



Color mismatch and observer metamerism between conventional liquid crystal displays and organic light emitting diode displays, Part II: Adjacent stimuli with a larger field of view

JIALU WU AND MINCHEN WEI* 

Color and Illumination Laboratory, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China

*minchen.wei@polyu.edu.hk

Abstract: In part I of this study [Opt. Express 29, 12292 (2021)], 50 human observers matched the color appearance of six color stimuli produced by four smartphone displays, including one conventional liquid crystal display (LCD) and three organic light emitting diode (OLED) displays, to those produced by a reference smartphone OLED display. The matching and reference stimuli had a field of view (FOV) around 4.77° and were 5.72° apart. In this experiment, we carefully designed and built a new apparatus to make the two stimuli adjacent to each other with an FOV around 20.2° . This not only made the viewing condition in the experiment similar to the typical viewing condition of smartphone displays, but also allowed for an easier color matching, resulting in smaller intra- and inter-observer variations. The performance of the four CMFs, however, were not significantly changed with the increase of the FOV. The CIE 2006 2° CMFs still had the best performance in characterizing the color matches, which did not support the recommendation of using 10° CMFs for stimuli with an FOV beyond 4° . Meanwhile, for the pairs of stimuli with matched color appearance, the LCD display always had the greatest chromaticity differences and degrees of observer metamerism among the four displays, regardless of the CMFs. In particular, the chromaticities of the stimuli produced by the LCD display were always shifted towards the $-u' + v'$ direction in the CIE 1976 $u'v'$ chromaticity diagram, when calculated using the CIE 1931 CMFs. This implies that the neutral colors shown on LCD displays would have a yellow-green tint on OLED displays, if they were calibrated to the same chromaticities using the CIE 1931 2° CMFs.

© 2021 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Color matching functions (CMFs) play a very important role in color science and management and are widely used to derive various colorimetric metrics, such as chromaticities, correlated color temperature (CCT), color differences, color rendition metrics, and color appearance attributes [1]. Industries heavily rely on various metrics derived using CMFs for color specification, quality control, and algorithm development. For example, the specification of the chromaticities is critically important to the lighting and display industries, so that the color consistency can be guaranteed among individual products and also among different product series.

Therefore, substantial efforts have been made to derive CMFs, so that pairs of color stimuli with the same chromaticities have matched color appearance. Two color matching experiments were carried out between 1929 and 1931 by Wright [2,3] and Guild [4] and led to a set of CMFs, which was proposed by the International Commission on Illumination (CIE) in 1931. It is known as the CIE 1931 2° CMFs and is the most widely used set in all the color-related industries and applications. This set of CMFs, however, is mainly for characterizing the matches of color stimuli with a field of view (FOV) around 2° , since the color matching experiments carried out by Wright and Guild used a 2° bipartite setup. In order to characterize color matches for common viewing conditions, Stiles and Burch [5] and Speranskaya [6] carried out color matching experiments

#444807

Journal © 2021

<https://doi.org/10.1364/OE.444807>

Received 30 Sep 2021; revised 1 Nov 2021; accepted 15 Nov 2021; published 30 Nov 2021

using a 10° bipartite setup, which resulted in a new set of CMFs proposed by CIE in 1964—CIE 1964 10° CMFs. In particular, the CIE 1931 2° CMFs are recommended for characterizing stimuli with an FOV between 1° and 4° , while the CIE 1964 10° CMFs are recommended for characterizing stimuli with an FOV beyond 4° [1]. The above two CMFs were developed purely based on the psychophysical color matching experiments. In 1991, a CIE technical committee (i.e., TC1-36) was established with a goal to “establish a fundamental chromaticity diagram of which the coordinates correspond to physiological significant axes”. The committee developed a model to characterize the cone fundamentals for normal observers with an FOV from 1° to 10° in 2006 [7]. The model was further extended to derive the corresponding CMFs and chromaticity diagrams, with the CMFs for the 2° and 10° directly provided for easy use in practice, in 2015 [8]. In this article, these two CMFs are referred as the CIE 2006 2° and 10° respectively. These four CMFs are the standard CMFs, though other sets or models [9,10] have been developed and proposed.

Meanwhile, researchers and scientists also carried out psychophysical experiments to compare and verify the performance of these standard CMFs [11–24]. In these experiments, the spectral composition of the color stimuli, in terms of spectral shape (i.e., narrow-band versus broad-band) and peak wavelength, was commonly varied as the independent variable, since it has been found to significantly affect the performance of CMFs and an optimal set of CMFs should work well for any spectral composition. The color stimuli with different spectral compositions were produced using various types of facilities, such as cathode-ray tube (CRT) displays, liquid crystal displays (LCD), organic light emitting diode (OLED) displays, light emitting diode (LED) lamps, projectors, broadband lamps with filters, and illuminated color samples [11–24]. The observers adjusted the color appearance of one stimulus to match that of the reference stimulus, and then the chromaticities of the two stimuli were used to compare the performance of different CMFs. The findings about the performance of the 2° and 10° CMFs, however, were not consistent. For example, though the 10° CMFs are recommended for stimuli with an FOV greater than 4° , part I of our study found that the 2° CMFs had better performance than the 10° CMFs [24]; Li et al. found the 10° CMFs had better performance [17]. More importantly, no past study held the spectral compositions to be fixed while varied the FOV as an independent variable, which makes it difficult to attribute the poor performance of CMFs to the FOV.

In part I of the study [24], we investigated the color mismatch and observer metamerism between conventional LCD and OLED displays. Fifty human observers completed color matching tasks for two 4.77° color stimuli, which were 5.72° apart, produced by displays. The CIE 2006 2° CMFs were found to have better performance than the other three CMFs, which did not support the recommendation made by the CIE. In order to further investigate whether such a result was due to the relatively small FOV, we increased the FOV of the stimuli to 20.2° in this experiment, which was also similar to the typical viewing condition when viewing smartphone displays in daily life. Moreover, the two stimuli in this experiment were placed adjacent to each other, allowing for an easy match of the color appearance, which did not happen in part I of the study and in most past studies.

2. Methods

The experiment was carried out in Color and Illumination Laboratory at The Hong Kong Polytechnic University, with the Institutional Review Board approving the experiment protocols and procedures.

