Research Article | Vol. 31, No. 17 / 14 Aug 2023 / Optics Express 28241

Optics EXPRESS

Color difference evaluations on metameric color stimuli by observers of three age groups

MIN HUANG,[1] XINYUAN GAO,[1] MINCHEN WEI,[2,3,*] YUETONG SHEN,[1,2] YU WANG,[1] AND XIU LI[1]

[1] *Printing and Packaging Engineering, Beijing Institute of Graphic Communication, Beijing, China*
[2] *Color, Imaging, and Illumination Laboratory, The Hong Kong Polytechnic University, Kowloon, Hong Kong*
[3] *Color, Imaging, and Metaverse Research Center, The Hong Kong Polytechnic University, Kowloon, Hong Kong*
[*] *minchen.wei@polyu.edu.hk*

Abstract: Color matching functions (CMFs), which are used to derive tristimulus values and chromaticities, play a very important role in color characterization, calibration, and specifications. Great efforts have been made to investigate how CMFs can characterize the color matches and mismatches between stimuli with different spectral compositions under different sizes of field of view (FOV). Few study specifically investigated how to better characterize the CMFs for different observer ages. In this study, we carefully designed a series of color stimuli using different CMFs based on our two past studies, and asked 51 young, middle-aged, and senior observers to evaluate the color differences. The three sets of CIE 2006 CMFs (i.e., 10°, 4°, and 2°) with an age of 20-, 40-, and 70-year were found to characterize the perceived color differences for the young, middle-aged, and senior observers. In addition, it was found that the characterization of a group of observers can also be performed using these three sets of CMFs based on the distributions of the observers, which had a better correlation to the perceived color differences than the individual colorimetric model.

1. Introduction

Color characterization and calibration play an extremely important role in different applications, such as lighting, camera, display, and printing. Tristimulus values and chromaticity coordinates are the basic quantities that are used to derive various colorimetric quantities and metrics. Stimuli having the same tristimulus values (a.k.a., same chromaticities and luminance) are expected to have matched color appearance (i.e., no perceived color difference) when viewed under the same condition. Both tristimulus values and chromaticities are calculated using the spectral power distribution (SPD) of the stimulus, which characterizes the radiant power produced by the stimulus between 380 and 780 nm, and the color matching functions (CMFs), which characterizes the color matching mechanism and the responses of the three cone receptors in the human visual system within the visible spectrum. Therefore, stimuli with different SPDs can produce identical tristimulus values and perceived color appearance, which are known as metameric stimuli. More importantly, color specification and calibration are performed using the tristimulus values and chromaticities, so that systems with different SPDs can be specified and calibrated to produce the same colors [1].

Currently, the CIE 1931 CMFs are the most widely used CMFs in practice and specified in various standards, though it was developed based on only 17 observers for stimuli with a field of view (FOV) of 2° [2,3]. In 1964, the CIE recommended a new set of CMFs (i.e., CIE 1964 10° CMFs) for characterizing color matches under an FOV beyond 4° [1]. Due to the metameric failure introduced by these two sets of CMFs (i.e., the stimuli having the same tristimulus values or chromaticities do not match in color appearance), a CIE technical committee TC1-36

#495090 https://doi.org/10.1364/OE.495090
Journal © 2023 Received 12 May 2023; revised 12 Jul 2023; accepted 26 Jul 2023; published 9 Aug 2023

recommended a model to estimate the cone fundamentals for normal human observers [4] in 2006 and later derived the corresponding CMFs and chromaticity diagrams in 2015 [5]. In particular, this model considers two important parameters—age and FOV—which have significant impacts on the cone fundamentals and CMFs, with the age between 20 and 80 years and FOV between 1° to 10° [4,5].

Though it is widely accepted that age and FOV are the two most important parameters affecting the CMFs, much greater efforts have been made to quantify the effect of FOV, which is more closely related to different applications [6–12]. Among the various studies, two specifically compared the color matching results under two FOVs. Hu and Houser completed a Maxwell color matching experiment with two FOVs—10° and 102° × 50° [7]; Wu and Wei completed a color matching experiment with two FOVs—4.77° and 20.2° [11,12]. In contrast, only a limited number of studies specifically investigated the effect of age, though great number of studies have shown the physiological factors, such as lens density, change with age [13–15]. In 2020, we carried out an experiment to investigate how the young and senior observers have different performance in perceiving color matches and mismatches, and found a significant difference between the two age groups [16]. In 2022, we carried out another experiment to compare the performance of the CIE 1931 2°, 1964 10°, and CIE 2006 CMFs in characterizing the perceived color differences of stimuli with two FOVs (i.e., 2.9° and 8.6°) [17]. The results suggested that the CIE 1931 2° and 1964 10° CMFs had better performance in characterizing the senior and young observers respectively.

In this study, we aim to further investigate the effect of observer age by including three age groups (i.e., young, middle-aged, and senior). Fifty-one observers between 19 and 83 years of age, with 17 in each age group, completed the color difference evaluations on 26 sets of stimuli, with each stimulus having an FOV of 8.2°. All the stimuli were carefully designed and calibrated using different sets of CMFs, and each set was evaluated five times by each observer.

2. Methods

2.1. Apparatus

Three identical spectrally tunable LED devices with a diffuse light emitting area were used to produce stimuli in the experiment. The light emitting area of each device was 5 cm × 5 cm, with the uniformity and stability characterized in our past study [17]. Figure 1 shows the SPDs of the LED channels measured using a PhotoResearch PR-670 spectroradiometer. The three devices were placed next to each other, as illustrated in Fig. 2, with the center one producing reference stimuli and the other two producing test stimuli. During the experiment, the observer was seated 35 cm in front of the devices, so that each stimulus had an FOV around 8.2°.

2.2. Color stimuli

The reference stimuli were always produced by an L1 primary set (i.e., R1-G2- B1). The test stimuli shown on the sides were produced by L2 and L3 primary sets respectively (i.e., L2: R1-G1-B1, L3: R1-G2-B2), with one of the primaries being different from the reference stimuli, as summarized in Table 1 and the gamut shown in Fig. 3.

Two groups of stimuli (Stimuli 1 to 20 and Stimuli 21 to 26) were carefully designed, with Table 2 summarizing the chromaticities with the corresponding CMFs and Table 3 summarizing the color differences between the test stimuli to the reference stimuli. The luminance levels of the reference and test stimuli were calibrated to the same, all of which were above 8 cd/m^2.

Stimuli 1 to 20 were designed, so that each of the four sets of CMFs resulted in the same chromaticities of the reference and test stimuli regardless of the primary sets, with five stimuli for each set of CMFs. These five stimuli are recommended by the CIE. These four sets of CMFs included the CIE 1931 2°, 1964 10°, CIE 2006 4° (20-years), and CIE 2006 4° (70-years).

