Correlation Between Visual and Colorimetric Scales Ranging from Threshold to Large Color Difference†

Haisong Xu,[1,2]* Hirohisa Yaguchi,[1] Satoshi Shioiri[1]

[1] Department of Information and Image Sciences, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan

[2] State Key Laboratory of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027, China

Received 17 July 2001; accepted 30 August 2001

Abstract: Psychophysical experiments of color discrimination threshold and suprathreshold color-difference comparison were carried out with CRT-generated stimuli using the interleaved staircase and constant stimuli methods, respectively. The experimental results ranged from small (including threshold) to large color difference at the five CIE color centers, which were satisfactorily described by chromaticity ellipses as equal color-difference contours in the CIELAB space. The comparisons of visual and colorimetric scales in CIELAB unit and threshold unit indicated that the colorimetric magnitudes typically were linear with the visual ones, though with different proportions in individual directions or color centers. In addition, color difference was generally underestimated by the Euclidean distance in the CIELAB space, whereas colorimetric magnitude was perceptually underestimated for threshold unit, implying the present color system is not a really linear uniform space. Furthermore, visual data were used to test the CIELAB-based color-difference formulas. In their original forms CIEDE2000 performed a little better than CMC, followed by CIELAB, and with CIE94 showing the worst performance for the combined data set under the viewing condition in this study. © 2002 Wiley Periodicals, Inc. Col Res Appl, 27, 349–359, 2002; Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/col.10081

Key words: color difference; color discrimination threshold; color-difference comparison; method of staircase; method of constant stimuli; equal color-difference contour; color-difference formula

INTRODUCTION

Industrial color-difference evaluation has been one of the main topics in the field of color science and technology for more than three decades. Since the CIELAB color space[1] was established for promoting uniformity of color-difference practice, significant advances have been made in objective color-difference research, especially, in the last 10 years. Among these achievements are the representative color-difference formulas—CMC,[2] CIE94,[3] and the latest, CIEDE2000,[4] recently proposed by the CIE TC1-47 Hue and Lightness-Dependent Correction to Industrial Color-Difference Evaluation—and, more importantly, the generation of a series of reliable color-difference data sets, such as RIT–DuPont,[5] Witt,[6] Leeds,[7] and BFD.[8] The majority of these studies looked at small color difference because that is mainly what the color industries deal with. A few of the studies investigated large color differences, but studies looking at on small and large color differences at the same time were few until now. Today large color difference is become more and more important in applications such as color reproduction, industrial design, and color communication.

All color-difference evaluation studies have aimed toward a final goal of developing a universal standard of color-difference evaluation for most industrial applications, in accordance with the CIE guidelines[9] for further research, to improve the correlation between psychological data and prediction models. Existing color-difference formulas, orig-

†A part of this article was presented at Color 01, the 9th Congress of AIC, Rochester, New York, June 24–29, 2001.

*Correspondence to: Haisong Yu; E-mail: chsxu@zju.edu.cn

inally developed to fit some experimental data sets, can only give out satisfactory predictions of different applications within a specified range of color difference. An ideal color-difference model would presumably be expected to perform well for a range from small through moderate to large color differences. Although this goal may be impossible to achieve in practice, it deserves a big effort to reach for as an ultimate objective. The basic issue of color-difference prediction is how to make the colorimetric magnitude represent the visual one. One of the most important aspects is the correlation between color-difference scales measured or predicted by a model and their perceived counterparts, a question of much interest in building color-difference evaluation models but one that remains to be resolved.

The aim of this study was to investigate the correlation between visual scales and colorimetric magnitudes ranging from color discrimination threshold to large color difference (maximum average size was 12.0 CIELAB ΔE units). A new visual data set was obtained from specially designed psychophysical experiments of color discrimination threshold and suprathreshold color-difference comparison using CRT-generated stimuli, described below in detail.

EXPERIMENTAL

Viewing Conditions

The study was divided into two experiments, color discrimination threshold and color-difference comparison. In both experiments the test stimuli were generated on a Sony Multiscan G500 monitor controlled by a Cambridge Research Systems VSG 2/4 graphics board, with 15-bit luminance-calibrated lookup tables. Observations were performed in a darkened booth at a viewing distance of 500 mm from CRT to an observer's eyes.

The experiments were carried out in the CIELAB color space, and the five basic CIE color centers[9,10] of gray, red, yellow, green, and blue were selected as the test color centers. The CIELAB values of these color centers are listed in Table I. The stimuli to be measured were evenly distributed along 12 directions every 30° in the a^*b^* plane and along 8 directions every 45° in the a^*L^* and b^*L^* planes from the test color centers. The experiment for each color center was separated into three sessions, one for each test plane, that is, the a^*b^*, a^*L^*, and b^*L^* planes. Each session, which began with a 3-min dark adaptation and a 1-min background adaptation, lasted no more than 25 min to avoid observer fatigue. Eight observers with normal color vision took part in the experiments. Of these, 4 subjects performed observations at the gray center, with the other color centers each observed by 3 subjects. All observers were students at Chiba University in Japan and were naive to the purposes of the experiments; most had no previous experience doing such observations.

TABLE I. CIELAB chromaticity parameters of the test color centers. The CIE1931 Standard Colorimetric Observer was used in the calculations.