2.1. Apparatus

A black metal viewing box, with dimensions of 50 cm (width) \times 28.5 cm (depth) \times 28.5 cm (height), was designed and built for this experiment. Two 5.5 cm \times 10 cm openings were cut on the ceiling and back panels. A holder was designed to fix the smartphone at each opening, with

the center area of the smartphone display appearing at the opening to produce a color stimulus. A $10\text{ cm} \times 10\text{ cm}$ opening was cut on the front panel as the viewing window. A mirror with a tilt angle of 45° was carefully placed and fixed inside the box, so that the two stimuli produced by the two displays can be viewed adjacent to each other simultaneously through the viewing window. Specifically, the top stimulus was produced by the display placed above the ceiling panel, which was viewed through the reflection of the mirror; the bottom stimulus was produced by the display placed behind the back panel, which was viewed directly. Each stimulus had a size of $5\text{ cm} \times 5\text{ cm}$ and a sharp dividing edge was between the two stimuli, which was carefully designed to achieve a higher accuracy of the color matching, in comparison to part I of this study [24]. It is worthwhile to mention that the mirror was carefully manufactured to have a flat spectral reflectance distribution across the visible spectrum, with an average reflectance factor around 93%.

During the experiment, the viewing box was placed in front of the observer, with the front panel 28 cm away from the observer's eye position. The observer fixed his or her chin on a rest, but can change the height of the chair, so that the two stimuli appeared to have the same size, with an FOV around 20.2° , to the observer. Figure 1 shows the design of the viewing box, the experiment setup, and the stimuli viewed by the observer during the experiment.

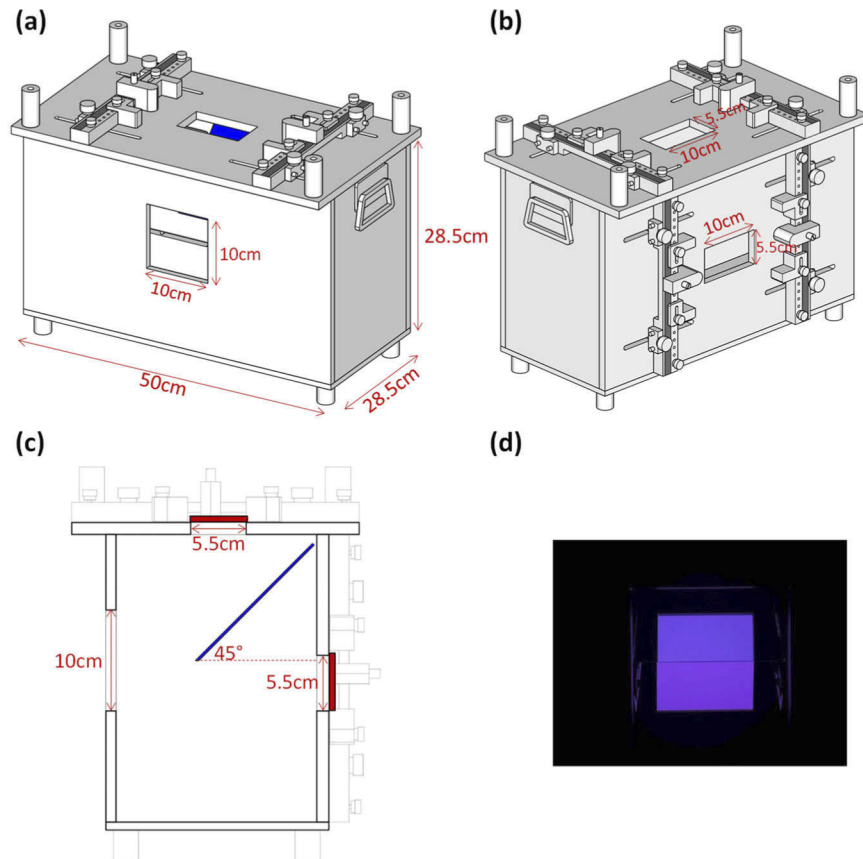


Fig. 1. Apparatus and experiment setup. (a) Perspective front view of the apparatus; (b) Perspective back view of the apparatus; (c) Cross section view of the apparatus and the positions of the two smart phones; (d) Photograph of the two stimuli viewed from the observer's eye position. (Note: the mirror is shown in blue in (a) and (c)).

2.2. Display, color stimuli, display calibration, control program, and experimental procedure

The five smart phone displays used in part I of the study [24], including one LCD display and four OLED displays, were used to produce the color stimuli. The OLED displays were carefully selected from eight OLED displays, so that they had differences in their primaries, in terms of the peak wavelengths and the shape of the spectral power distributions (SPDs), and noticeable perceived color differences. The OLED display with the smallest color gamut was selected as the reference display, so that the comparison can be made between the LCD and OLED displays and also among the OLED displays. The gamut and the SPDs of the primaries used in these five displays are shown in Figs. 1 and 2 in part I of the study [24].

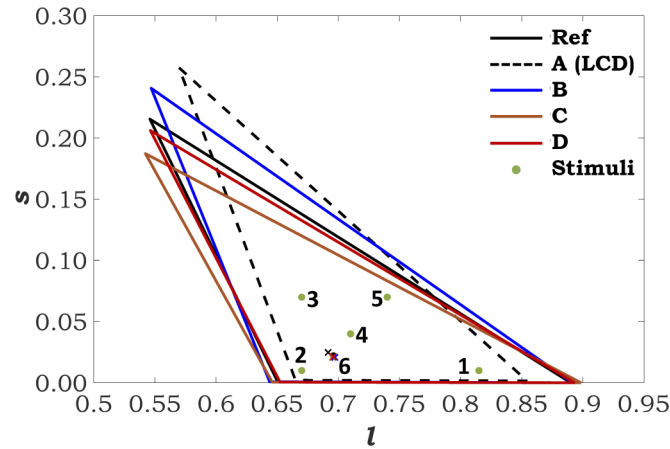


Fig. 2. Color gamuts and white points of the displays in the MacLeod-Boynton chromaticity diagram for 2° FOV and the chromaticities of the reference stimuli shown on the reference display for color matching. (Note: the numbers shown on the figure represent Stimuli 1 to 6, the crosses are the white points of the displays).