Optics EXPRESS

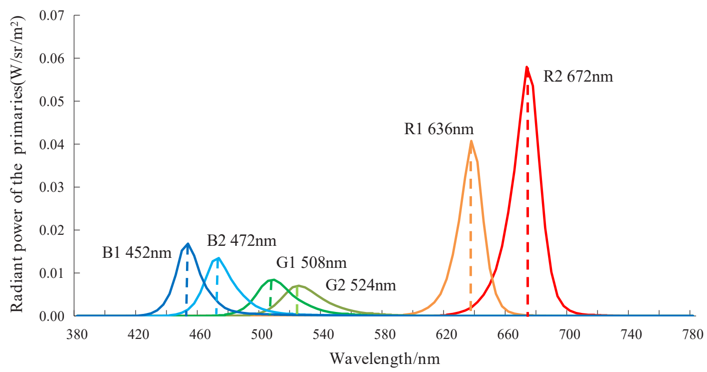


Fig. 1. Spectral power distributions (SPDs) of the six primaries in the spectrally tunable LED device measured using a PhotoResearch PR-670 spectroradiometer.

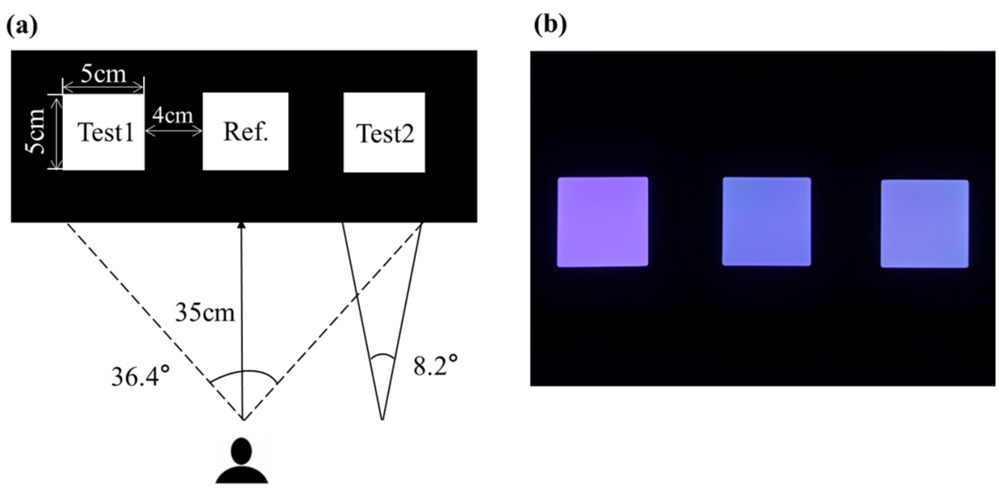


Fig. 2. Illustration of the experiment setup. (a) Arrangement of the LED devices; (b) Photograph captured at the observer's eye position.

Table 1. Summary of the primaries and colorimetric characteristics of the primary sets used to produce the stimuli and the standard color spaces

	Name	Shifted primary	Color gamut area	
			CIE 1931 *xy*	CIE 1976 *u'v'*
Primary sets	L1(R1-G2-B1)	-	0.1956	0.0950
	L2(R1-G1-B1)	G1	0.1879	0.1000
	L3(R1-G2-B2)	B2	0.1819	0.0561
Standard color spaces	Adobe RGB		0.1512	0.0758
	DCI P3		0.1520	0.0815
	Rec 2020		0.2116	0.1118

Optics EXPRESS

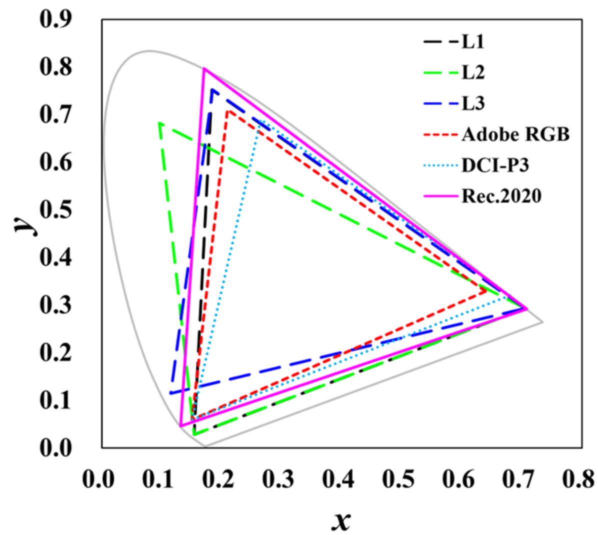


Fig. 3. Color gamuts enclosed by the three primary sets, together with the standard gamuts, in the CIE 1931 *xy* chromaticity diagram, with the chromaticities calculated using the CIE 1931 2° CMFs.

Table 2. Summary of the chromaticities of the stimuli using different CMFs. For Stimuli 1 to 20, all the three stimuli were designed to have the same chromaticities with the corresponding CMFs. For Stimuli 21 to 26, one primary set was designed to have the same chromaticities as the reference and the other primary set was designed to have large chromaticity differences to the reference using the two CMF sets, with the chromaticities of the reference stimuli underlined

CMFs	Stimulus No	Color	(*x*,*y*)	CMFs	Stimulus No	Color	(*x*,*y*)
CIE 1931 2° CMFs	1	Gray	(0.348,0.357)	CIE 1964 10° CMFs	6	Gray	(0.347,0.359)
	2	Red	(0.516,0.349)		7	Red	(0.509,0.353)
	3	Yellow	(0.407,0.446)		8	Yellow	(0.408,0.447)
	4	Green	(0.267,0.398)		9	Green	(0.277,0.397)
	5	Blue	(0.215,0.193)		10	Blue	(0.215,0.204)
CIE 2006 4° 20-years-CMFs	11	Gray	(0.349,0.366)	CIE 2006 4° 70-years-CMFs	16	Gray	(0.338,0.377)
	12	Red	(0.510,0.358)		17	Red	(0.507,0.365)
	13	Yellow	(0.410,0.455)		18	Yellow	(0.401,0.476)
	14	Green	(0.275,0.408)		19	Green	(0.264,0.422)
	15	Blue	(0.213,0.208)		20	Blue	(0.211,0.209)

Stimulus No	Color	CIE 1931 2° CMFs (*x*,*y*)		CIE 1964 10° CMFs (*x*,*y*)	
		L2	L3	L2	L3
21	White1	(0.353,0.339)	(0.342,0.358)	(0.343,0.359)	(0.339,0.380)
22	White2	(0.343,0.357)	(0.347,0.335)	(0.334,0.379)	(0.341,0.358)
23	Green1	(0.368,0.397)	(0.359,0.417)	(0.362,0.416)	(0.361,0.431)
24	Green2	(0.358,0.418)	(0.361,0.401)	(0.353,0.438)	(0.361,0.416)
25	Blue1	(0.278,0.277)	(0.272,0.290)	(0.271,0.293)	(0.267,0.322)
26	Blue2	(0.271,0.291)	(0.280,0.259)	(0.265,0.307)	(0.271,0.294)

Table 3. Summary of the color differences, in terms of ΔE^*_{ab} between the test and reference stimuli using the different CMFs.

