QUALITY-OF-EXPERIENCE EVALUATION OF 8K ULTRA-HIGH-DEFINITION TELEVISION

Yoshiaki Shishikui

Meiji University, Tokyo, Japan

ABSTRACT

When evaluating extremely high-quality videos that reach maximum human perceptual limits, the conventional perceptual-level evaluation methods may seem insufficient. Hence, a paradigm shift in the evaluation methods is required. In this study, we propose a cognitive-level evaluation method for evaluating high-quality videos using impression strength as an index for the quality of experience (QoE) of an 8K ultra-high-definition television (UHDTV). Based on the evaluation scheme, we quantitatively demonstrated the difference in the quality of high-quality images close to the perceptual limit of human vision, and the results validated the hypothesis of viewers experiencing strong psychological effects when watching 8K videos. In addition, the strength of higher-order impressions such as freshness and deliciousness change with changes in the signal bandwidth and luminance level, indicating that the QoE of a UHDTV can be modulated using signal processing techniques.

Index Terms— QoE, subjective evaluation, 8K UHD TV, impression

1. INTRODUCTION

Given that visual mediums obtained from image and video signal processing systems are for people to experience and appreciate, a subjective evaluation is essential for the design, development, and improvement of processing systems. Due to the inherent complexities of subjective evaluation methods, researchers have alternatively proposed objective evaluation methods [1, 2]. For instance, the subjective assessment scores of perceptual degradation and the objective evaluation values, such as PSNR, are widely used to evaluate the performance of compression coding schemes.

Because people can perceive the differences between objects in high-definition images and real life, most image-quality-evaluation techniques focus on the overall fidelity, degradation, and preferences [3–7], which commensurate with the quality of conventional visual systems. Hence, the evaluation scores of the conventional visual systems based on the fidelity criteria do not saturate, indicating adequate sufficiency for comparing the performances of the systems.

This research was partially supported by Meiji University under the Research Project Grant (B) by Institute of Science and Technology

The ultra-high-definition (UHD) TV [8] is an immersive technology that shrinks the line between the real and hyperreal and is rapidly gaining attention worldwide. Almost a decade ago, researchers determined the parameters for displaying 2D images with the “ultimate” quality on 8K-UHD systems based on the psychophysical evidence of the perceptual limits of human vision [9]. In particular, the number of pixels was determined by conducting subjective evaluation experiments on the sense of “being there” [10–12]. As a result, many viewers watching an 8K UHDTV video expressed experiencing a strong psychological effect [13, 14]. However, there is little quantitative evidence that confirms the psychological impact on viewers watching 8K content. Considering the 8K-UHDTV video system is designed based on the perceptual limits of human vision, the conventional metrics, i.e., perceptual fidelity-based evaluation, would be saturated. Therefore, to evaluate the difference in the quality of images near the perceptual limit, a suitable metric is required.

Considering that viewers do not appraise the quality of test charts in a specified viewing condition in daily life and enjoy the immersive experiences from a TV screen, the quality of the viewing experience should be measured directly with more natural stimuli. Therefore, in this study, we propose a method that evaluates the quality of experience (QoE) when viewing an 8K-UHDTV using the strength of common impressions generated from familiar images.

In a previous study [15], the psychological effects of viewing 4K-UHDTV images were evaluated. Images of varying resolutions (4K/2K/1K) were presented to subjects to elicit their ratings of higher- and lower-order impressions (Table 1). We further utilized the proposed methodology to evaluate the QoE of an 8K-UHDTV. The results proved that the proposed cognitive-level evaluation scheme was effective in evaluating ultra-high-quality images and verifying the value of 8K resolution based on the new metrics.

We also examined the effect of the physical parameters such as signal bandwidth and luminance level on the QoE, considering previous studies only evaluated the effect of physical parameters on the overall preference [16–19]. The results showed that the 8K resolution generated an enhanced impression as compared to 4K, and the luminance affects the strength of certain impressions.

2. EXPERIMENTS

We conducted subjective evaluation experiments in which subjects rated their impressions on test images of different resolutions and display luminance levels. Table 1 lists the experimental conditions. We set the viewing distance to 1.5 H (H: picture height), which is the design viewing distance of 4K [20]. The room conditions conform to the laboratory environment of ITU-R BT.500 [3, 21].

2.1. Test images

The test images were images of food and outdoor scenarios, as shown in Fig. 1. We selected test-image subjects that would elicit viewer impressions, such as "beautiful" and "delicious."