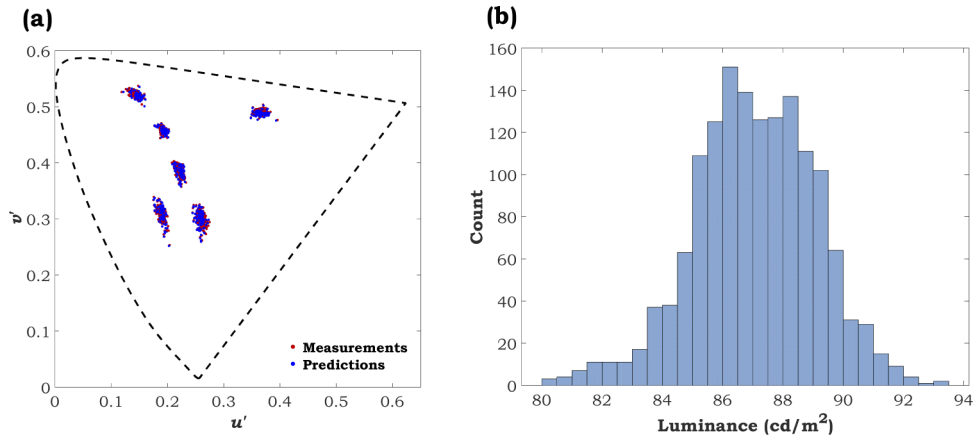


Fig. 3. Accuracy of the control program, in terms of chromaticities and luminance. (a) Chromaticities of the adjusted stimuli derived from the measured SPDs and the predictions using the control program in the CIE 1976 $u'v'$ chromaticity diagram; (b) Histogram of luminance of the adjusted stimuli derived from the measured SPDs, with an average of 87.06 cd/m^2 .

During the experiment, the reference display was fixed behind the back panel to produce the bottom stimulus, while one of the four test displays was placed above the ceiling panel to produce the top stimulus, which was purposely designed to allow the experimenter to easily change the display placed above the ceiling panel during the experiment. Since the mirror had a flat spectral reflectance distribution across the visible spectrum, the reflection was expected to reduce the luminance but maintain the chromaticities. This was verified by measuring the SPDs of the 20 stimuli produced by each of the four displays with and without the mirror. Therefore, we changed

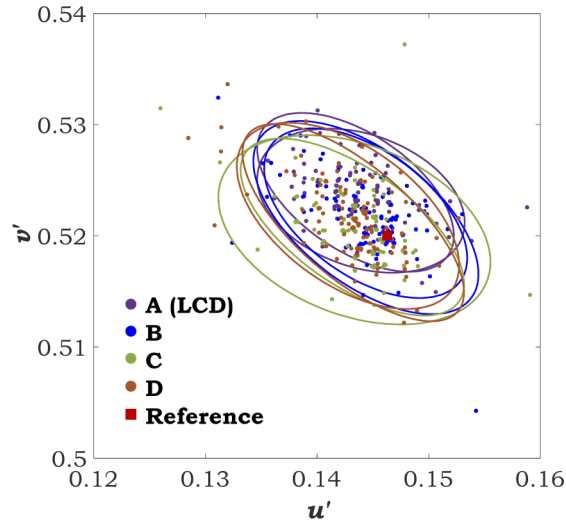


Fig. 4. Chromaticities, together with the 95% confidence error ellipses, of the repeated matches performed by each observer for Stimulus 2 using the four test displays.

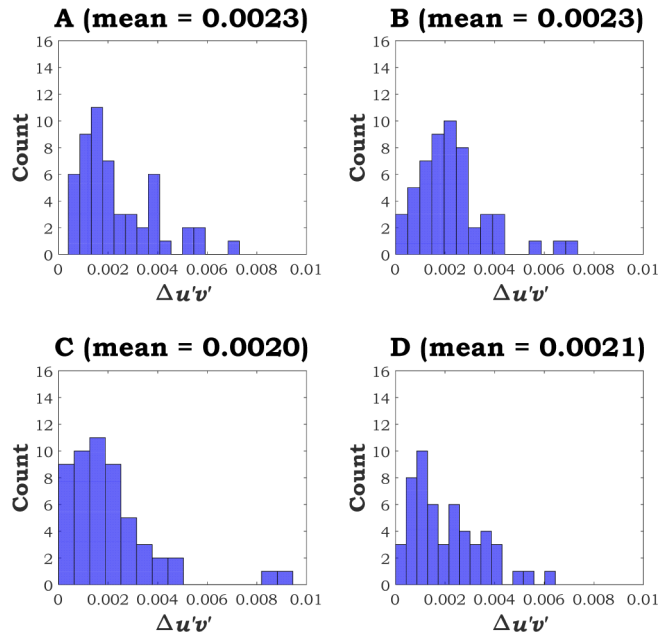


Fig. 5. Histogram of the MCDM values of the intra-observer variations for each test display.

the luminance level of the stimuli to 88 cd/m² from 93 cd/m² in part I [24], so the look-up-tables (LUTs) and control program can be directly used in this experiment. In contrast, the six stimuli shown on the reference display were recalibrated to have a luminance of 88 cd/m² using the spectroradiometer. Figure 2 shows the chromaticities of the six reference stimuli and the color gamuts of the displays in the MacLeod-Boynton chromaticity diagram for 2° FOV. Figure 3 can show the rough locations of the chromaticities of the six stimuli in the CIE 1976 $u'v'$ diagram.

The experiment procedure was identical to part I [24] of the study. Briefly speaking, each observer adjusted the color appearance of seven stimuli (Stimuli 1 to 6 with a repeated adjustment