CMFs	Stimulus No	Ref. and L2	Ref. and L3	CMFs	Stimulus No	Ref. and L2	Ref. and L3
CIE 1931 2° CMFs	1	0.32	0.48	CIE 1964 10° CMFs	6	0.40	0.39
	2	0.53	0.34		7	0.71	0.75
	3	0.34	0.28		8	0.62	0.77
	4	0.46	0.57		9	0.40	0.33
	5	0.40	0.16		10	0.54	0.48
CIE 2006 4° 20-years-CMFs	11	0.59	0.63	CIE 2006 4° 70-years-CMFs	16	0.70	0.59
	12	0.37	0.23		17	0.35	0.17
	13	0.29	0.41		18	0.66	0.41
	14	0.30	0.37		19	0.31	0.36
	15	0.46	0.38		20	0.25	0.30

Stimulus No	CIE 1931 2° CMFs		CIE 1964 10° CMFs	
	Ref. and L2	Ref. and L3	Ref. and L2	Ref. and L3
21	14.70	0.57	0.73	13.25
22	0.33	15.58	14.38	0.36
23	8.92	0.62	0.64	5.01
24	0.24	6.25	8.66	0.33
25	7.93	0.46	0.26	13.20
26	0.27	16.35	7.64	0.50

Stimuli 21 to 26 were designed, so that the two sets of CMFs (i.e., CIE 1931 2° and CIE 1964 10° CMFs) resulted in one of the two primary sets, L2 or L3, have the same chromaticities to the reference, and the other primary set have much larger chromaticity difference to the reference stimuli. In particular, for Stimuli 21, 23, and 25, the CIE 1931 2° CMFs resulted in the same chromaticities to the L3 and reference sets and very large differences to the L2 and reference sets, while the CIE 1964 10° CMFs resulted in the same chromaticities to the L2 and reference sets and very large differences to L3 and reference sets. In contrast, for Stimuli 22, 24, and 26, the CIE 1931 2° CMFs resulted in the same chromaticities to the L2 and reference sets and very large differences to the L3 and reference sets, while the CIE 1964 10° CMFs resulted in the same chromaticities to the L3 and reference sets and very large differences to L2 and reference sets. This was purposely designed based on the results of our past study [17] that the CIE 1931 2° CMFs had better performance for the senior observers and the CIE 1964 10° CMFs had better performance for the young observers.

2.3. Experimental procedures

Upon arrival, the observer completed the Ishihara Color Vision Test and a general information survey, and was then seated 35 cm in front of the three LED devices, with his or her sagittal plane aligned to the center of the device presenting the reference stimuli. The general illumination in the space was then switched off, and the observer can only see the three stimuli in his or her FOV. Three stimuli were then presented to the observer, and the observer was asked to compare the color appearance of the two test stimuli on the sides to the reference stimulus at the center, and judge which of the two test stimuli had a larger color difference to the reference stimulus, which was a forced choice. After the observer spoke out his or her judgment, the experimenter

Optics EXPRESS

recorded it. The three stimuli were switched to the next set, during which the observer's eyes were covered by an eye mask. The observer evaluated each of the 26 sets five times, with a total of 130 judgments, which was used for characterizing the intra-observer variations. The order of the 130 sets was randomized, and the locations of the two test stimuli within each set were also randomized.

2.4. Observers

Fifty-one observers, organized in three age groups (i.e., young, middle-aged, and senior) with 17 observers in each age group completed the experiment. All the observers had normal color vision, as tested using the Ishihara Color Vision Test. Table 4 summarizes the information of the observers in each age group. In total, 6,630 judgments were made by the observers.

Table 4. Summary of the information of the observers in each age group

Age group	Number of observers	Range of ages	Average age	Std. Dev. of age
Young	17 (10 males and 7 females)	19-27	22.8	2.09
Middle-aged	17 (9 males and 8 females)	37-45	41.8	2.59
Senior	17 (8 males and 9 females)	63-83	70.4	4.50

3. Results

3.1. Intra- and inter-observer variations

The intra-observer variation was characterized based on the repeated judgments made by each observer. For each set of the stimuli, the judgment that was made for more than three out of the five times (i.e., greater than 60%) was considered as the final judgment for each observer. The average misjudgment for the 26 stimuli sets was then calculated for each observer. The inter-observer variation was characterized based on the judgments made by each observer and those of an average observer (i.e., the judgments made by more than 50% of the observers). The percentage of the misjudgment was then calculated for each observer. Table 5 summarizes the intra- and inter-observer variations for the three age groups.

Table 5. Summary of the intra- and inter-observer variations, in terms of the percentage of the misjudgments

		Young	Middle-aged	Senior
Intra-observer	Min	0.0	0.0	0.0
	Max	40.0	40.0	40.0
	Mean	11.0	13.3	11.3
Inter-observer	Min	6.2	16.2	14.6
	Max	25.4	40.8	40.0
	Mean	16.5	28.3	22.4

It can be observed that the intra-observer variations were similar across the three age groups, and the intra-observer variations were generally smaller than the inter-observer variations. The young observers had the smallest inter-observer variations.

3.2. Comparisons among the three age groups

The probability of the observers who selected the test stimulus produced by the L2 primary set (i.e., *P*) was calculated and summarized in Table 6. In addition, the last three columns label the age groups having inconsistent judgments (i.e., whether *P* values are consistently greater or smaller than 0.5). Figure 4 shows the relationship between the observers' judgments and color differences calculated using the corresponding CMFs. It was expected that the P values should always be around 0.5 for Figs. 4(a) to (d) due to the similar calculated color differences, and the P values should always be around 0 or 1 for Figs. 4(e) and (f) due to the large calculated color differences.