Color center	L^*	a^*	b^*	C^*	$h°$
Gray	61.65	0.11	0.04	0.12	20
Red	44.38	36.91	23.33	43.67	32
Yellow	86.65	−6.92	47.15	47.66	98
Green	56.09	−32.13	0.44	32.13	179
Blue	35.60	4.83	−30.18	30.56	279

Color Discrimination Threshold Experiment

Stimuli. In the color discrimination threshold experiment, the test stimulus was a square array of four 1° squares with a small black line separation of 0.1° visual angle. The squares, with a black frame of 0.1°, were presented on a 6° background set as the center color. During the experiment only one of the four squares was set as the test color; the remaining three squares held the same color as the background. The visual subtended angle of the test stimulus at the center of the CRT was 2.3°. Because it was less than 4°, the CIE1931 Standard Colorimetric Observer was used in the calculations. As illustrated in Fig. 1, the pattern was surrounded by a bright border of 8°, with a luminance of 100 cd/m^2 and a chromaticity of D65. This border was displayed to define the white point for the test pattern and to have the CRT stimuli appear as simulated surface colors[11] or related colors rather than as aperture colors. Outside the white border was a black screen.

Procedure. Every trial of observations lasted 2 s, including two periods of 200 ms of background color and black gaps at the beginning and end of the 1200-ms presentation of the test stimulus. The response time generally was less than 2 s for all observers, so observer judgment was not influenced by the limited presenting time of test stimuli under such a condition where there is a temporal gap. The background color and black gaps between trials effectively prevented the possible cues from affecting observer judgment caused by the color-changing process and an observer's adaptation to the stimulus color. During gaps all areas, including the background and the four-square array, were covered with black except for the surrounding border, which remained to hold the complete adaptation of the observer to the white point.

In each trial the test color was presented on one of the four squares selected randomly by the software, with the other three squares remaining the same color as the background. The test color was determined according to the predicted color distance from the test color center using the psychophysical staircase method. Each test color was assessed 3 times by each observer, as mentioned earlier. To avoid possible observer bias resulting from the presenting sequence of test stimuli, an interleaved staircase method was used. Each session involved a group of stimuli in four randomly selected directions with a random presenting sequence. In the initial trial of the staircase, an obviously discriminable step was presented. Then the step size decreased systematically until a criterion value, which had been determined in pilot experiments to produce an efficient

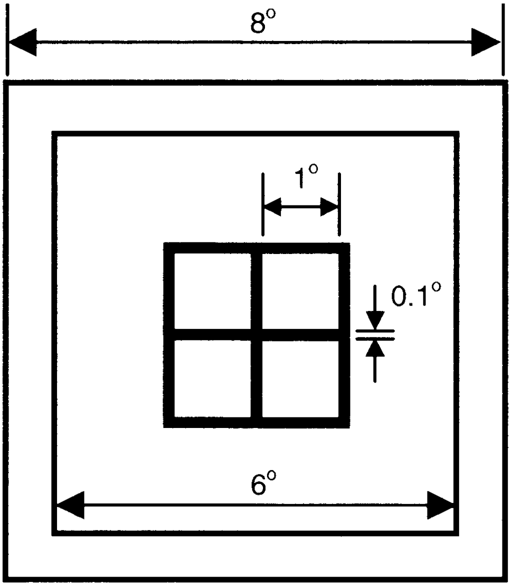


FIG. 1. The test stimulus arrangement used in color discrimination threshold experiment. A four-square array, with 0.1° separation and frame, was presented on a 6° background of the center color, surrounded by a bright border of 8° visual angle, out of which was a black screen. See text for details.

staircase, was reached that would generate 10 reversals. The averages of the 10 reversals were calculated as thresholds.

The visual task of the observer was to judge the position of the square where a color different from the background color was perceived and then to press the corresponding key on the keyboard as his or her response; this stored the result and started the next trial.

Color-Difference Comparison Experiment

Stimuli. Figure 2 illustrates the arrangement of the test stimulus pattern in the color-difference comparison experiment. The stimulus construction was almost the same as that used in the color discrimination experiment, except that at the center of the 6° background were two color pairs instead of the square array and that the background was set as Munsell N5 neutral gray with a chromaticity of D65. The two color pairs, designated as the reference and test pairs, consisted of two 1° squares in the upper and lower positions, with a black frame of 0.1° and a separation of 0.5° visual angle between them. The total visual angle of these two pairs was 2.7° (width) × 2.3° (height), so this also met the criteria for applying the CIE1931 colorimetric system to calculations.

One color of the reference pair was selected as gray with a chromaticity of D65 and a lightness (CIE Y) of 30, the same as CIE gray, and the second color differed from it along the $+L^*$ axis in the CIELAB space. Thus, the color difference of the reference pair was only the luminance variation ($+\Delta L^*$) or called gray scales along the $+L^*$ axis of the CIELAB space. For the test pairs the color differences were the selected color distances from the test color centers according to a step size predetermined by a pilot experiment. Hence, the test pairs were formed by the center colors and those stimuli evenly distributed around them in the CIELAB color space.

Procedure. In the color-difference comparison experiment, every trial of observations began with a 200-ms gap and ended on receiving the response from the observer, with no limitation for the observer's judgment time, so the length of a trial was different for individual observers. During gaps only the reference and test pairs were shut off with black, while the surrounding border and background stayed so the observer could maintain complete adaptation to the white point and background throughout the entire experiment.

This experiment was designed using the principles of the psychophysical method of constant stimuli. Thus, the measured stimuli were chosen, through pilot experiments, such that the color differences of the test pairs ranged from "always judged to be greater than the reference" to "never judged to be greater than the reference," with the majority lying between these two extremes.[10] The task of the observer was to judge visually whether the color difference of a test pair was greater or lesser than that of the reference pair and then to press one of the two keys on the keyboard