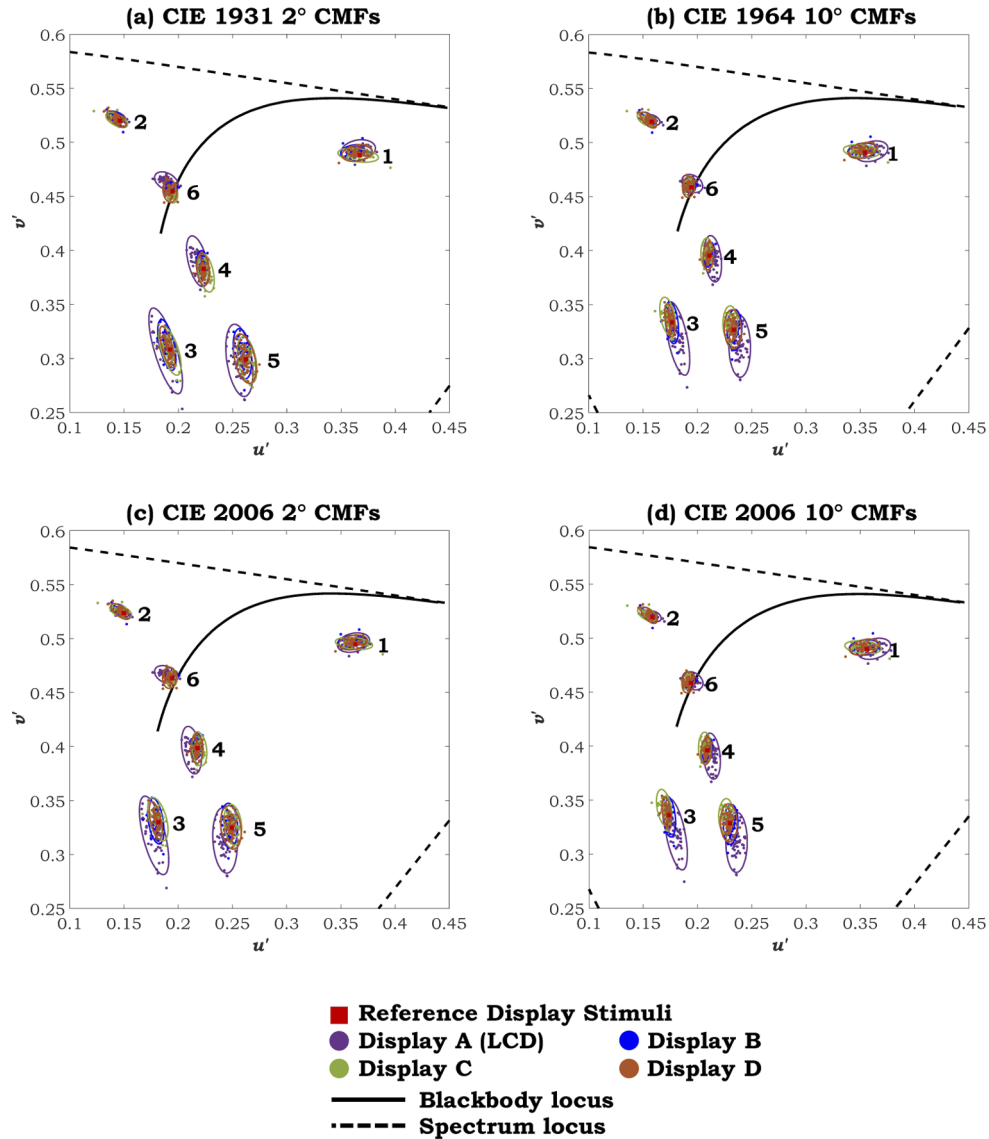


Fig. 6. Chromaticities of the stimuli adjusted by the 53 observers using the four displays to match the color appearance of the reference stimuli shown on the reference display, together with the fitted 95% confidence error ellipses, calculated using the four CMFs. (a) CIE 1931 2° CMFs; (b) CIE 1964 10°; (c) CIE 2006 2° CMFs; (d) CIE 2006 10° CMFs. (Note: the numbers shown on the figures represent Stimuli 1 to 6).

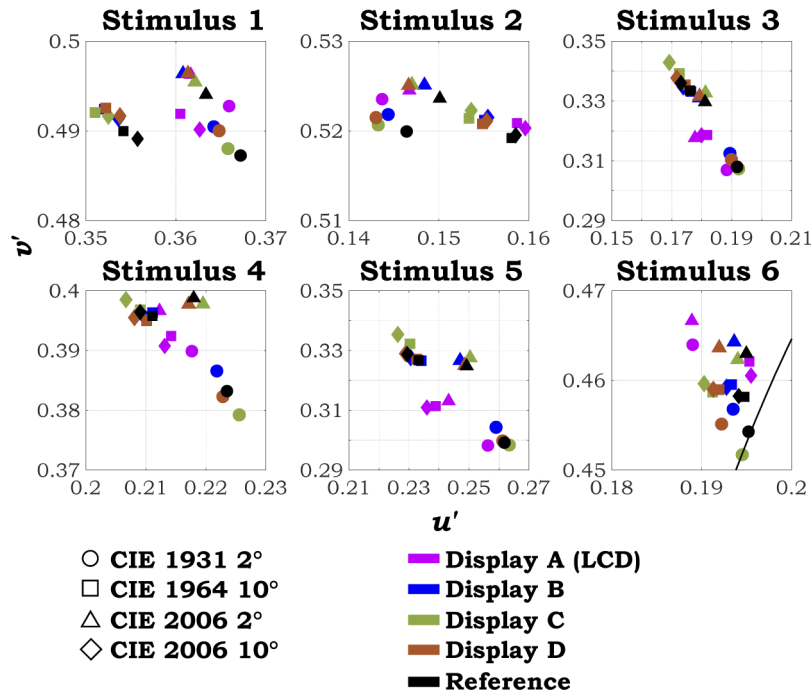


Fig. 7. Average chromaticities of the stimuli adjusted by the 53 observers using the four displays and the corresponding reference stimuli shown on the reference display calculated using the four CMFs. (Note: for illustration purposes, the axis ranges are different, but each grid corresponds to 0.01 unit in u' and v' , the blackbody locus shown with Stimulus 6 was plotted using the CIE 1931 2° CMFs).

Table 1. Average difference $\Delta u' v'$ between the average chromaticities adjusted by the observers of the six color stimuli using each display and the chromaticities of the reference stimuli shown on the reference display calculated using the four CMFs.

	Display A	Display B	Display C	Display D	Average
CIE 1931 2°	0.00660	0.00403	0.00230	0.00243	0.00384
CIE 1964 10°	0.00802	0.00159	0.00461	0.00229	0.00413
CIE 2006 2°	0.00738	0.00241	0.00222	0.00234	0.00359
CIE 2006 10°	0.00913	0.00172	0.00537	0.00226	0.00462

of Stimulus 2 for evaluating the intra-observer variations), which were produced by each test display and appeared at the top, to match the color appearance of the corresponding reference stimuli that were produced by the reference display and appeared at the bottom. The order of the seven stimuli was randomized and the order of the four test displays was also randomized.

2.3. Observers

Fifty-three observers (28 males and 25 females) between 19 and 38 years of age (mean = 24.1, std. dev. = 3.28) were recruited for the experiment. All the observers had normal color vision, as tested using the Ishihara Color Vision Test. The observers were recruited from the same observer pool as part I of the study [24], but not all the observers participated in both experiments.