Table 6. Summary of the color difference judgments

		Probability of the observers selecting the stimulus produced by L2 (P)			Inconsistent judgments		
Stimulus No.		Young	Middle-aged	Senior	Young vs Middle-aged	Young vs Senior	Middle-aged vs Senior
1	Gray	0.06	0.44	0.36			
2	Red	0.25	0.33	0.44			
3	Yellow	0.94	0.95	0.91			
4	Green	0.09	0.27	0.53		√	√
5	Blue	0.00	0.02	0.06			
6	Gray	0.29	0.52	0.20	√		√
7	Red	0.39	0.64	0.61	√	√	
8	Yellow	0.75	0.93	0.99			
9	Green	0.14	0.54	0.44	√		√
10	Blue	0.01	0.00	0.00			
11	Gray	0.62	0.58	0.20		√	√
12	Red	0.59	0.48	0.53	√		√
13	Yellow	0.92	0.99	0.82			
14	Green	0.51	0.54	0.16		√	√
15	Blue	0.06	0.07	0.01			
16	Gray	0.22	0.60	0.64	√	√	
17	Red	0.39	0.49	0.59		√	√
18	Yellow	0.96	1.00	0.98			
19	Green	0.08	0.28	0.59		√	√
20	Blue	0.01	0.01	0.08			
21	White1	0.04	0.48	0.93		√	√
22	White2	0.96	0.60	0.13		√	√
23	Green1	0.09	0.53	0.96	√	√	
24	Green2	0.94	0.95	0.32		√	√
25	Blue1	0.11	0.41	0.46			
26	Blue2	0.75	0.42	0.05	√	√	
Sum					**7**	12	11

For Stimuli 1 to 20, though the reference and test stimuli were designed to have the same chromaticities, there still existed small color differences. It can be observed that such small color differences still introduced obvious perceived color differences. Among the four stimulus groups, only the fourth group (i.e., Stimuli 16 to 20) resulted in expected judgments, with the stimuli produced by the L2 set having larger perceived color differences. In contrast, the judgments on the second group (i.e., Stimuli 6 to 10) contradict to the expectation, with a smaller color difference judged to have larger perceived color difference. Moreover, the general trends among

Optics EXPRESS

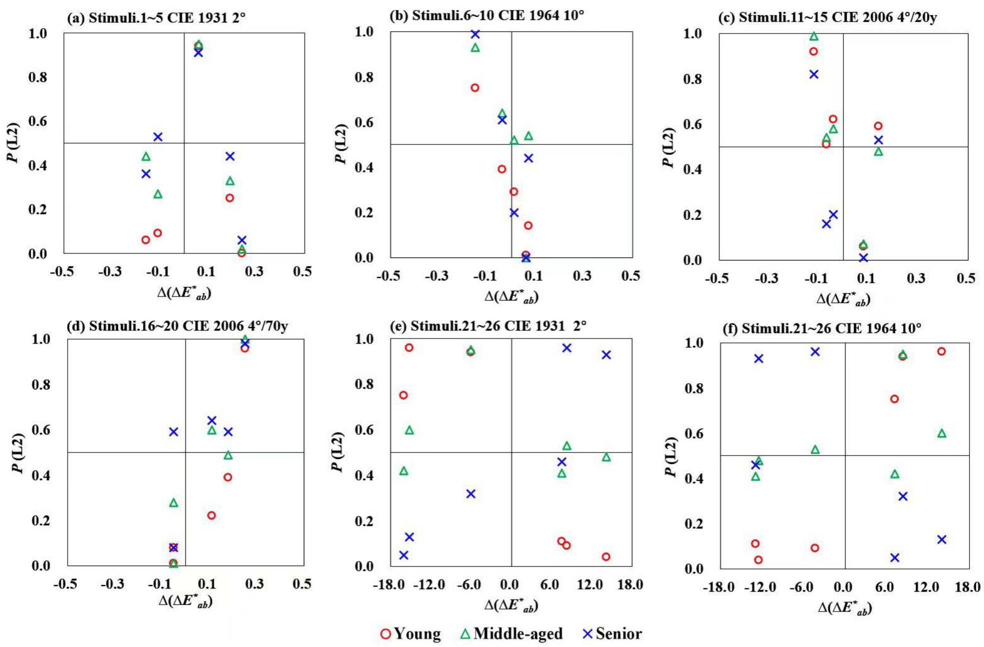


Fig. 4. Scatter plot of the observers' judgements, in terms of the probability of the judgments selecting the test stimuli produced by L2 (i.e., P) to have larger perceived color differences, versus the differences of the color differences $\Delta(\Delta E^*)$ (i.e., $\Delta(\Delta E^*) = \Delta E^*_{\text{Ref}-\text{L2}}$ - $\Delta E^*_{\text{Ref}-\text{L3}}$). The color differences were calculated using the CMFs that were selected to designed the stimuli. (a) Stimuli 1 to 5 with the CIE 1931 2° CMFs; (b) Stimuli 6 to 10 with the CIE 1964 10° CMFs; (c) Stimuli 11 to 15 with the CIE 2006 4° 20-year CMFs; (d) Stimuli 16 to 20 with the CIE 2006 4° 70-year CMFs; (e) Stimuli 21 to 26 with the CIE 1931 2° CMFs; (f) Stimuli 21 to 26 with the CIE 1964 10° CMFs.

the three age groups were similar, with the young and middle-aged observers making more consistent judgments. For all the yellow and blue stimuli, all the three age groups always made consistent judgements.

For Stimuli 21 to 26, it can be found that the judgments made by the young and senior observers were generally inconsistent, which was consistent to how the stimuli were designed. In particular, the color differences calculated using the CIE 1931 2° and 1964 10° CMFs were well correlated to the perceived color differences by the senior and young observers respectively, which corroborated our recent study [17]. In contrast, for the middle-aged observers, around half of the observers judged the stimuli produced by L2 primary set were larger.

4. Discussion

4.1. Performance of the CMFs

In addition to analyzing the observers' judgments based on the color differences calculated using the CMFs that were selected for designing the stimuli, the analyses were also performed using all 32 different CMFs, including the CIE 1931 2°, CIE 1964 10°, and CIE 2006 1° to 10° with the ages of 20, 40, and 70. This resulted in 26 pairs of $\Delta(\Delta E^*)$ (i.e., $\Delta(\Delta E^*) = \Delta E^*_{Ref-L2}$-$\Delta E^*_{Ref-L3}$) values and P value for each set of CMFs. Figure 5 shows the variations of color differences caused by the different CMFs (note: for the CIE 2006 CMFs, only the results calculated using 1°

and 10° are shown, for illustrating the range of variations), which clearly illustrates the obvious effect of using different CMFs.