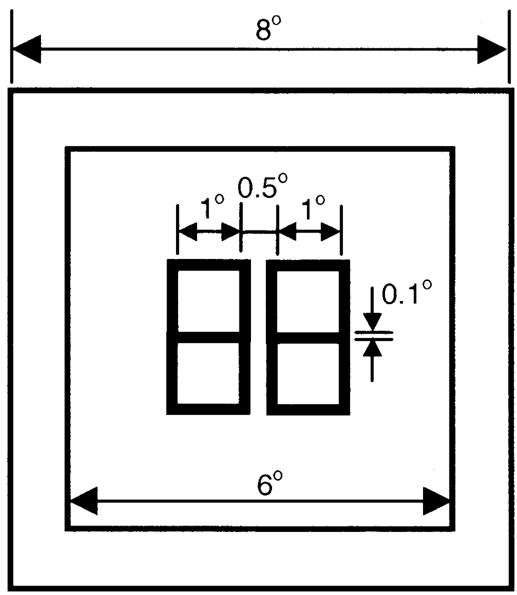


FIG. 2. The test stimulus paradigm used in the color-difference comparison experiment. The construction was almost the same as that in Fig. 1, except that two color pairs were presented at the center of the background, whose color was set as Munsell N5 instead of the test center color. See text for details.

corresponding to the two judgment categories. The judgments were repeated sufficient times to assign each test pair a probability of being judged to have a color difference greater than the reference pair. An iterative algorithm called probit analysis,[5,12–14] which is a maximum-likelihood model that relates experimental response functions to occurrence probability estimates, was used to find the most precise estimate at a tolerance of 50% rejection probability. This corresponded to the color-difference value visually equivalent to the reference color difference.

The color difference of the reference pair was set as 4.0, 8.0, and 12.0 CIELAB ΔE units, the reference scales used in the present study. Each test pair was assessed 20 times over two sessions by each observer. The presentation of the color stimuli—the arrangement of the left and right positions of the reference and test pair and of the upper and lower positions of the two color pairs—was randomly ordered by the computer program during the experiment's various trials to avoid an eventual judgment bias of the observers resulting from the presenting positions of test stimuli. In each direction about every color center for each reference scale, nine test pairs, corresponding to nine color stimuli distributed that direction, were compared with the reference pair. Using the statistical method of probit analysis, the equal color-difference contours for each reference visual scale, 4.0, 8.0, and 12.0 CIELAB ΔE units, were obtained, which will be analyzed next.

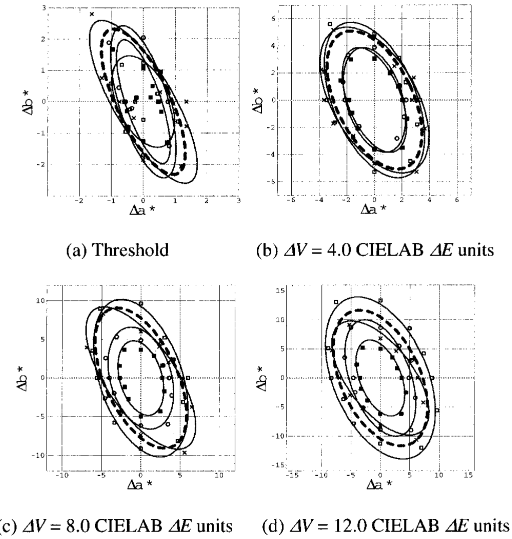


FIG. 3. Chromaticity ellipses in the a^*b^* plane at CIE gray color center corresponding to (a) threshold and reference visual scales of (b) 4.0, (c) 8.0, and (d) 12.0 CIELAB ΔE units. The raw data are plotted in different symbols for 4 observers: YT (○), NM (×), KF (□), and HT (■), and the dashed line in each panel indicates the ellipse fitted to the pooled data of all observers: (a) threshold, (b) ΔV = 4.0 CIELAB ΔE units, (c) ΔV = 8.0 CIELAB ΔE units, (d) ΔV = 12.0 CIELAB ΔE units.

RESULTS AND DISCUSSION

Validity of Visual Data

Although it is believed that an ellipse is not necessarily the best mathematical model for conceptualizing the visual data of color discrimination,[15,16] the experimental results in the present study were first plotted as chromaticity ellipses in the CIELAB space for comparison. This was done not only because ellipses have been widely employed to describe the contours of equally perceived color differences around a given center in a color space[17–20] since the pioneering work of MacAdam[21] but also because the thresholds and equal color-difference contours obtained from the experiments in this study can be well represented and summarized in ellipses. An example of such chromaticity ellipses in the a^*b^* plane at the gray center is given in Fig. 3.

As Nagy *et al.* pointed out,[22] it has proved difficult to compare discrimination contours from different observers and different chromaticities in the CIE diagram because of the multivariate nature of this comparison. Fortunately, Wyszecki and Fielder[23] proposed a simple and meaningful solution to this problem: to pool all the data from different occasions and fit a single ellipse to the pooled data rather than to compute means. Thus, the chromaticity ellipses fitted to the pooled data from all observers in the same viewing condition, such as the dashed ellipses shown in Fig. 3, were used in the following analysis. In addition, a performance factor (*PF*), first devised by Luo and Rigg[24] and then modified to *PF*/3 by Guan and Luo[25] as given in Eq. (1), was adopted in the present study to ease the comparison between two sets of data.

$$PF/3 = 100\left[(\gamma - 1) + V_{AB} + \frac{CV}{100}\right]\bigg/3, \qquad (1)$$

where

$$CV = \frac{\sqrt{\frac{1}{N}\sum(X_i - fY_i)^2}}{\bar{X}} \times 100, \qquad (2)$$

and

$$f = \frac{\sum X_iY_i}{\sum Y_i^2},$$

$$\log_{10}\gamma = \sqrt{\frac{1}{N}\sum\left[\log_{10}\left(\frac{X_i}{Y_i}\right) - \overline{\log_{10}\left(\frac{X_i}{Y_i}\right)}\right]^2}, \qquad (3)$$

$$V_{AB} = \sqrt{\frac{1}{N}\sum\frac{(X_i - FY_i)^2}{X_iFY_i}}, \qquad (4)$$

and

$$F = \sqrt{\sum\frac{X_i}{Y_i}\bigg/\sum\frac{Y_i}{X_i}},$$

where N is the number of compared pairs and X_i and Y_i are values of pair i. A higher *PF*/3 value implies worse agree-

TABLE II. Chromaticity ellipse fitting accuracy in correlation coefficient (*r*) and PF/3 measures.