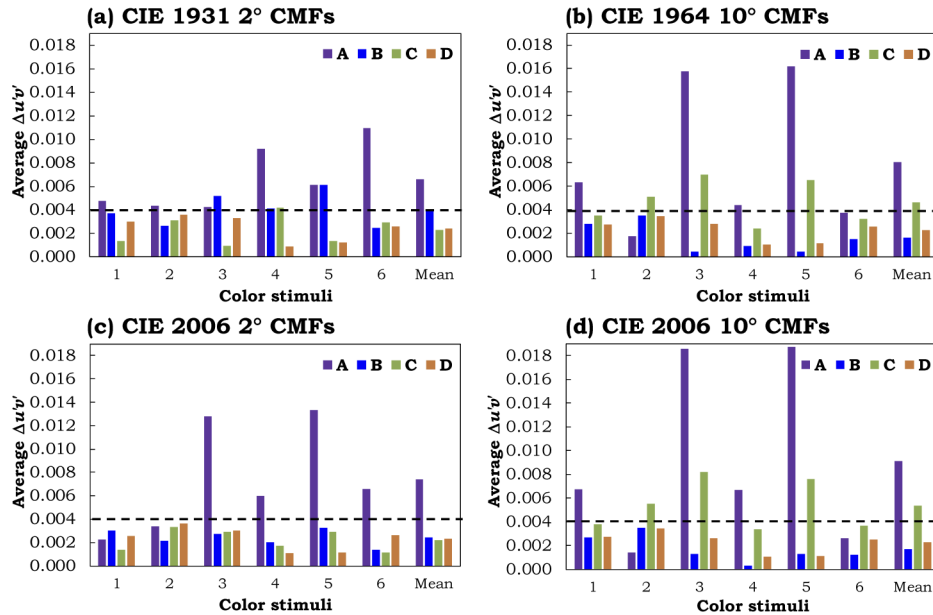


Fig. 8. Chromaticity differences $\Delta u'v'$ between the average chromaticities of the stimuli adjusted by the 53 observers and the chromaticities of the reference stimuli shown on the reference display calculated using the four CMFs. (a) CIE 1931 2° CMFs; (b) CIE 1964 10° CMFs; (c) CIE 2006 2° CMFs; (d) CIE 2006 10° CMFs. (Note: the dash line labels 1 just-noticeable color difference, 1 JND. However, it was developed only based on the CIE 1931 2° CMFs and may not be applicable to the other three CMFs).

3. Results and discussions

3.1. Verification of the control program accuracy

Though the control program allowed the observer to change the color appearance of the stimulus by changing its chromaticities along the u' and v' axes in the CIE 1976 $u'v'$ chromaticity diagram, it actually changed the RGB digital values through the LUTs. After the experiment, the RGB combinations adjusted by the observers were used to reproduce the stimuli on the corresponding displays, and the SPDs were measured from the observer's eye position. The chromaticities (u', v') of all the adjusted stimuli derived from the measured SPDs and from the predictions using the control program are shown in Fig. 3(a), with the chromaticity difference $\Delta u'v'$ values between 0.0006 and 0.0052 and an average $\Delta u'v'$ of 0.0026. Figure 3(b) shows the histogram of the luminance values of the adjusted stimuli derived from the measured SPDs, with an average of 87.06 cd/m². These suggested that the control program and calibration used in part I of the study [24] also produced reliable data in this experiment.

3.2. Intra- and inter-observer variations

Both the intra- and inter-observer variations were characterized using the mean color difference from the mean (MCDM) in the CIE 1976 $u'v'$ chromaticity diagram, which was similar to part I [24]. Figure 4 shows the chromaticities, together with the 95% confidence error ellipses, of the repeated matches made by each observer for Stimulus 2 using the four test displays. Both the size and orientation of the ellipses overlapped well. The histograms of the MCDM values of four displays are shown in Fig. 5, with the average MCDM values being 0.0023, 0.0023, 0.0020, and 0.0021 for Displays A, B, C, and D respectively. These MCDM value were smaller than

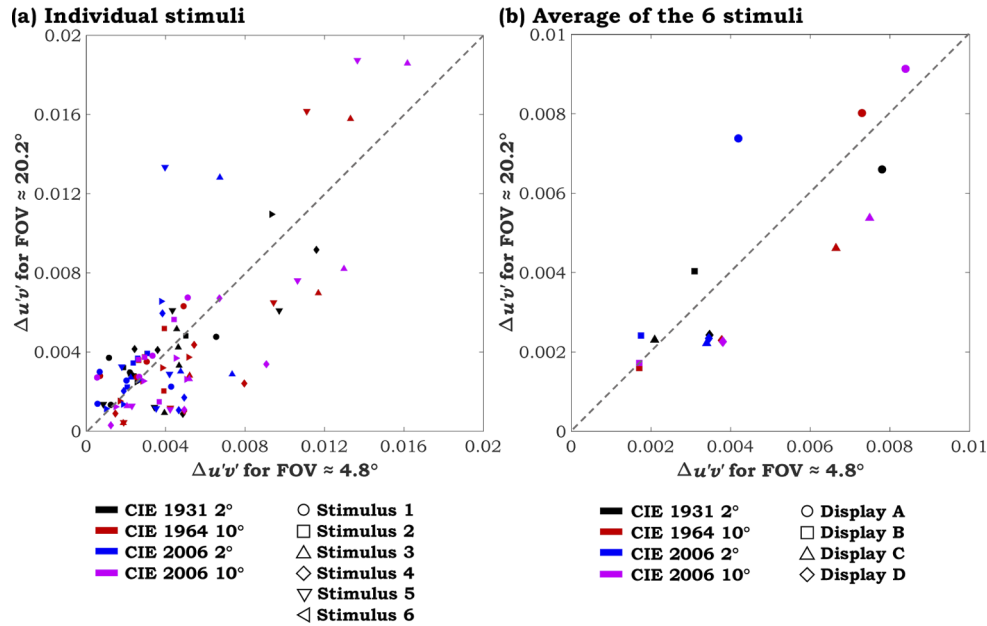


Fig. 9. Scatter plots of the $\Delta u'v'$ values calculated using four CMFs between this experiment with an FOV $\approx 20.2^\circ$ and part I with an FOV $\approx 4.8^\circ$ [24]. (a) Individual stimuli (6 stimuli \times 4 displays \times 4 CMFs = 96 data points); (b) Average of the six stimuli for each display (4 displays \times 4 CMFs = 16 data points).

0.004 units of $u'v'$ (i.e., ≈ 1 unit of just-noticeable color difference, JND) [25]. In particular, they were smaller than those in part I of the study [24], which was believed due to the fact that the new apparatus allowed the two stimuli to appear adjacent to each other in the experiment. The MCDM values for characterizing the inter-observer variations were between 0.0022 and 0.0095, with an average of 0.0047, which was also smaller than those in part I [24]. Therefore, the experiment results were believed to be highly reliable.