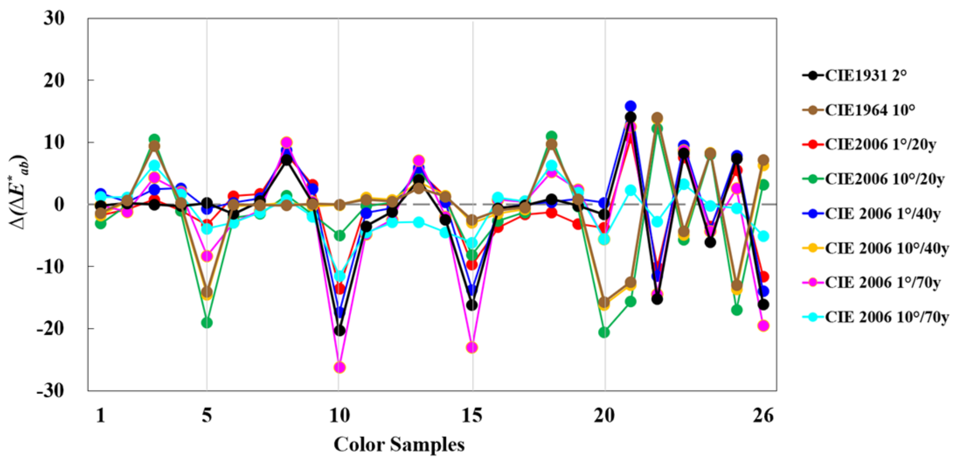


Fig. 5. The variations of the color differences $\Delta(\Delta E^{*})$ (i.e., $\Delta(\Delta E^{*}) = \Delta E^{*}_{\mathrm{Ref-L2}}$ - $\Delta E^{*}_{\mathrm{Ref-L3}}$) caused by the different CMFs for each set of stimuli.

Both linear regression and logistic regression were performed to quantify the relationship between the color difference judgments (i.e., P) and the calculated color differences (i.e., $\Delta(\Delta E^{*})$). The linear correlation coefficient R^2 and the quadratic loss function expressed using Eqs. (1) and (2) were used for the linear and logistic regressions respectively, with the values summarized in Tables 7 and 8.

$$h(x_i) = \frac{1}{1 + e^{-x_i}} \tag{1}$$

$$L(p_i, x_i) = -\frac{1}{n}\sum_{i=1}^{n}(p_i \times Log(h(x_i) + (1 - p_i) \times Log(1 - h(x_i)) \tag{2}$$

where x_{i} represents the difference between the two color difference values ΔE^{*}_{Ref-L2} and ΔE^{*}_{Ref-L3} (i.e., $\Delta(\Delta E^{*}) = \Delta E^{*}_{Ref-L2}$- ΔE^{*}_{Ref-L3}) and P represents the probability of the observers judging the stimulus produced by the L2 primary set had the larger perceived color difference to the reference stimulus.

Table 7. Summary of the linear correlation coefficient R^2 between the calculated color differences and the judgments on the perceived color differences, with the best performance underlined. The closer the value to 1, the better the performance

Group	CIE 1931 2°	CIE 1964 10°	CIE 2006									
			1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
Young	0.019	0.594	0.001	0.138	0.492	0.565	0.611	0.617	0.633	0.648	0.660	0.667
Middle-aged	0.112	0.346	0.123	0.237	0.551	0.753	0.625	0.469	0.416	0.384	0.365	0.371
Senior	0.552	0.008	0.729	0.679	0.616	0.560	0.592	0.646	0.693	0.716	0.704	0.656

It can be seen from Tables 7 and 8 that the best sets of CMFs for each observer group were different when characterized using the two methods. Based on the linear correlation coefficient R^2, the CIE 2006 10°, 4°, and 1° were the best for the three age groups respectively; based on the values calculated using the loss function, the CIE 2006 5°, 4°, and 9° were the best for the three age groups respectively. For the young and senior observers, however, the variations of the

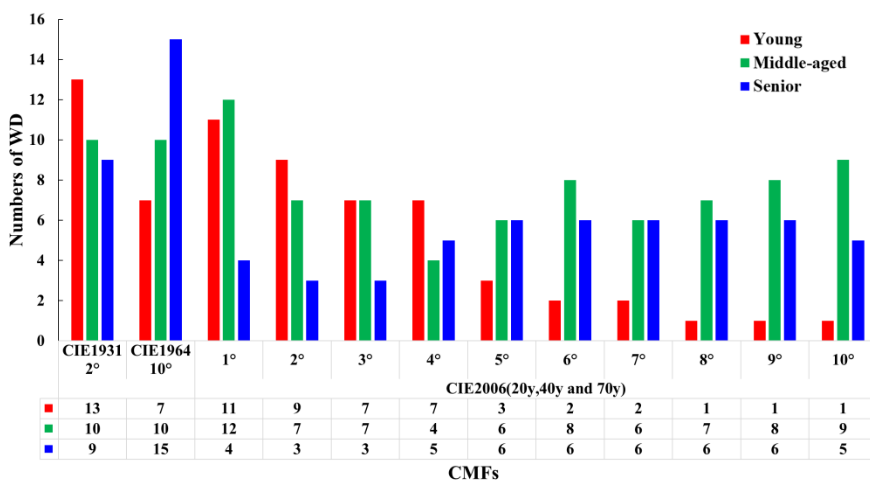
Optics EXPRESS

Table 8. Summary of the values calculated using the loss function of the logistic regression between the calculated color differences and the judgments on the perceived color differences. The closer the value to 0, the better the performance.

Group	CIE 1931 2°	CIE 1964 10°	CIE 2006									
			1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
Young	1.273	0.281	0.999	0.501	0.253	0.229	0.221	0.237	0.242	0.241	0.243	0.247
Middle-aged	0.866	0.601	0.821	0.633	0.388	0.291	0.328	0.404	0.469	0.530	0.581	0.605
Senior	0.359	1.028	0.334	0.322	0.331	0.349	0.329	0.307	0.291	0.286	0.285	0.298

values are smaller when the FOV changes from 5° to 10° and from 1° to 9°. And the CIE 1931 2° and the 1964 10° CMFs were much worse than the optimal solutions of the CIE 2006 CMFs.

In addition to the correlation analyses, the performance of the CMFs was also investigated based on wrong decision judgments. For a certain set of stimuli, the difference of the two color differences $\Delta(\Delta E^*)$ (i.e., $\Delta(\Delta E^*) = \Delta E^*_{Ref-L2}$- ΔE^*_{Ref-L3}) was used to predict whether the test stimulus produced by L2 was more different from the reference, which was then compared to the probability of the observers judging the stimulus produced by the L2 primary set had the larger perceived color difference. Figure 6 summarizes the number of wrong decisions for the 26 stimuli sets. For the three age groups, the CIE 2006 8°-10°, 4°, and 2°-3° CMFs had the best performance for the young, middle-aged, and senior observers respectively.



	CIE1931 2°	CIE1964 10°	1°	2°	3°	4°	5°	6°	7°	8°	9°	10°
Young	13	7	11	9	7	7	3	2	2	1	1	1
Middle-aged	10	10	12	7	7	4	6	8	6	7	8	9
Senior	9	15	4	3	3	5	6	6	6	6	6	5

Fig. 6. Performance of the various CMFs based on the wrong decisions for the 26 stimuli sets (i.e., whether the difference of the calculated color differences correctly characterize the judgments made by the observers).