Test plane	*a*b**		*a*L**		*b*L**		All planes	
Color center	*r*	PF/3	*r*	PF/3	*r*	PF/3	*r*	PF/3
Gray	0.992	16.8	0.956	18.6	0.967	18.1	0.972	18.3
Red	0.968	19.9	0.980	16.8	0.967	18.3	0.971	19.0
Yellow	0.972	19.5	0.986	14.7	0.989	11.7	0.981	16.4
Green	0.975	22.2	0.979	19.3	0.956	21.1	0.971	21.1
Blue	0.985	17.6	0.981	17.8	0.986	12.9	0.985	16.5

ment between data sets. For example, a *PF*/3 of 30 indicates a disagreement of about 30%.

Observer Accuracy

As mentioned above, pooled ellipses were used to represent the color-difference contours, so the agreement between the results of individual observers and the calculated values according to the equations of pooled ellipses can be considered the interobserver accuracy. The mean interobserver error was found to be about 30 *PF*/3 units, with a range from 20 *PF*/3 units, for the most accurate observer, to 47 *PF*/3 units, for the least accurate. This corresponds to a standard deviation for the pooled ellipse in each situation of about 10.6 *PF/3* units ($30/8^{1/2}$) for the 8 observers.

Compared with related studies, this observer accuracy falls between the results of Guan and Luo using small color differences[25] (40 *PF/3* units) and large color differences[26,27] (27 and 24 *PF/3* units) because the present study involved color differences from threshold to large suprathreshold. Thus, this observer variation can be considered typical for such experiments, confirming the believability and validity of the experimental results in this study.

Chromaticity Ellipse-Fitting Accuracy

The errors fitting the chromaticity ellipses to the visual data of individual observers for each test plane and all color centers were also calculated, as listed in Table II, as *PF*/3 counts and correlation coefficients (*r*). At the gray center the fitting accuracy in the a^*b^* plane was better than in the a^*L^* and b^*L^* planes, but this was not true for the other color centers. The fitting error for the gray center in the a^*b^* plane was the smallest among all the color centers. However, the total errors for all planes, which ranged from 16.4 to 21.1 *PF*/3 units, are quite good compared with the observer accuracy of 30 units, and all correlation coefficients were greater than 0.95. Thus, it can be said that the ellipse is appropriate and effective as a mathematical model to describe the visual data of the present study.

Equal Color-Difference Contours

The parameters of pooled ellipses representing the equal color-difference contours were expressed in the semimajor axis (*A*), the ratio of semiaxes (*A/B*), the orientation angle (θ), and the square root of the ellipse area, as given in Table III for the a^*b^* plane. Table III also lists the ratios of $(\pi AB)^{1/2}$ of suprathreshold contours relative to threshold contours at each color center. From a comparison with Witt's threshold ellipses[19] using surface colors, done by calculating the corresponding *A/B* and θ parameters in the a^*b^* plane from the ellipsoid coefficients without gap and lightness of achromatic surround, it was found that the present threshold ellipses were more elongated at the gray and red centers but the opposite at the yellow, green, and blue centers, while the orientations (θ) were similar to some degree. For the suprathreshold color differences the chromaticity ellipses in this study were somewhat more elongated than those of Guan and Luo[25,27] on surface colors, but the orientations (θ) basically fell between their results for small and large color differences. A number of studies[11,12,19,25,27] on parametric effects on color-difference evaluation have shown that color discrimination is influenced by the temporal characteristics[28] and spatial conditions[29] of stimulus presentation, such as exposure time, gap, and background color. Hence, the discrepancies between the present results and these other studies may be a result of the different spatial and temporal conditions of the stimulus pattern and the different stimulus color modes, that is, CRT-generated stimuli and surface colors. In addition, the chromaticity parameters of the test color centers and the psychophysical methods of color-difference comparison employed in these studies were also different.

At a given color center, the orientations (θ) and shapes (*A/B*) of ellipses corresponding to threshold and all visual scales were almost identical except for the green center, for which the parameters were unstable and the ellipse fitting error was the largest (21.1 *PF*/3 units) among the five centers. The shapes (*A/B*) of the threshold contours for the gray, red, and green centers were more elongated than those of the suprathreshold ones but not for the yellow and blue centers, the opponent colors of the yellow-blue mechanism. The ellipse-fitting accuracies of the yellow and blue centers, the brightest and darkest colors among the five centers, were the best (16.4 and 16.5, respectively) of all the color centers. Moreover, the *A/B* values of the blue center were almost the same for the threshold and various visual scales and in general were the maximums among all five centers, indicating that the blue region is very different from other color regions in the CIELAB color space.

The areas of threshold ellipses were close to each other in size for all color centers, with that of the blue center the smallest, whereas the sizes of the suprathreshold ellipses

TABLE III. Parameters of pooled chromaticity ellipses in a^*b^* plane for threshold and visual scales of 4.0, 8.0, and 12.0 CIELAB ΔE units.