3.3. Performance of the four CMFs in characterizing color matches

The chromaticities of the stimuli adjusted by all the observers using the four displays and the 95% confidence error ellipses were calculated using the four CMFs, as shown in the CIE 1976 $u'v'$ chromaticity diagram (Fig. 6). Figure 7 shows the average chromaticities of the stimuli adjusted by the observers (i.e., the center of each ellipse in Fig. 6) and the chromaticities of the corresponding reference stimuli using the four CMFs, with Fig. 8 and Table 1 summarizing the chromaticity differences $\Delta u'v'$. In general, the distributions of the chromaticities and ellipses shown in Fig. 6 were similar to those in part I of the study [24].

In color matching experiments, a smaller $\Delta u'v'$ is always preferred, since an optimal set of CMFs should result in a $\Delta u'v'$ value of zero for the color stimuli having matched color appearance. Overall, the CIE 2006 2°, CIE 1964 10°, and CIE 1931 2° CMFs resulted in similar average $\Delta u'v'$ values around 0.004, which was smaller than the value of 0.005 of the CIE 2006 10° CMFs. Figure 9 shows the $\Delta u'v'$ values calculated using the four CMFs in this experiment versus those in part I (i.e., FOV of 20.2° versus 4.8°) [24]. In general, the data points were scattered around the diagonal line, suggesting the similar results between the two experiments. For Display D, which had the most similar primaries as the reference display, the four data points were below the diagonal line, as shown in Fig. 9(b), suggesting smaller chromaticity differences for stimuli with matched color appearance when using the four CMFs for the larger FOV. When applied

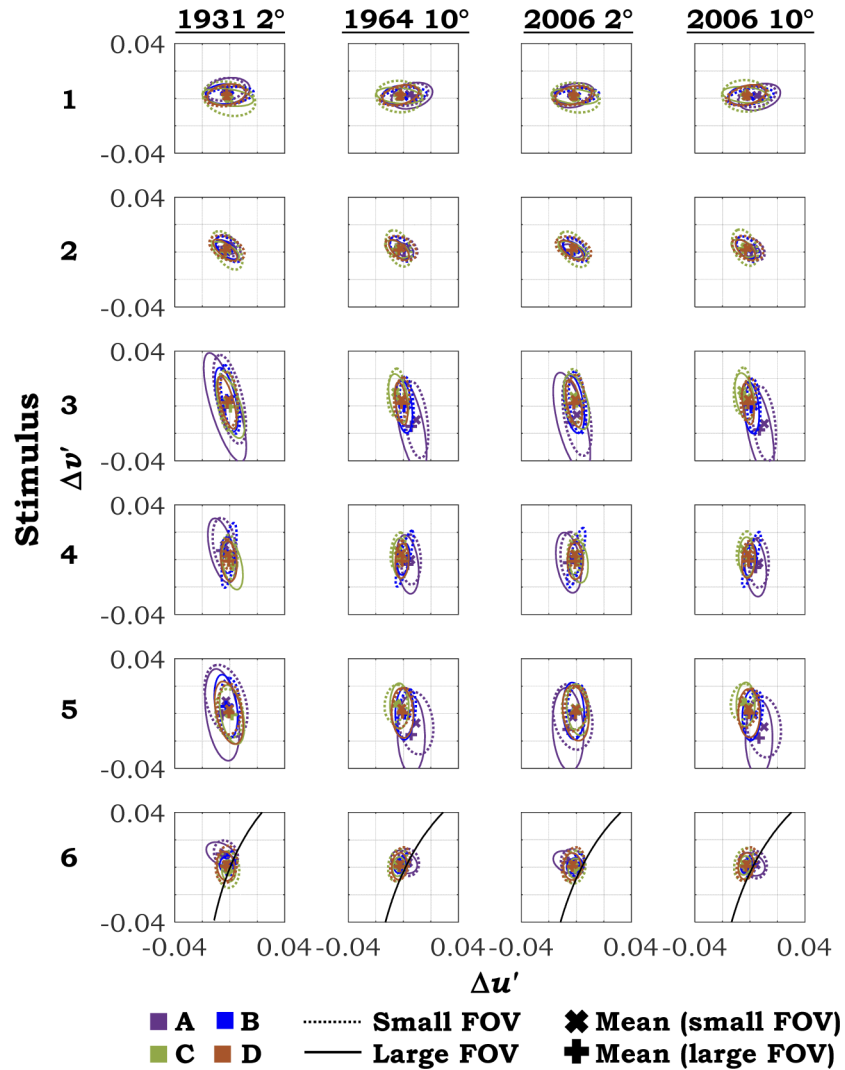


Fig. 10. Average chromaticities of the stimuli adjusted by the 53 observers in this experiment and the 50 observers in part I (ellipse centers) [24] and the fitted 95% confidence error ellipses in relative to the chromaticities of the corresponding reference stimuli shown on the reference display, plotted using the four CMFs. (note: the blackbody loci calculated using the four CMFs are plotted with Stimulus 6).

on Display A, the CIE 2006 2° CMFs resulted in a larger chromaticity difference from ~ 0.004 to ~ 0.008 units for the larger FOV, while the two 10° CMFs resulted in similar chromaticity differences for the two FOVs (i.e., closer to the diagonal line). For the three OLED displays (i.e., Displays B, C, and D), the CIE 2006 2° CMFs still had the best performance for all the six color stimuli on average. Thus, though the size of the stimuli were increased in this experiment, the results still did not support the CIE's recommendation of using 10° CMFs for characterizing the color of stimuli with an FOV greater than 4°.

Display A (LCD) always had the largest average $\Delta u'v'$ values regardless of the CMFs, which should be due to the significantly different primaries from the reference display. In contrast, Display D always had the smallest average $\Delta u'v'$ values regardless of the CMFs and all the