Based on the above analyses, the CIE 2006 10°, 4°, and 2° CMFs were selected for the young, middle-aged, and senior observers, with the input age set to 20, 40, and 70 years, and Fig. 7 shows the three sets of CMFs. It can be seen that the CMFs shifted towards longer wavelengths and the sensitivities at the short wavelengths are reduced, with the increase of age. Figure 8 shows the scatter plots of the observers' judgments and differences of the calculated color differences using the selected CMFs, together with the fitted lines derived from the linear and logistics regressions.

4.2. Chromaticities of the stimuli calculated using the selected CMFs

The chromaticities of all the stimuli are then plotted using the selected CMFs for the three age groups, with Figs. 9 and 10 showing those of Stimuli 1 to 20 and Stimuli 21 to 26 respectively.

Optics EXPRESS

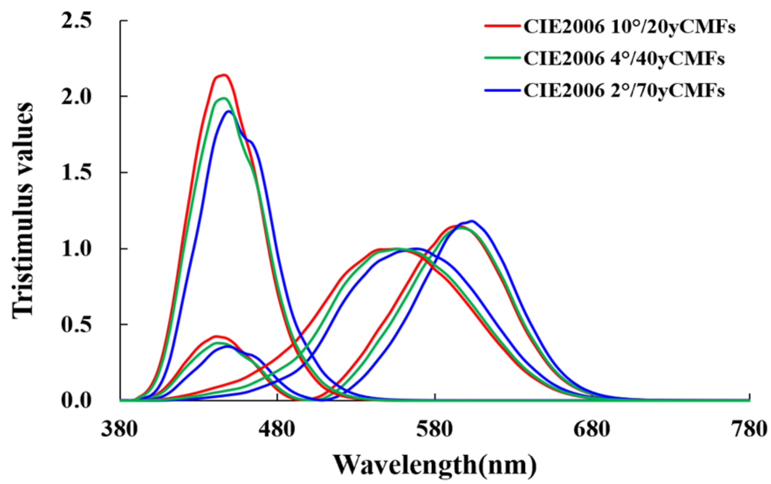


Fig. 7. Color matching functions (CMFs) selected as the optimal sets for the three age groups.

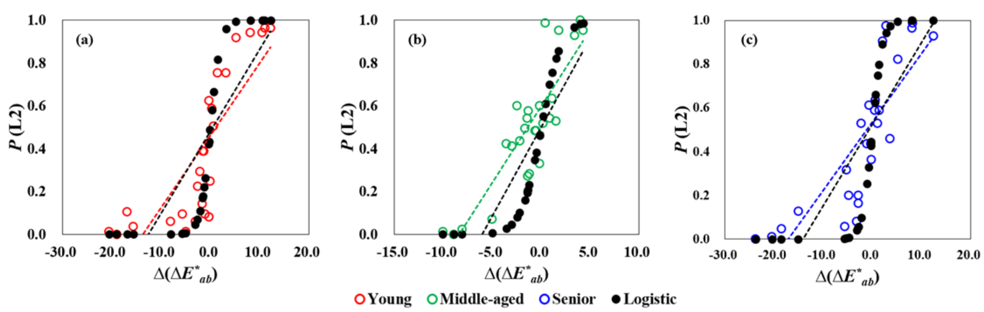


Fig. 8. Scatter plots of the observers' judgements and the differences of the calculated color differences for the three age groups, together with fitted lines derived from the linear and logistic regressions. (a) Young; (b) Middle-aged; (c) Senior.

As shown in Fig. 9, the chromaticities of the red stimuli were generally very closer to each other regardless of the observer's age groups, which corroborated the observers' judgments that 40% to 60% of the observers selecting the stimuli produced by L2 appeared similar to the reference stimuli. In contrast, the chromaticities of the yellow and blue stimuli had the largest variations, which was consistent with the observers' judgments. In particular, the chromaticity distributions suggest that the yellow stimuli produced by the L2 primary set and the blue stimuli produced by the L3 primary set would appear more greenish, and the yellow produced by the L3 primary set and the blue stimuli produced by the L2 primary set would appear more pinkish to the young observers. In contrast, the senior observers could perceive the opposite conditions.

The chromaticity distributions shown in Fig. 10 also support the experiment results. For the young and senior observers, either the L2 or L3 primary set can always introduce a larger chromaticity distance to the reference stimulus, by comparing the two datapoints with different shapes. In contrast, the L2 and L3 primary sets always produced the similar chromaticity distance to the reference stimuli for the middle-aged observers.

Optics EXPRESS

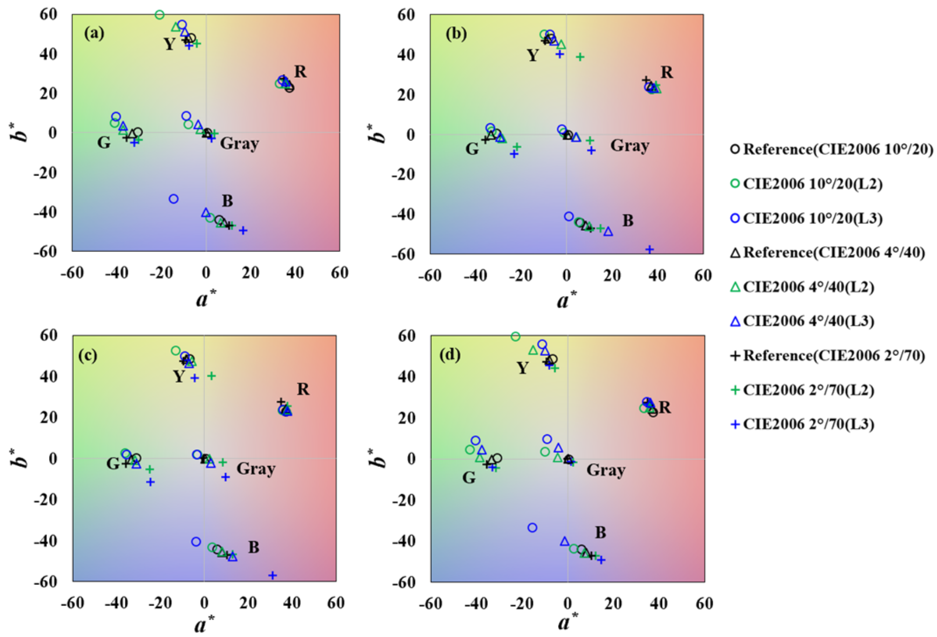


Fig. 9. Chromaticities of Stimuli 1 to 20 calculated by the three CMFs selected for the three age groups in the a^*-b^* plane of the CIELAB color space. (a) Stimuli 1 to 5; (b) Stimuli 6 to 10; (c) Stimuli 11 to 15; (d) Stimuli 16 to 20.