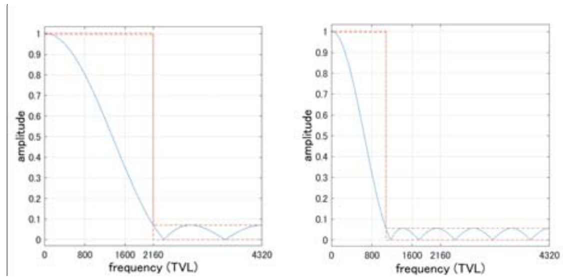
The images were captured on a camera (Canon EOS 5DS with 8,688 × 5,792 pixels) and recorded in RAW format. The images were developed in 10 bits per pixel using the gamma and color gamut method of ITU-R BT.2020 and trimmed to the 8K format (7,680 × 4,320 pixels).

Table 1. Experimental Conditions

Test images	Cake, fruit, nikuman, sashimi, wheel, lady, hanabi, dog
Resolution	8K, 4K-equivalent, 2K-equivalent
Display and viewing distance	Canon prototype (55-inch, 8,192 × 4,320 pixels) Peak luminance level: 100/300/500 cd/m^2 1.5 H (96 cm) (H: picture height)
Evaluation terms	Lower-order impression: three-dimensional, high-resolution, vivid, glossy Higher-order impression: beautiful, delicious, fresh, real, being-there

(a) Cake (b) Fruit
(c) Nikuman (d) Sashimi
(e) Wheel (f) Lady
(g) Hanabi (h) Dog

Fig. 1. Test images



(a) 4K equivalent (b) 2K equivalent

Fig. 2. Frequency and impulse responses of the LPF

From the 8K images, we developed 4K- and 2K-equivalent images using a low-pass filter (LPF) that mimics the modulation transfer function (MTF) characteristics of 4K and 2K cameras, respectively. The LPF to generate 2K-equivalent images were specifically designed as per the MTF characteristics of an actual 2K camera reported in [22]. The original 8K images along with the 4K- and 2K-equivalent images were displayed on an 8K display. Fig. 2 shows the MTF of the LPF used.

2.2. Experimental procedure

The trial series was, conducted with 18 subjects aged between 20–22 years, with normal or corrected-to-normal visual acuity tested using the Tumbling E eye chart. Informed consent was obtained from the subjects before conducting the experiment.

During a trial, images with eight content types and three bandwidths were randomly presented to the subjects to ensure the ratings were not affected by the order in which the images were presented. During the 30-second presentation, the subjects were asked to view the images and fill out a questionnaire based on nine impressions (evaluation terms are listed in Table 1) on a seven-grade scale ranging from "strongly disagree" to "strongly agree." However, "being there" was excluded from the food image, whereas "fresh" and "delicious" were excluded from the outdoor scene images.

We investigated the extent to which the common impressions in daily life are altered by the physical factors. We selected the following evaluation terms for expressing popular values: "looks beautiful," "looks delicious," "looks fresh," "looks real," and "being there," which are evaluation terms of higher-order impressions obtained by combining various image features. The evaluation terms used for lower-order impressions were "high resolution," "color vividity," "glossiness," and "three-dimensional (3D)."

2.3. Analysis

We calculated the average values of the evaluation results and their 95% confidence intervals based on the t-distribution. Furthermore, a t-test was conducted to determine any significant differences in the impressions between the resolutions. Moreover, we conducted multiple t-tests to compare the impression scores between the adjacent resolutions (i.e., between 8K and 4K and between 4K and 2K), for each image content. To avoid multiple comparison problems, we adjusted the false discovery rates (FDRs) for each p-value using the Benjamini–Hochberg method [23].

