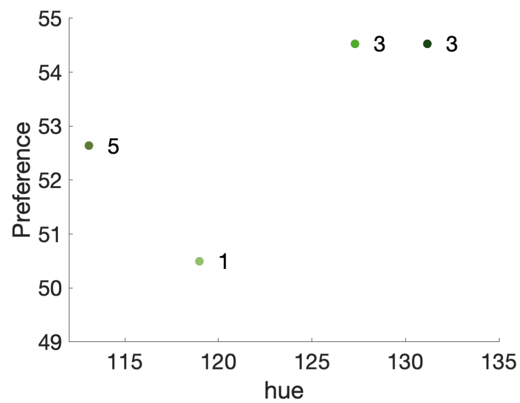


Figure 10: Preference per Hue

Note: Green represents the color of Animation, Movie (Dark), and Soccer content under the same display setting and lighting condition. The x-axis is the hue of different shades of green on the test display.

General Discussion

This project was designed to evaluate the influence of various factors on people's video quality preferences. Our findings suggested that TV display settings (Picture Mode), the intensity of ambient light, and the video content have significant effects on TV viewers' preferences.

In our experiment design, the illumination level of ambient light compared to that of the TV was generally lower (15–350 Lux). This could potentially diminish the influence of correlated color temperature (CCT) on the study subjects' responses.

We conducted appearance- and color-based image analyses per each experimental condition to get a deeper insight into video quality preference determinants. The analysis results indicate that chroma, lightness, and memory colors are essential in understanding people's preferences. In general, people prefer brighter and more chromatic images. Moreover, memory colors, such as grass and skin tones, significantly influence their perception and preferences of the visual quality of displayed content.

For display characterization, the gamma-offset-gain (GOG) model was applied, which is based on the principles of additivity and scalability. Under the same display setting, the assumption is valid. Additionally, the display uniformity was assessed by measuring a set of colors, specifically red, green, blue, and a randomly chosen color, lime. For each color, we conducted five measurements: one at the center and four at the corners of the display. The calculated $\Delta E2000$ values for these measurements were 1.27, 0.84, 0.56, and 0.89, respectively, with an average of 0.89. The uniformity of the display is within the satisfactory level. The GOG model conducted around the center area of the display could represent the whole display. Overall, the minor variations observed were unlikely to influence our results significantly.

The effectiveness of using genre to categorize video types needs to be examined. For instance, our results indicate a discernible difference in viewer preferences between animated and real-life scenes. This suggests that the content within these scenes is significantly independent from the genre itself. Another example is the presence of skin tones, which emerged as a critical feature in video content regardless of their genre. Observers' memory of the grass and the sunset, which in this case are from different genres, is the key to the preference – how the naturalness of these colors aligns with the viewer's memory.

Conclusion

We uncovered factors impacting human-perceived video quality and investigated the reasons why. We found that the intensity of the room illumination, the video content, and the display settings significantly impact TV viewer's picture quality perception and preferences. Our results also indicated that the conventional content genre categorization does not sufficiently capture the specific visual quality TV viewers may desire. People perceive images/videos differently based on the characteristics of each scene, even in the same content. Furthermore, we found that people generally prefer brighter and more chromatic videos in general. The naturalness of the objects' colors and memory colors, such as skin tones and grass, significantly impact their preferences.

Acknowledgements

This research has been supported by Samsung. The video clips used in this research were the Animation content from [20], Movie (Dark) and Movie (Bright) content from the American Society of Cinematographers [21], and Game and Sports content were purchased from Adobe Stock.

References

[1] YungKyung Park. Image Quality for TV Genre Depending on Viewers Experience. *Journal of Broadcast Engineering*, 26(3):308–320, 2021.

[2] Jan Kufa and Tomas Kratochvil. Visual Quality Assessment Considering Ultra HD, Full HD Resolution and Viewing Distance. In *2019 29th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pages 1–4, 2019.

[3] Jake Wetzel and Angie Hernandez. Effects of Display Media Type, Ambient Lighting, Viewing Distance, Viewing Offset, Object Type, and Object Size on Large-Screen Display Legibility. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 54, pages 2155–2159. SAGE Publications Sage CA: Los Angeles, CA, 2010.

[4] Andrei Chubarau, Tara Akhavan, Hyunjin Yoo, Rafal K Mantiuk, and James Clark. Perceptual image quality assessment for various viewing conditions and display systems. 2020.

[5] YeSeul Baek, Hyesun Han, Youngshin Kwak, Jae Sung Park, Junhee Woo, Jiman Kim, and Seongwoon Jung. The effect of ambient lighting on the preferred color temperature of television. *Electronic Imaging*, 2021(16):327–1, 2021.