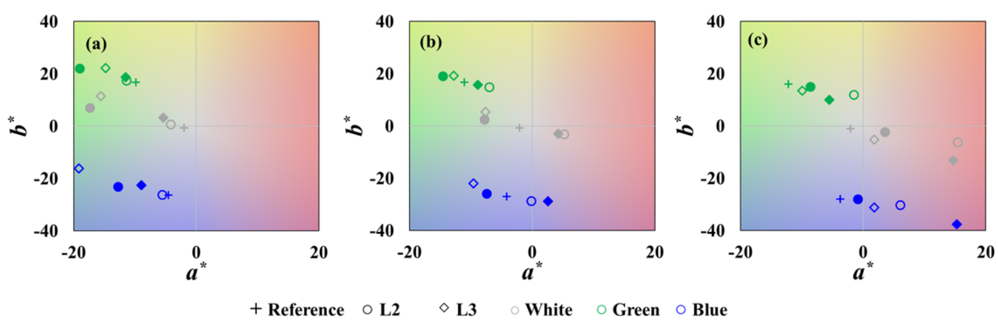


Fig. 10. Chromaticities of Stimuli 21 to 26 calculated by the three CMFs selected for the three age groups in the a^*-b^* plane of the CIELAB color space. (a) Young observers; (b) Middle-aged observers; (c) Senior observers.

4.3. Characterization for a group of observers

In addition to the three sets of CMFs selected for characterizing the three age groups, the characterization can also be performed based on the individual colorimetric model proposed by Asano [18]. Figure 11 shows the comparison between the three selected sets and the 1000 individual CMFs which were modeled based on Monte Carlo simulation, an FOV of 10°, and the 2010 US Census age distribution (i.e., 20.1% between 0 and 14 years old, 66.9% between 15 and 64 years old, and 13% beyond 65 years old, with a mean of 37 years old).

Using the three selected sets of CMFs and the 1000 simulated CMFs, the difference of the color difference between the test stimulus produced by the L2 primary set and the reference and that between the one produced by the L3 primary set and the reference—$\Delta(\Delta E^*)$ (i.e., $\Delta(\Delta E^*) = \Delta E^*_{Ref-L2}$- ΔE^*_{Ref-L3})—can be calculated for each of the 26 sets of stimuli, as shown

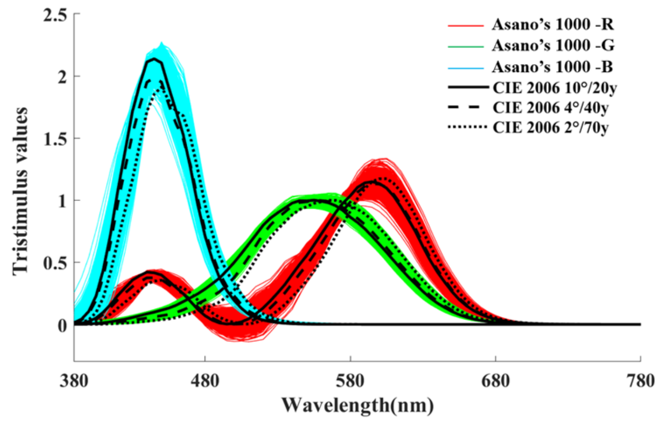


Fig. 11. Comparison among the three selected CMFs for the three age groups and the 1000 CMFs for individual observers derived using the Asano's individual colorimetric model.

in Fig. 12. The discrepancies are expected, since these four sets of CMFs represent different observer groups.

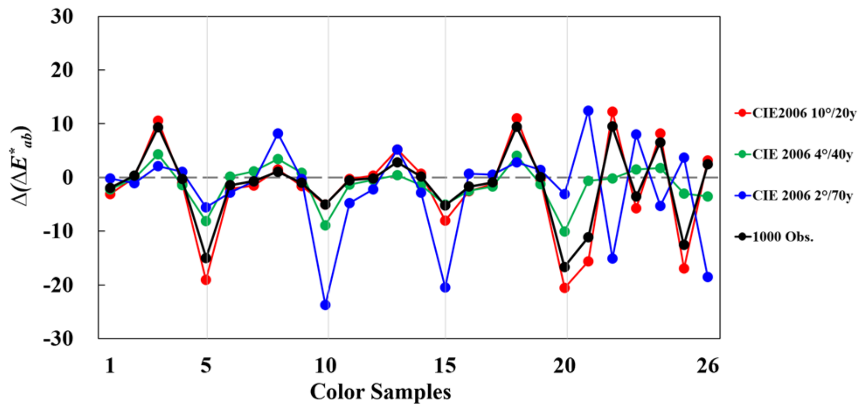


Fig. 12. Comparisons among the differences of the color difference between the test stimulus produced by the L2 primary set and the reference and that between the one produced by the L3 primary set and the reference using the four sets of CMFs for each of the 26 sets of stimuli.

Thus, to characterize a group of observers, we propose to use Eq. (3) using the selected three sets of CMFs, based on the distributions of the observer age groups.

$$\Delta(\Delta E^*_{iab}) = \frac{Y}{Count} \times \Delta(\Delta E^*_{Yiab}) + \frac{M}{Count} \times \Delta(\Delta E^*_{Miab}) + \frac{S}{Count} \times \Delta(\Delta E^*_{Siab}) \tag{3}$$

where Y, M, and S are the number of the young, middle-aged, and senior observers in the group, Count is the total number of the observers (i.e., $Y + M + S$), and i represents the one of the 26 stimulus sets.

This method was then used to characterize a group of observers, with 1/3 of young, 1/3 of middle-aged, and 1/3 of senior observers, for the 26 sets of stimuli. Meanwhile, we adopted the individual observer model proposed by Asano to derive 999 CMFs, with 333 for each of the young, middle-aged, and senior age group, as summarized in Table 9. The calculated differences $\Delta(\Delta E^*)$ for the 26 sets of stimuli is shown in Fig. 13, with Fig. 14 showing the scatter plot and

Table 9. Summary of age information for the 999 observers for deriving the individual CMFs using the Asano's model

Different age groups	Number of Obs.	Age ranges of Obs.	Average age of Obs.	Deviation of Obs. age
Young	333	19~26	22.2	1.93
Middle-aged	333	37~45	39.9	2.57
Senior	333	65~77	69.9	3.42

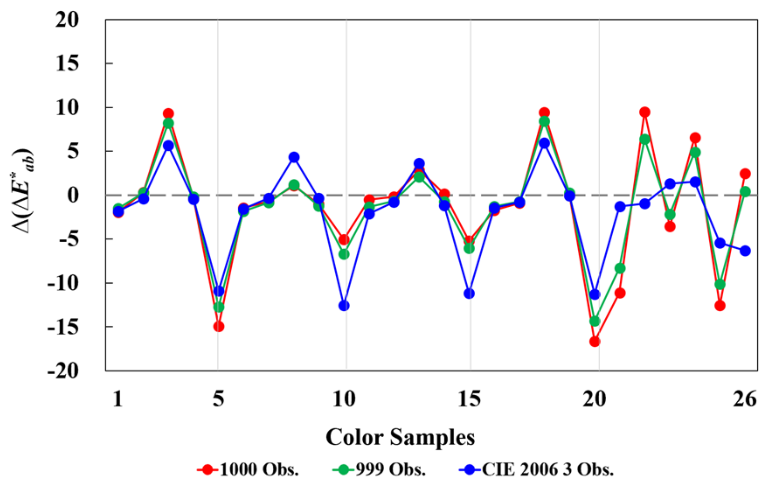


Fig. 13. Comparisons among the differences of the color difference between the test stimulus produced by the L2 primary set and the reference and that between the one produced by the L3 primary set and the reference for each of the 26 sets of stimuli for a group of observers derived using the three methods.