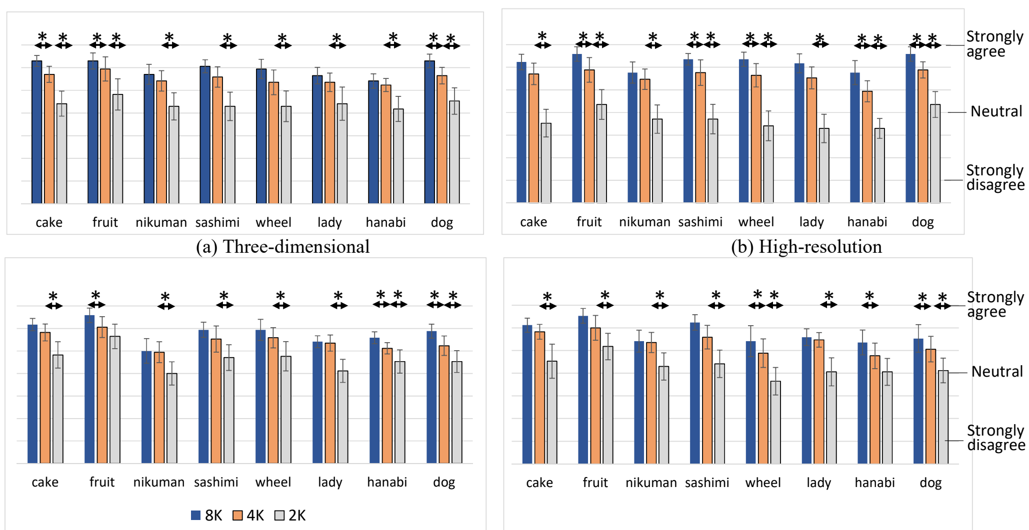
The strength of an impression may vary depending on the image content and display luminance. To examine the effect of these factors, we performed a two-way analysis of variance (ANOVA2) on the scores of the 8K images and further conducted multiple comparison tests.

3. RESULTS

3.1. Impression strength for different image resolutions

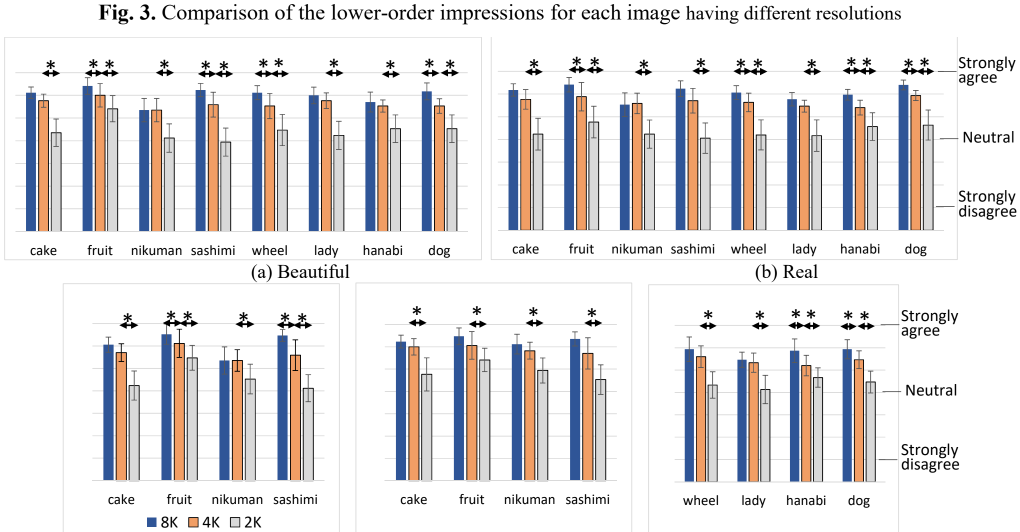
Fig. 3 summarizes the results of the lower-order impressions for the 8K, 4K, and 2K images with a display peak luminance of 300 cd/m^2. The asterisk (*) indicates a significant difference between the images with adjacent resolutions ($p < 0.05$, FDR corrected.) Fig. 4 displays the results of the higher-order impressions. The results demonstrate significant differences between the 8K, 4K, and 2K images in many cases among the combinations of impression terms and image content.

The followings were derived from Figs. 3 and 4:
- Significant differences were observed between the 8K, 4K,



(a) Three-dimensional (b) High-resolution (c) Vivid (d) Glossy

Fig. 3. Comparison of the lower-order impressions for each image having different resolutions



(a) Beautiful (b) Real (c) Fresh (d) Delicious (e) Being-there

Fig. 4. Comparisons of the higher-order impressions for each image having different resolutions

Table 2. Results of ANOVA2

*: $p < 0.05$ **: $p < 0.01$

	Luminance	Content	Interaction
Three-dimensional		**	
Resolution	**	**	
Vivid		**	
Glossy		**	
Beautiful		**	
Real		**	
Fresh	**	**	
Delicious	*	**	
Being there		**	

and 2K images for lower- and higher-order impressions.

- The differences were image-dependent and were greater in the "fruit" and "hanabi" images and smaller in the "nikuman" and "lady" images.
- The average impression scores for the 8K images reached the "agree" level.