[6] Oleg Gofaizen, Olena Osharovska, Mikola Patlayenko, Volodymyr Pyliavskyi, Svitlana Kiiko, and Valentyna Solodka. Image Sharpness Control in Modern Television Systems. In *2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T)*, pages 57–60. IEEE, 2020.

[7] Anush Krishna Moorthy, Lark Kwon Choi, G De Veciana, and A Bovik. Mobile video quality assessment database. In *IEEE ICC Workshop on Realizing Advanced Video Optimized Wireless Networks*, volume 6, pages 652–671, 2012.

[8] Stefan Winkler. Issues in vision modeling for perceptual video quality assessment. *Signal processing*, 78(2):231–252, 1999.

[9] Rajiv Soundararajan and Alan C Bovik. Video quality assessment by reduced reference spatio-temporal entropic differencing. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(4):684–694, 2012.

[10] Berta García Castiella et al. Factors to consider in improving the flat panel tv display user's experience on image quality terms. 2021.

[11] DaEun Park and YungKyung Park. The characteristic of image quality perception of elderly group: according to the comparison with adult group. *Journal of Information Display*, 19(3):111–120, 2018.

[12] Fei Gao, Yi Wang, Panpeng Li, Min Tan, Jun Yu, and Yani Zhu. Deepsim: Deep similarity for image quality assessment. *Neurocomputing*, 257:104–114, 2017.

[13] Guangtao Zhai and Xiongkuo Min. Perceptual image quality assessment: a survey. *Science China Information Sciences*, 63:1–52, 2020.

[14] Margaret H Pinson and Stephen Wolf. Comparing subjective video quality testing methodologies. In *Visual Communications and Image Processing 2003*, volume 5150, pages 573–582. SPIE, 2003.

[15] Scot R Fernandez, Mark D Fairchild, and Karen Braun. Analysis of observer and cultural variability while generating preferred color reproductions of pictorial images. *Journal of Imaging Science and Technology*, 49(1):96, 2005.

[16] Hosub Lee and Youngmin Park. Data-driven Approaches for Discovery and Prediction of User-preferred Picture Settings on Smart TVs. In *ACM International Conference on Interactive Media Experiences*, pages 134–143, New York, NY, USA, 2021. Association for Computing Machinery.

[17] Hosub Lee. Uncovering and Understanding Smart TV Users' Picture Quality Preferences via Big Data Analytics. In *2023 IEEE International Conference on Big Data (Big Data)*, pages 1764–1773, 2023.

[18] Chang Su, Chenguang Liu, and Pradeep Nagesh. Perceptual color enhancement based on properties of responses of human vision system to color stimulus, U.S. Patent 11,557,265, Jan. 2023.

[19] Mark D Fairchild. *Color appearance models*. John Wiley & Sons, 2013.

[20] Frank Kemper. The witch. `https://pixabay.com/videos/witch-magic-cartoon-fairy-child-88302/`.

[21] The American Society of Cinematographers. Standard Evaluation Material 2. `https://dpel.aswf.io/asc-stem2/`.

Author Biography

Eddie Pei received his BS in Engineering Management from Qingdao University of Technology. Project management certifi-

cate from Hubei University of Technology. His MS in Applied Statistics from Rochester Institute of Technology. He is currently a PhD student in color science at RIT. His research interests center around color appearance, machine learning, and color perception.

Hosub Lee is a research scientist at the Visual Solution Lab of Samsung Research America, Inc. He was also a research intern at Intel Corporation and a research staff member at Samsung Advanced Institute of Technology (SAIT). He received a bachelor's degree in computer science from Korea University in 2005, a master's degree in computer science and engineering from Seoul National University in 2007, and a Ph.D. in information and computer science from the University of California, Irvine in 2019. His research interests lie in user modeling, personalization, and human-AI interaction.

ELENA A FEDOROVSKAYA received a M.S. in psychology, Ph.D. degree in psychophysiology and a M.S. degree in Applied Mathematics, all from Lomonosov Moscow State University, Russia. She is currently a research professor at the Munsell Color Science Lab (MCSL), Rochester Institute of Technology. Prior to joining MCSL, Elena held a position of the Paul and Louise Endowed Chair at the School of Media Sciences, RIT, and worked at Eastman Kodak Research Laboratories. Her work focuses on understanding and modeling human perception of images and media, including multisensory integration in extended reality systems.