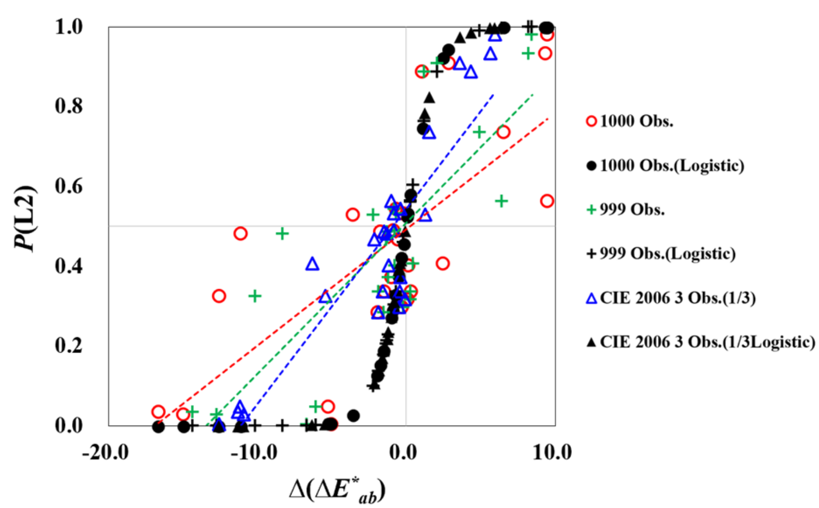


Fig. 14. Scatter plots of the observers' judgements and the differences of the calculated color differences for the group of observers derived using the three methods, together with fitted lines derived from the linear and logistic regressions.

fitted lines derived from the linear and logistic regressions. Table 10 summarizes the performance of three methods for characterizing the group of the observers. It can be clearly seen that the revised age distribution was able to improve the performance of the individual model proposed by Asano, but our proposed method still had the best performance.

Table 10. Summary of the linear regression and logistic regression, as illustrated in Fig. 14, for comparing the performance of the three methods in characterizing the group of observers

R^2 of linear regression			Loss function of logistic regression		
1000 Obs	999 Obs	CIE 2006 3 Obs.	1000 Obs	999 Obs	CIE 2006 3 Obs.
0.516	0.608	<u>0.805</u>	0.509	0.418	<u>0.321</u>

5. Conclusion

The study was designed to better understand how CMFs can characterize the perceived color differences for observers at different age groups, which was seldom investigated in the past. A series of color stimuli, with a field of view of 8.2°, were designed and calibrated using different sets of CMFs based on our past studies. Three groups of observers (i.e., young, middle-aged, and senior), with 17 observers in each age group, evaluated the color differences between the two test stimuli and the reference stimulus, and selected the one producing a larger perceived color difference. In total, 26 sets of stimuli were evaluated, and each set was evaluated five times. The results suggested that the standard CMFs (i.e., CIE 1931 2°, 1964 10°, and CIE 2006 4° CMFs) cannot characterize the perceived color differences. Instead, the CIE 2006 10°/20-year, 4°/40-year, and 2°/70-year CMFs were found to better characterize the young, middle-aged, and senior observers. Moreover, a method based on these three sets of CMFs and the distribution of the observer ages was proposed to characterize the perceived color differences of a group of observers. In short, this study clearly reveals the importance of observer age in characterizing the performance of CMFs and color specifications.

Funding. Research Grants Council, University Grants Committee (PolyU R5006-21); BIGC College's Scientific Research Project (BIGCEc202302).

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. CIE, "*Fundamental chromaticity diagram with physiological axes – part 1*," in *CIE 170-1*:2006 (CIE, 2006).
5. CIE, "*Fundamental chromaticity diagram with physiological axes – part 2: spectral luminous efficiency functions and chromaticity diagrams*," in *CIE 170-2:*2015 (CIE, 2015).
6. A. Borbely and J. Schanda, "Colour matching using LEDs as primaries," Color Res. Appl. **29**(5), 360–364 (2004).
7. X. Hu and K. W. Houser, "Large-field color matching functions," Color Res. Appl. **31**(1), 18–29 (2006).
8. Y. Asano, M. D. Fairchild, L. Blonde, and P. Morvan, "Color matching experiment for highlighting interobserver variability," Color Res. Appl. **41**(5), 530–539 (2016).
9. H. Xie, S. P. Farnand, and M. J. Murdoch, "Observer metamerism in commercial displays," J. Opt. Soc. Am. A **37**(4), A61–A69 (2020).
10. J. Li, P. Hanselaer, and A. G. Smet K, "Impact of Color-Matching Primaries on Observer Matching: Part I–Accuracy," Leukos. **18**(2), 104–126 (2022).
11. 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).
12. J. Wu and M. Wei, "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," Opt. Express **29**(25), 41731–41744 (2021).

Optics EXPRESS

13. J. M. Artigas, A. Felipe, A. Navea, A. Fandiño, and C. Artigas, "Spectral transmission of the human crystalline lens in adult and elderly persons: Color and total transmission of visible light," Invest. Ophthalmol. Vis. Sci. **53**(7), 4076–4084 (2012).
14. T. T. J. M. Berendschot, "Lens aging in relation to nutritional determinants and possible risk factors for age-related cataract," Arch. Ophthalmol. (Chicago, IL, U. S.) **120**(12), 1732 (2002).
15. D. V. Norren and J. J. Vos, "Spectral transmission of the human ocular media," Vision Res. **14**(11), 1237–1244 (1974).
16. M. Huang, Y. Xi, J. Pan, R. He, and X. Li, "Colorimetric observer categories for young and aged using paired-comparison experiments," IEEE Access **8**, 219473–219482 (2020).
17. M. Huang, Y. Wang, M. Wei, Y. Li, X. Gao, and X. Li, "Effect of observer age and stimulus size on the performance of CIE color matching functions," Opt. Express **30**(10), 16973–16986 (2022).
18. Y. Asano, "Individual colorimetric observers for personalized color imaging," PhD dissertation at Rochester Institute of Technology (2015).