Color center	Visual scale	A	A/B	θ(deg)	$(\pi AB)^{1/2}$	Ratio[a]
Gray	Threshold	2.53	2.98	115	2.60	
	4.0	5.46	1.96	115	6.91	2.66
	8.0	9.73	2.10	114	11.90	4.58
	12.0	12.47	1.84	114	16.28	6.26
Red	Threshold	3.19	2.84	76	3.36	
	4.0	8.79	1.61	75	12.28	3.65
	8.0	14.68	1.69	83	20.02	5.96
	12.0	19.64	1.93	78	25.07	7.46
Yellow	Threshold	2.80	2.04	92	3.47	
	4.0	9.19	2.00	103	11.53	3.32
	8.0	16.50	2.25	98	19.50	5.62
	12.0	23.37	2.31	101	27.24	7.85
Green	Threshold	3.26	3.42	136	3.12	
	4.0	13.67	2.11	169	16.69	5.35
	8.0	17.24	1.68	153	23.60	7.56
	12.0	24.38	1.80	135	32.19	10.32
Blue	Threshold	2.11	2.67	110	2.29	
	4.0	10.29	2.85	120	10.81	4.72
	8.0	15.37	2.75	123	16.43	7.17
	12.0	18.30	2.66	120	19.89	8.69

[a] Ratio of $(\pi AB)^{1/2}$ of the chromaticity contour corresponding to ratio of suprathreshold color difference to that of threshold at each color center.

were very different for individual centers. This indicates that the color-difference metrics in CIELAB space is not uniform perceptually. In addition, the $(\pi AB)^{1/2}$ ratios of suprathreshold to threshold contours increased with the visual scales proportionately in some way. Among all color centers, the areas of equal color-difference contours at the gray and blue centers were the smallest, which should be because of the achromatic background of Munsell N5, making the largest perceived differences (with the highest visual sensitivities) of both lightness and chromatic stimuli for the gray and blue centers, the nearest colors to the background.[30]

The shapes of all contours were ellipses, not circles, which shows worse local uniformity in the individual color regions. On the other hand, the areas of the ellipse for all color centers were different from each other. The areas at the green center usually were the largest, followed by red and yellow, both of which were very similar in size, and then the blue center, with the contour areas for the gray center the smallest. This shows that the CIELAB color space lacks overall uniformity, as can be seen from the pooled ellipses in Fig. 4.

Comparison of Visual Scales and Colorimetric Magnitudes

To compare the color-difference metrics with the perceived ones, the colorimetric magnitudes (ΔE) in each test direction of all measurement planes for the five color centers were plotted against the corresponding visual scales (ΔV) of 4.0, 8.0, and 12.0 CIELAB ΔE units. Figure 5 illustrates some of the plots for the a^* axis at the red and green centers and for the b^* axis at the yellow and blue centers in the a^*b^* plane. The comparisons were calculated in CIELAB ΔE and threshold units. It was found that the visual data could be linearly fitted, no matter which unit was used, CIELAB ΔE or threshold, for any in the same direction with high correlation coefficients. To facilitate the discussion, the regression lines fitted to the pooled data of individual observers in each direction were calculated, shown by the thick lines in Fig. 5. Table IV summaries the linear regression results only for the a^*, b^*, and L^* axes in all test planes. Almost all the correlation coefficients for individual observers in all directions, including those not listed in Table IV, were greater than 0.9, implying that linear fitting was both appropriate and sufficient for these visual data.

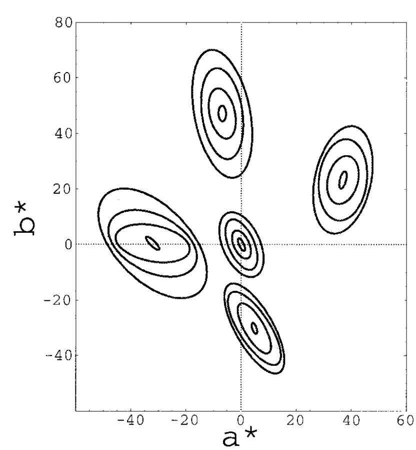
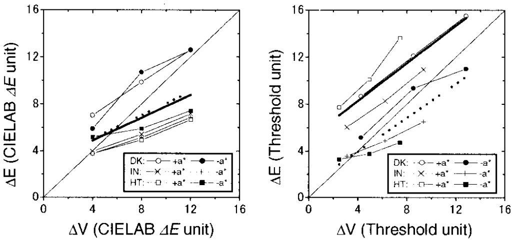
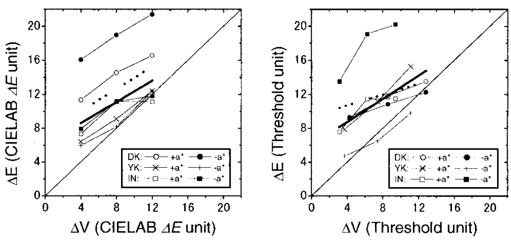


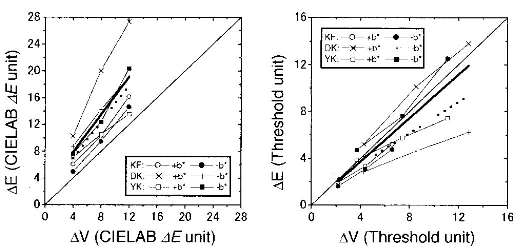
FIG. 4. Equal color-difference contours for all color centers plotted in the a^*b^* diagram.



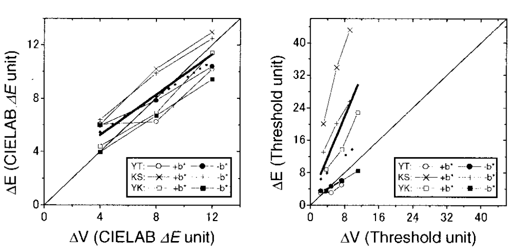
(a) a^* axis in a^*b^* -plane for CIE red center



(b) a^* axis in a^*b^* -plane for CIE green center



(c) b^* axis in a^*b^* -plane for CIE yellow center



(d) b^* axis in a^*b^* -plane for CIE blue center

FIG. 5. Comparison of colorimetric magnitudes against visual scales, in CIELAB ΔE and threshold units, for a^* axis at the red and green centers and for the b^* axis at the yellow and blue centers in the a^*b^* plane. The thick solid lines are the linear regression lines fitted to the pooled data from all observers in each positive direction, while the thick dotted lines are for the negative directions. The diagonal thin lines with slopes of 45° represent the ideal relation of visual and colorimetric scales. (a) a^* axis in the a^*b^* plane for the CIE red center, (b) a^* axis in the a^*b^* plane for the CIE green center, (c) b^* axis in the a^*b^* plane for the CIE yellow center, (d) b^* axis in the a^*b^* plane for the CIE blue center.