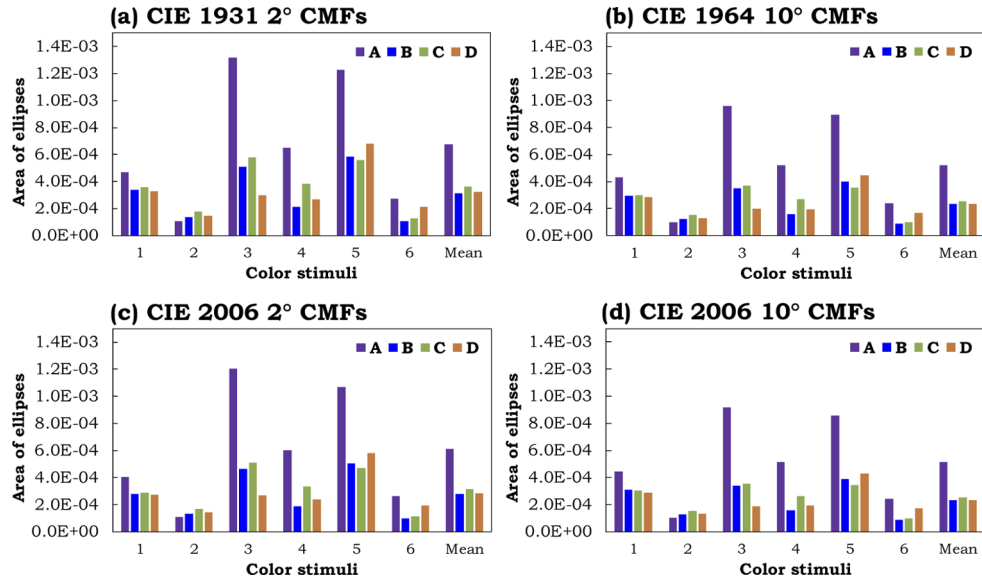


Fig. 11. Areas of the fitted 95% confidence error ellipses of the stimuli adjusted by the 53 observers calculated using the four CMFs, as shown in Fig. 10. (a) CIE 1931 2° CMFs; (b) CIE 1964 10° CMFs; (c) CIE 2006 2° CMFs; (d) CIE 2006 10° CMFs.

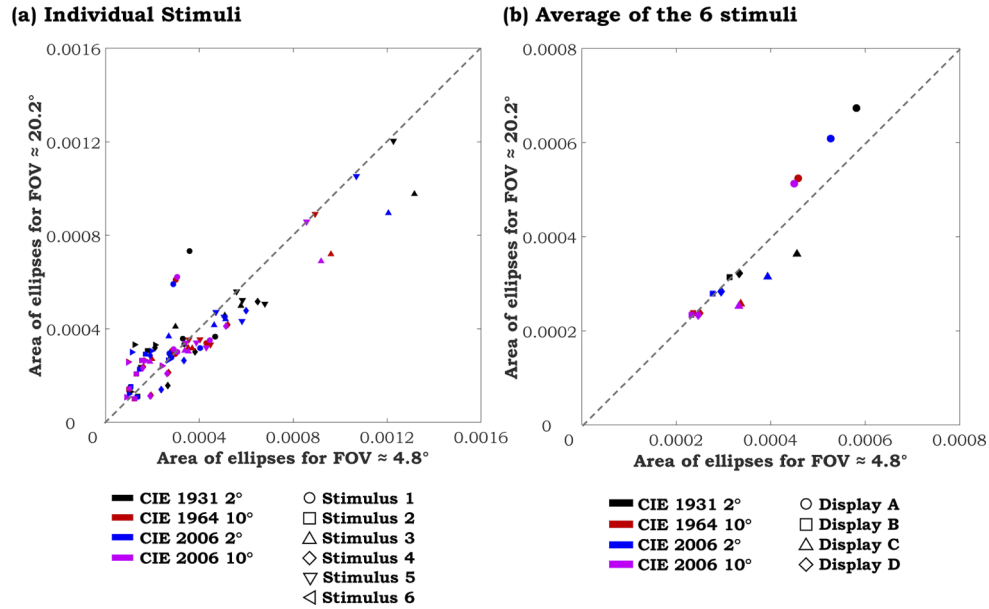


Fig. 12. Scatter plots of the areas of fitted 95% confidence error ellipses of the stimuli adjusted by the 53 observers using the four CMFs between this experiment with an FOV ≈ 20.2° and part I with an FOV ≈ 4.8° [24]. (a) Individual stimuli (6 stimuli × 4 displays × 4 CMFs = 96 data points); (b) Average of the six stimuli for each display (4 displays × 4 CMFs = 16 data points).

$\Delta u'v'$ values were consistently smaller than 0.004 (i.e., 1 JND), which was believed due to the very similar primaries as the reference display (note: Display D and the reference display were

manufactured by the same supplier). In particular, the average $\Delta u'v'$ values of the six stimuli for the four displays were 0.0066, 0.0040, 0.0023, and 0.0024 respectively when calculated using the CIE 1931 2° CMFs, which were 2.87, 1.74, 1.15, and 1.14 times of the MCDM values of the intra-observer variations summarized in Fig. 5, suggesting the meaningfulness of the comparisons among the displays. Therefore, it is obvious that the performance of the CMFs varied with the primary sets.

When the CIE 1931 2° CMFs, the most widely used CMFs for display calibration and characterization, were used, Display A (LCD) had the largest chromaticity differences $\Delta u'v'$, with all the values greater than 0.004 (i.e., 1 JND). As shown in Fig. 7, all the chromaticities of Display A (LCD) were shifted towards the $-u'+v'$ direction, in relative to those of the reference display. This implies that serious color mismatches would be perceived, if the color stimuli shown on Display A (LCD) and the reference display were calibrated to have the same chromaticities using the CIE 1931 2° CMFs. In particular, a stimulus appearing neutral on the LCD display will appear to have a yellow-green tint on the OLED display, while a stimulus appearing neutral on the OLED display will appear to have a pinkish tint on the LCD display. Both the color mismatch and the appearance of the tint were consistent to our observations. More importantly, the change of the CMFs cannot simultaneously reduce the $\Delta u'v'$ values of the six color stimuli. The decrease in the $\Delta u'v'$ values for Stimuli 2 (green), 4 (purple), and 6 (neutral) always came with the increase for Stimuli 1 (red), 3 (blue), and 5 (reddish purple). Thus, the performance of the CMFs also varied with the color stimuli.

3.4. Performance of the four CMFs in characterizing observer metamerism

The degree of observer metamerism can be characterized based on the size of the 95% confidence error ellipses shown in Fig. 6. In order to more clearly make the comparisons, we moved the chromaticities of the stimuli shown on the reference display to the origin, and plotted the 95% confidence error ellipse in Fig. 10, with the ellipse areas being summarized in Fig. 11.

It can be found that both the size and orientation of the ellipses were similar among the four CMFs. The degrees of observer metamerism were very similar among the three OLED displays (i.e., Displays B, C, and D), which were much smaller than that of Display A (LCD), especially for Stimuli 3 (blue), 4 (purple), and 5 (reddish purple). Figure 12 compares the ellipse areas between the results of this experiment and those in part I [24]. The datapoints were closer to the diagonal line than those shown in Fig. 9, suggesting little effect of the FOV on the degree of observer metamerism. It can be observed that the two 10° CMFs resulted in the smaller ellipses area, suggesting that these two CMFs had better performance in predicting the variations among the observers.