3.2. Effect of the peak display luminance

We used the ANOVA2 process to determine whether the significant differences were due to the display luminance, image content, or the interaction between them for each image and impression term. Table 2 summarizes the results, where "*" indicates $p < 0.05$ and "**" indicates $p < 0.01$. The p-values were corrected for multiple comparisons.

The effect of image content was significant for all impressions, whereas the display luminance was significant only for the "resolution," "fresh," and "delicious," as shown in Table 2. No interaction was observed between the two factors. Fig. 5 shows the results of multiple comparison tests conducted on the effect of luminance for the "resolution," "fresh," and "delicious." The results indicated that "high resolution" was significantly lower for 500 cd/m^2 as compared to 100 and 300 cd/m^2, whereas "fresh" and "delicious" were significantly lower for 100 cd/m^2 as compared to the rest.

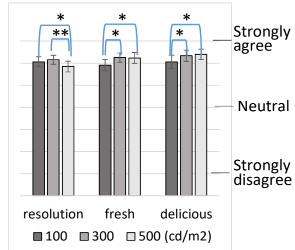


Fig. 5. Results of multiple comparison tests for analyzing the effect of luminance *: $p < 0.05$ **: $p < 0.01$

4. DISCUSSION

The viewing distance in the experiment was 1.5 H. The angular resolution of 4K from the viewing distance corresponded to the separable visual angle of a subject with normal vision. Therefore, the difference between 8K and 4K is at or above the perceptual limits. The difference between impression strengths evoked by 4K and 8K was shown using the proposed cognitive-level evaluation scheme, in which viewers rated images based on their higher-order impressions. Furthermore, 8K evoked multiple strong impressions and quantitatively substantiated frequently expressed opinions from the viewers on experiencing overwhelming sensations from the images. We also confirmed that the proposed evaluation method could provide reasonable QoE of high-quality images.

The comparison results of the multiple resolutions confirmed that both lower- and higher-order impressions were enhanced by improving the resolution. While physical changes in the resolution generally trigger the sense of resolution alone, in this experiment, it also affected the strength of the other lower-order impressions, indicating that the impressions were generated by a complex hierarchical process as per the functionality of the brain, and not from a simple process of visual features. Concerning the effect of display luminance level, while the sense of high resolution was slightly reduced in a fairly bright image, the senses of freshness and deliciousness increased with an increase in the luminance, as shown in Fig. 5. This result is consistent with intuition. These results confirm that the strength of specific impressions can be modulated using physical parameters, such as signal bandwidth and intensity, which further suggests the possibility of manipulating impressions using improved signal processing techniques.

Yet another challenge to be considered in future research is to derive an objective evaluation value that mimics the subjective evaluation results from the image signal. Regarding the perceptual-level evaluation, while several objective metrics were proposed in various literatures, such as PSNR, SSIM [24], VIF [25], and NIQE [26], conventional schemes such as simple frequency component analysis are insufficient and required multiple analytical studies to derive impression strength owing to the complex processing ability of the brain, as described above.

5. CONCLUSION

In this study, we proposed a cognitive-level assessment method to evaluate the QoE of ultra-high-quality images near human perceptual limits, which can replace perceptual-level methods. The results of this experiment confirmed the validity of the evaluation method. Furthermore, this method also demonstrated the value of 8K-UHDTV and the possibility of impression manipulation using signal processing.

For future research, we aim to model the QoE by conducting evaluation experiments by varying the physical factors such as signal conditions (e.g., dynamic range) and viewing conditions (e.g., lighting environment) to analyze their causal relationships with the physical factors and evoked impressions. We believe the outcomes would contribute to the optimization of signal processing in ultra-high-quality images, such as QoE-aware editing and compression.

6. REFERENCES

[1] M. Nadenau, S. M. Winkler, D. Alleysson and M. Kunt, "Human Vision Models for Perceptually Optimized Image Processing – A Review," Proc. of IEEE, 2000.

[2] G. Zhai and X. Min, "Perceptual image quality assessment: a survey," Sci. China Inf. Sci. 63, 211301, 2020.

[3] Recommendation ITU-R BT.500-13. "Methodology for the subjective assessment of the quality of television pictures," 2012.

[4] Recommendation ITU-T P.910, "Subjective video quality assessment methods for multimedia applications," 2008.

[5] Recommendation ITU-T P.911, "Subjective audiovisual quality assessment methods for multimedia applications," 1998.