In general, the slopes of the regression lines for all directions were far from 45°, which represents the ideal relationship of the visual and colorimetric scales. Only the fitting lines in the $+L^*$ direction in both the a^*L^* and b^*L^* planes at the gray center were satisfactory, coinciding with the ideal 45° line, as shown in Fig. 6. This occurred because only in this case were the reference and test pairs completely in the same state. Hence, the colorimetric magnitudes of the visual scales studied here, whose range went from moderate to large color differences, were typically linear with the visual scales, but the slopes were different for individual directions, a finding that essentially is consistent with the results in the studies of Witt[6,31] about small color differences.

For the CIELAB ΔE unit, to some degree, most lines in opposite directions had the same slopes—such as the $\pm a^*$ and $\pm b^*$ directions at the gray and yellow centers, the $\pm a^*$ directions at the red and green centers, and the $\pm L^*$ directions at the red and yellow centers—so they were nearly parallel; this was not true for the blue center. Hence, the symmetry of CIELAB space in the blue region is worse than in other color regions. For the L^* axis the regression lines in the $-L^*$ directions at the gray and blue centers, shown in Figs. 6 and 7, were obviously above those in the $+L^*$ directions, but it was the opposite for the red center, as can be seen in Fig. 8, in which the $-L^*$ fitting lines were below the $+L^*$ lines. Furthermore, at the green center the $-L^*$ fitting line was below the $+L^*$ line in the a^*L^* plane, whereas above the $+L^*$ line in the b^*L^* plane, as shown in Fig. 9. Only for the yellow center, shown in Fig. 10, did the regression lines in the $\pm L^*$ directions nearly coincide. Thus, except for the yellow center, the equally perceived color-difference contours were not symmetric about the a^* or b^* axes in the a^*L^* and b^*L^* planes. This shows the perceptual asymmetric effects of constant stimulus changes between $+\Delta L^*$ and $-\Delta L^*$ directions.

Most regression lines calculated in CIELAB ΔE units were below the 45° line, especially in the a^*b^* plane at the gray, blue, and red centers, whereas nearly the opposite was true for the threshold unit. This means the color difference was underestimated by the Euclidean distance compared with the visual scales in the CIELAB space, whereas in the relative comparison using the threshold unit, the colorimetric magnitude was perceptually underestimated, a finding consistent with other studies.[6,31] Therefore, the unit value representing the color-difference metrics is important, which is related to the visual uniformity of a color space. For opponent colors there was no robust rule found in the a^* axes at the red and green centers and in the b^* axes at the yellow and blue centers. However, almost all the fitted lines did not point to the origin of the axes but at the colorimetric ΔE axis, with a small positive intercept, which reflected the influence of the color discrimination threshold.

For the threshold unit perceived color differences were linear with the threshold steps, but in general the regression lines did not coincide with the ideal diagonal lines, as expected except for the yellow center, whose fitted lines were nearest to the 45° lines among all centers. This implies

TABLE IV. Summary of linear regressions between visual and colorimetric scales in a^*, b^*, and L^* axes of each test plane for all color centers.