Susan Farnand is an Associate Professor and the Graduate Program Director of the Program of Color Science at the Rochester Institute of Technology (RIT). Her research interests center around human color vision and perception and include color imaging, individual differences in color vision, visual attention, image quality metrics, color 3D printing, and color reproduction in archival applications. Prior to joining RIT, Dr. Farnand was a senior research scientist at Eastman Kodak Company working primarily on projects in perceptual image quality measurement and modeling. She holds a BS in Engineering from Cornell University and an MS in Imaging Science and a PhD in Color Science from RIT. Dr. Farnand recently completed her term as President of the Society for Imaging Science and Technology.

Appendix

Table 5: Video Index Table (Experiment I)

Content\Lighting	Dark Warm	Bright Warm	Dark Cool	Bright Cool
Animation	1,21	6,26	11,31	16,36
Game	2,22	7,27	12,32	17,37
Movie (Dark)	3,23	8,28	13,33	18,38
Nature	4,24	9,29	14,34	19,39
Sports	5,25	10,30	15,35	20,40

Note: 1–20 were shown under Standard Picture Mode.
Note: 21-40 were shown under Movie Picture Mode.

Table 6: Video Index Table (Experiment II)

Content\Lighting	Dark Warm	Bright Warm	Dark Cool	Bright Cool
Animation	41,61,81	46,66,86	51,71,91	56,76,96
Game	42,62,82	47,67,87	52,72,92	57,77,97
Movie (Dark)	43,63,83	48,68,88	53,73,93	58,78,98
Nature	44,64,84	49,69,89	54,74,94	59,79,99
Sports	45,65,85	50,70,90	55,75,95	60,80,100

Note: 41–60 were shown under Vividness 1.
Note: 61–80 were shown under Vividness 2.
Note: 81–100 were shown under Vividness 3.

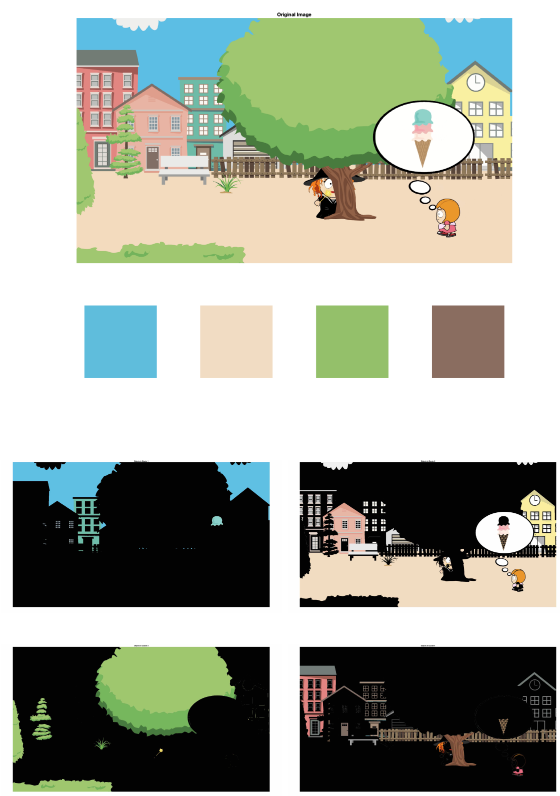


Figure 11: Color Segmentation Results (Animation Content)

Figure 12: Color Segmentation Results (Game Content)

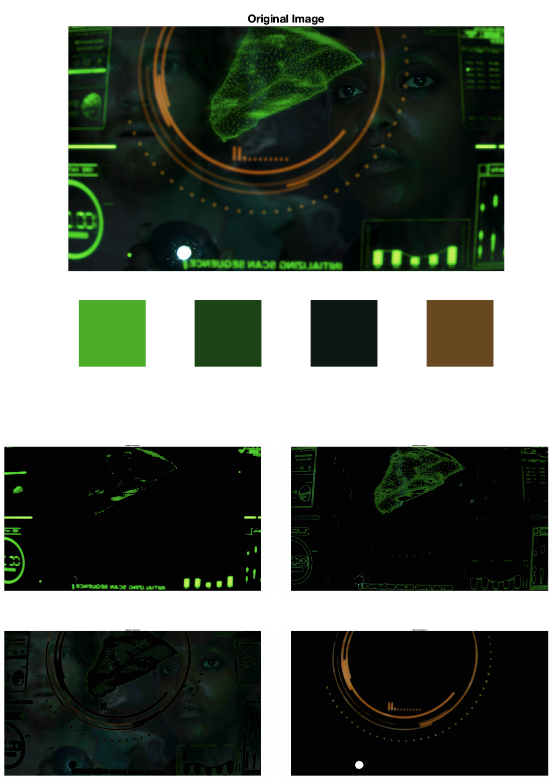


Figure 13: Color Segmentation Results (Movie (Dark) Content)

Figure 14: Color Segmentation Results (Nature Content)