[6] J. You, U. Reiter, M. M. Hannuksela, M. Gabbou, and A. Perkis, "Perceptual-based quality assessment for audio–visual services: A survey," Signal Processing: Image Communication, vol. 25, issue 7, pp. 482–501, 2010.

[7] W. Lu, X. Li, X. Gao, W. Tang, J. Li, and D. Tao, "A video quality assessment metric based on human visual system," Cogn. Comut 2, pp. 120–131, 2010.

[8] Recommendation ITU-R BT.2020. "Parameter values for ultrahigh definition television systems for production and international programme exchange," 2012.

[9] M. Sugawara, M. Emoto, K. Masaoka, Y. Nishida, and Y. Shishikui, "SUPER Hi-VISION for the next generation television," ITE Transactions on MTA, vol. 1, no. 1, pp. 27–33, 2013.

[10] T. Hatada, H. Sakata, and H. Kusaka, "Psychological Analysis of the 'Sensation of Reality' Induced by a Visual Wide-Field Display," SMPTE Motion Imaging Journal, vol. 89, no. 8, pp. 560–569, Aug. 1980.

[11] M. Emoto, K. Masaoka, M. Sugawara, and F. Okano, "Viewing angle effects from wide field video projection images on the human equilibrium," Displays, vol. 26, no. 1, pp. 9–14, Jan. 2005.

[12] K. Masaoka, M. Emoto, M. Sugawara, and Y. Nojiri, "Contrast effect in evaluating the sense of presence for wide displays," Journal of the Society for Information Display, vol. 14, no. 9, pp. 785–791, Sep. 2006.

[13] Y. Shishikui, Y. Fujita, and K. Kubota, "Super Hi-Vision —the star of the show!," EBU Technical Review, 2009-Q0, 2009, p. 1–13.

[14] M. Sugawara, S. Sawada, H. Fujinuma, Y. Shishikui, J. Zubrzycki, R. Weerakkody, and A. Quested, "Super Hi-Vision at the London 2012 Olympics," SMPTE Motion Imaging Journal, vol. 122, no. 1, p. 29–38, 2013.

[15] Y. Shishikui, and Y. Sawahata: "Effects of Viewing Ultra-High-Resolution Images With Practical Viewing Distances on Familiar Impressions," IEEE Transactions on Broadcasting, vol. 64, no. 2, June 2018, pp. 498–507, 2018.

[16] P. S. Guterman, K. Fukuda, L. M. Wilcox, and R. S. Allison, "Is Brighter Always Better? The Effects of Display and Ambient Luminance on Preferences for Digital Signage," SID Symp. Dig. Tech. Papers, vol. 41, no. 1, pp. 1116–1119, 2010.

[17] J. Kufa, and T Kratochvil, "Visual Quality Assessment Considering Ultra HD, Full HD Resolution and Viewing Distance," 29th International Conference Radioelektronika (RADIOELEKTRONIKA), 2019.

[18] P. Hanhart, P. Korshunov, and T. Ebrahimi, "Subjective Evaluation of Higher Dynamic Range Video," Proceedings of SPIE 9217, Applications of Digital Image Processing XXXVII, 2014.

[19] Y. Shi, Y. Tu, L. Wang, Y. Zhang, and X. Gao, "Influence of Display Luminance on Visual Discomfort in Dark Ambient Based on Haemodynamic Response," IEEE Photonics Journal pp. (99):1-1, July 2020.

[20] Recommendation ITU-R BT.2022, "General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays," 2012.

[21] R. Sitelo, J. Joskowics, M. Anedda, M. Murroni, and D. Giusto: "Subjective video quality assessments for 4K UHDTV," Proc. of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1–6, 2017.

[22] T. Nakamura, Y. Iwasaki, S. Tajima, K. Masaoka, T. Yamashita, and Y. Nishida: "MTF Measurement of a Super Hi-Vision Camera," ITE Winter Annual Convention, pp. 11–18, 2013.

[23] Y. Benjamini, and Y. Hochberg: "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," Journal of the Royal Statistical Society: Series B (Methodological), vol. 57, no. 1, pp. 298–300, Jan. 1995.

[24] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600-612, Apr. 2004.

[25] H. R. Sheikh, and A. C. Bovik, "Image Information and Visual Quality," IEEE Transactions on Image Processing, vol. 15, no. 2, Feb. 2006.

[26] A. Mittal, R. Soundararajan, and A.C. Bovik, "Making a "completely blind" image quality analyzer," IEEE Signal Process. Letters, vol. 20, no. 3, pp. 209-212, Mar. 2013.