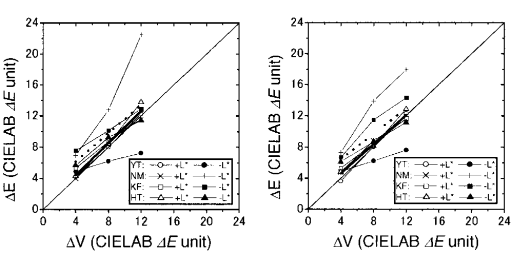
Color center	Test plane	Test direction	CIELAB ΔE unit		Threshold unit		Correlation coefficient (r)[a]
			Slope (deg)	Intercept	Slope (deg)	Intercept	
Gray	a^*b^*	$+a^*/-a^*$	22/21	0.98/1.24	49/54	1.52/3.99	0.976–1.000/0.958–1.000
		$+b^*/-b^*$	29/29	1.84/2.27	26/41	2.71/2.66	0.950–0.998/0.959–1.000
	a^*L^*	$+a^*/-a^*$	22/21	1.31/1.59	35/44	4.39/4.05	0.950–0.999/0.975–0.999
		$+L^*/-L^*$	46/42	0.34/2.53	44/33	0.21/4.13	0.989–0.999/0.990–1.000
	b^*L^*	$+b^*/-b^*$	30/29	1.87/1.96	38/38	2.73/2.20	0.967–1.000/0.929–0.996
		$+L^*/-L^*$	43/39	0.87/3.15	47/17	0.19/8.34	0.995–1.000/0.989–1.000
Red	a^*b^*	$+a^*/-a^*$	26/27	2.94/3.04	39/36	5.06/1.07	0.994–1.000/0.970–0.998
		$+b^*/-b^*$	50/33	4.15/4.64	31/21	2.35/1.06	0.967–0.986/0.917–0.985
	a^*L^*	$+a^*/-a^*$	24/23	2.67/3.34	40/42	1.60/0.45	0.966–0.988/0.948–1.000
		$+L^*/-L^*$	24/22	1.68/1.55	25/25	2.92/6.97	0.962–0.998/0.940–0.987
	b^*L^*	$+b^*/-b^*$	48/32	5.21/5.20	25/20	2.94/2.38	0.889–0.972/0.899–0.968
		$+L^*/-L^*$	23/22	2.46/1.90	41/33	1.43/2.47	0.972–0.997/0.941–1.000
Yellow	a^*b^*	$+a^*/-a^*$	35/34	0.81/0.91	46/37	0.75/1.06	0.946–1.000/0.989–1.000
		$+b^*/-b^*$	54/54	2.31/1.49	43/34	0.01/0.82	0.996–0.999/0.989–0.999
	a^*L^*	$+a^*/-a^*$	30/29	1.76/1.32	43/39	3.84/1.80	0.965–1.000/0.962–0.996
		$+L^*/-L^*$	35/36	2.07/1.53	36/36	0.06/0.20	0.999–0.999/0.985–0.999
	b^*L^*	$+b^*/-b^*$	49/53	3.20/1.11	29/33	2.43/1.83	0.991–1.000/0.994–1.000
		$+L^*/-L^*$	40/45	1.59/−0.07	43/34	−0.96/0.39	0.991–0.997/0.988–1.000
Green	a^*b^*	$+a^*/-a^*$	32/33	6.09/7.42	34/18	5.99/9.45	0.867–0.998/0.932–0.999
		$+b^*/-b^*$	43/28	1.09/2.74	55/35	−1.28/1.78	0.983–1.000/0.934–0.997
	a^*L^*	$+a^*/-a^*$	32/32	6.54/5.28	29/20	7.28/5.97	0.984–1.000/0.958–0.999
		$+L^*/-L^*$	34/26	2.14/2.92	27/42	1.93/2.57	0.947–0.996/0.964–0.996
	b^*L^*	$+b^*/-b^*$	47/29	1.13/3.42	57/34	−0.88/3.36	0.986–1.000/0.974–1.000
		$+L^*/-L^*$	34/40	2.57/1.99	38/60	1.96/−0.68	0.985–0.995/0.983–1.000
Blue	a^*b^*	$+a^*/-a^*$	21/27	2.52/2.06	52/43	3.88/7.69	0.962–0.997/0.969–0.993
		$+b^*/-b^*$	37/34	2.23/2.79	68/45	1.56/4.00	0.885–0.994/0.996–1.000
	a^*L^*	$+a^*/-a^*$	29/33	0.55/0.67	69/53	−2.87/2.61	0.947–1.000/0.950–0.982
		$+L^*/-L^*$	31/41	1.77/0.82	33/59	1.42/0.59	0.963–1.000/0.945–1.000
	b^*L^*	$+b^*/-b^*$	37/33	1.19/2.65	63/47	−1.01/3.76	0.907–0.988/0.951–0.985
		$+L^*/-L^*$	28/32	2.68/2.43	38/49	2.15/3.34	0.971–0.996/0.949–0.999

[a] The range of correlation coefficients for individual observers in each case.

that the metrics adopted in the present color system are not really linear or uniform but only subadditive with respect to the threshold step. Therefore, as proposed by Kuehni,[32] it is truly needed that color-difference metrics be rooted in a sound theory based on deep insight into the mysteries of human color vision.

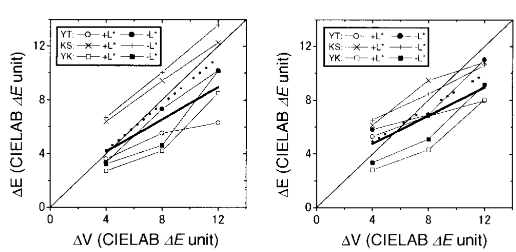
Evaluation of Color-Difference Formulas

The visual data obtained from the present experiments were also used to test the three advanced color-difference formulas—CMC, CIE94, and CIEDE2000—and the basic CIELAB equation. Comparisons between color differences (ΔE) predicted by individual formulas with their original forms—that is, $k_L = k_C = k_H = 1$—and the corresponding



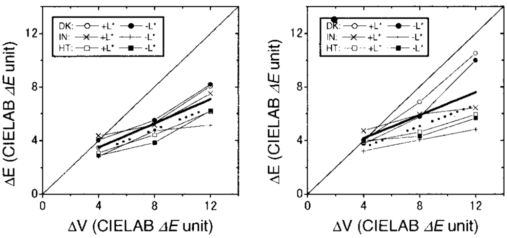
(a) a^*L^* -plane (b) b^*L^* -plane

FIG. 6. Relation of visual and colorimetric scales in CIELAB ΔE units for L^* axis at the CIE gray center: (a) a^*L^* plane, (b) b^*L^* plane.



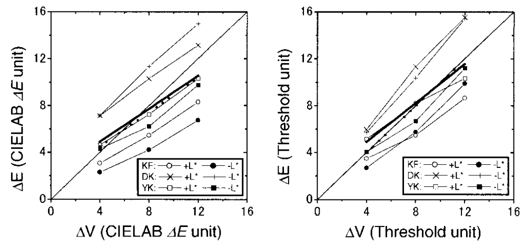
(a) a^*L^* -plane (b) b^*L^* -plane

FIG. 7. Visual data, as in Figure 6, for the L^* axis in the CIE blue center: (a) a^*L^* plane, (b) b^*L^* plane.



(a) a^*L^* -plane (b) b^*L^* -plane

FIG. 8. Visual data, as in Fig. 6, for the L^* axis in the CIE red center: (a) a^*L^* plane, (b) b^*L^* plane.

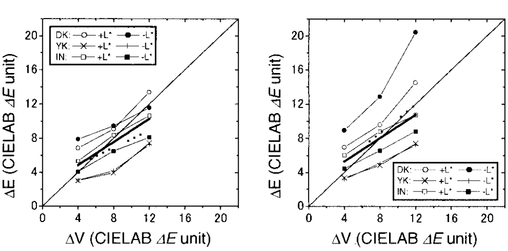


(a) a^*L^* -plane (b) b^*L^* -plane

FIG. 10. Visual data, as in Fig. 6, in L^* axis for the CIE yellow center: (a) a^*L^* plane, (b) b^*L^* plane.

visual scales (ΔV) were carried out in terms of an *PF*/3 measure. The resulting performances were sorted for all sub- and combined data sets of separate test planes at all color centers and were ordered most accurate to least accurate predictions with respect to *PF*/3 units, as listed in Table V.