4. Conclusion

In this experiment, 53 human observers performed color match of six color stimuli using four smartphone displays (i.e., one LCD display having an sRGB color gamut and three OLED displays having P3 color gamuts) and a reference OLED smartphone display. The stimuli were carefully selected in a physiological-based chromaticity diagram. In comparison to the experiment setup in part I, a new apparatus was carefully designed and built. It not only made the color stimuli produced by the test and reference displays to appear adjacent to each other, allowing an easier color match for the observers, but also increased the field of view (FOV) from 4.77° to 20.2°, making the display size more similar to the typical viewing condition. Both the inter- and intra-observer variations, as characterized using the mean color difference from the mean (MCDM) values, were smaller than those in part I.

Large differences were found between the LCD and the reference OLED displays. The four CMFs always had the worst performance in characterizing the color matches between the LCD and the reference OLED displays. And the LCD display always introduced the greatest degree of

observer metamerism. This suggested the importance of the color primaries, in terms of the peak wavelengths and spectral shapes, to perceived color matches and observer metamerism.

On the other hand, the performance of the four CMFs, in terms of characterizing color matches and observer metamerism, were very similar to those in part I. When using the CIE 1931 2° CMFs to characterize the color matches, the chromaticities of the stimuli produced by the LCD display were significantly different from those produced by the OLED displays, with the chromaticities being shifted towards the $-u' + v'$ direction in the CIE 1976 $u'v'$ chromaticity diagram. This suggests that if the LCD and OLED displays are calibrated to produce stimuli with the same chromaticities using the CIE 1931 2° CMFs, the stimuli shown on the OLED display will have a green-yellow tint if those shown on the LCD displays appear neutral. In addition, though the stimuli in this experiment had an FOV of 20.2°, the CIE 2006 2° CMFs still had the best performance among the four CMFs for all the displays, in terms of the average chromaticity differences among stimuli with matched color appearance, which still failed to support the recommendation using 10° for stimuli with an FOV beyond 4°. The performance of the CMFs also varied with the color stimuli.

Funding. Research Grants Council, University Grants Committee (PolyU 152063/18E).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. CIE, "Colorimetry, 4th edition," in *CIE 015:2018* (CIE, 2018).
2. W. D. Wright, "A re-determination of the trichromatic coefficients of the spectral colors," *Trans. Opt. Soc.* **30**(4), 141–164 (1929).
3. W. D. Wright, "A re-determination of the mixture curves of the spectrum," *Trans. Opt. Soc.* **31**(4), 201–218 (1930).
4. J. Guild, "The colorimetric properties of the spectrum," *Phil. Trans. R. Soc. Lond. A* **230**(681-693), 149–187 (1931).
5. W. S. Stiles and J. M. Burch, "N.P.L. color-matching investigation: Final report," *Opt. Acta* **6**(1), 1–26 (1959).
6. N. I. Speranskaya, "Determination of spectral color coordinates for twenty-seven normal observers," *Optics Spectrosc.* **7**, 424–448 (1959).
7. CIE, "Fundamental chromaticity diagram with physiological axes – part 1," in *CIE 170-1:2006* (CIE, 2006).
8. CIE, "Fundamental chromaticity diagram with physiological axes – part 2: spectral luminous efficiency functions and chromaticity diagrams," in *CIE 170-2:2015* (CIE, 2015).
9. Y. Asano, "Individual colorimetric observers for personalized color imaging," PhD dissertation at Rochester Institute of Technology (2015).
10. A. Sarkar, "Identification and assignment of colorimetric observer categories and their applications in color and vision sciences," PhD dissertation at University of Nantes (2011).
11. M. Huang, G. Cui, Y. Liu, and H. Liu, "Analysis of observers metamerism differences for different retinal cone visual responses," *Spectros. Spect. Anal.* **35**(10), 2802–2809 (2015).
12. Y. Hu, M. Wei, and M. R. Luo, "Observer metamerism to display white point using different primary sets," *Opt. Express* **28**(14), 20305–20323 (2020).
13. X. Hu and K. W. Houser, "Large-field color matching functions," *Color Res. Appl.* **31**(1), 18–29 (2006).
14. K. W. Houser and X. Hu, "Visually matching daylight fluorescent lamp light with two primary sets," *Color Res. Appl.* **29**(6), 428–437 (2004).
15. B. Oicherman, M. R. Luo, B. Rigg, and A. R. Robertson, "Effect of observer metamerism on color matching of display and surface colors," *Color Res. Appl.* **33**(5), 346–359 (2008).
16. B. Bodner, N. Robinson, R. Atkins, and S. Daly, "Correcting metameric failure of wide color gamut displays," *Dig. Tech. Pap. - Soc. Inf. Disp. Int. Symp.* **49**(1), 1040–1043 (2018).
17. J. Li, P. Hanselaer, and K. A. G. Smet, "Impact of color matching primaries on observer matching: Part I-accuracy," *Leukos*, published online, DOI: 10.1080/15502724.2020.1864395.
18. J. Li, P. Hanselaer, and K. A. G. Smet, "Impact of color matching primaries on observer matching: Part II-observer variability," *Leukos*, published online, DOI: 10.1080/15502724.2020.1864396.
19. C. Y. Bai and L. C. Ou, "Observer variability study and method to implement observer categories for novel light source projection system," *Color Res. Appl.* **46**(5), 1019–1033 (2021).
20. Y. Seo, E. Lee, Y. Yi, B. Choi, and S. Jo, "Color correction method based on spectral distribution for solving metameric failure in wide color gamut displays," *Dig. Tech. Pap. - Soc. Inf. Disp. Int. Symp.* **52**(1), 454–457 (2021).
21. R. L. Alfvén and M. D. Fairchild, "Observer variability in metameric color matches using color reproduction media," *Color Res. Appl.* **22**(3), 174–188 (1997).

22. D. C. Rich and J. Jalijali, "Effects of observer metamerism in the determination of human color-matching functions," *Color Res. Appl.* **20**(1), 29–35 (1995).
23. H. Xie, S. P. Farnand, and M. J. Murdoch, "Observer metamerism in commercial displays," *J. Opt. Soc. Am. A* **37**(4), A61–A69 (2020).
24. J. Wu, M. Wei, Y. Fu, and C. Cui, "Color mismatch and observer metamerism between conventional liquid crystal displays and organic light emitting diode displays," *Opt. Express* **29**(8), 12292–12306 (2021).
25. CIE, "Chromaticity Difference Specification for Light Sources," in *CIE TN 001:2014* (CIE, 2014).