The *PF*/3 values in the a^*b^* plane were different from in the a^*L^* or the b^*L^* plane for all formulas, implying imbalances in the four formulas for predicting lightness and chromatic differences. For instance, the CIEDE2000 outperformed all other equations in the a^*b^* plane but not in the a^*L^* and b^*L^* planes for its accurately predictable gray, yellow, and blue centers, so its relative performance in predicting lightness differences was not as qualified as well as it was for chromatic ones under the viewing condition in this study. At the green center in any test plane, the CIEDE2000 and CIE94 were worse than the CMC and CIELAB, though the *PF*/3 differences were not as large. This shows that the performances of the two newer formulas in the green region were not improved over the two older ones. For all formulas the performances at the blue and red centers were almost the worst for all color centers, so red and blue should be the most difficult colors for visual prediction. However, the visual data at the blue center were well predicted by CIEDE2000 with an accuracy (24.8 *PF*/3 units) that was—though a little worse than for yellow (19.6) and gray (21.4)—obviously better than that predicted by CIELAB (33.5), CMC (36.2), and CIE94 (38.8). This confirms that the rotation item involved in the CIEDE2000 equation effectively improves the uniformity and predicting performance for the color differences in the blue region. In general, the performances of all formulas at the gray and yellow centers were better than those at the red and blue ones, with the green center between them. For the combined data set, the CIEDE2000 (32.6 *PF*/3 units), with its complexity of calculation, performed slightly better than CMC (32.8), followed by CIELAB (38.0) and then CIE94 (39.4), which performed the worst. The CMC performed best only at the green center, but its strong ability to predict large color differences made it able to perform well for the combined data set of all five centers, which ranged from moderate to large color difference. The CIE94 predicted visual data the worst at all centers except for the red center, the least accurate center for CIEDE2000 and other formulas. However, given the relatively small observer variation (30 *PF*/3 units), it did not perform satisfactorily for any of the four color-difference formulas.



(a) a^*L^* -plane (b) b^*L^* -plane

FIG. 9. Visual data, as in Fig. 6, for the L^* axis in the CIE green center: (a) a^*L^* plane, (b) b^*L^* plane.

The distributions of color pairs predicted to scatter in successive intervals of 1 ΔE unit across the whole data set by individual color-difference formulas, shown in Fig. 11, told the same story. The predicted color differences concentrated better for CIEDE2000 than for CMC and better for CIE94 than for CIELAB around the corresponding visual ones of 4.0 CIELAB ΔE units. But the peaks corresponding to visual scales of 8.0 and 12.0 CIELAB ΔE units were farther from the visual ones for CIEDE2000 than they were for CMC and farther from the visual ones for CIE94 than they were for CIELAB, whereas among all the formulas CIELAB had the widest scattering range of color pairs.

CONCLUSIONS

An investigation was undertaken of the correlation between visual scales and colorimetric magnitudes for ranges from color discrimination threshold to large color difference. The color discrimination threshold and color-difference comparison experimental designs were based, respectively, on the interleaved staircase and constant stimuli psychophysical methods, using CRT-generated stimuli around the five CIE color centers. The resulting equal color-difference contours

TABLE V. Performance sorting of color-difference formulas by PF/3 measure. The number in parentheses after each item indicates the corresponding PF/3 units.

Color center	Test plane	Order of color-difference formulae from best to worst performance			
Gray	a^*b^*	CIEDE2000 (20.4)	CMC (25.0)	CIE94 (26.6)	CIELAB (26.8)
	a^*L^*	CIEDE2000 (22.8)	CMC (23.2)	CIELAB (33.5)	CIE94 (34.7)
	b^*L^*	CMC (16.8)	CIELAB (16.8)	CIEDE2000 (17.6)	CIE94 (18.0)
	All planes	CIEDE2000 (21.4)	CMC (23.4)	CIELAB (29.7)	CIE94 (30.7)
Red	a^*b^*	CIELAB (27.0)	CIE94 (36.2)	CIEDE2000 (36.9)	CMC (37.2)
	a^*L^*	CMC (19.7)	CIELAB (21.0)	CIEDE2000 (23.1)	CIE94 (27.5)
	b^*L^*	CIE94 (19.7)	CIEDE2000 (22.0)	CMC (27.3)	CIELAB (39.8)
	All planes	CIE94 (34.2)	CIEDE2000 (34.2)	CMC (35.9)	CIELAB (39.2)
Yellow	a^*b^*	CIEDE2000 (13.9)	CIE94 (16.4)	CMC (18.0)	CIELAB (31.2)
	a^*L^*	CIEDE2000 (18.2)	CIELAB (19.8)	CMC (22.4)	CIE94 (36.0)
	b^*L^*	CMC (23.5)	CIELAB (24.6)	CIEDE2000 (25.9)	CIE94 (39.1)
	All planes	CIEDE2000 (19.6)	CMC (22.4)	CIELAB (30.0)	CIE94 (38.1)
Green	a^*b^*	CMC (28.0)	CIELAB (28.8)	CIEDE2000 (29.2)	CIE94 (30.0)
	a^*L^*	CIELAB (28.0)	CMC (28.2)	CIEDE2000 (32.6)	CIE94 (35.7)
	b^*L^*	CIELAB (16.0)	CMC (19.5)	CIEDE2000 (20.1)	CIE94 (21.8)
	All planes	CMC (26.6)	CIELAB (27.7)	CIEDE2000 (29.5)	CIE94 (31.8)
Blue	a^*b^*	CIEDE2000 (23.1)	CMC (35.7)	CIE94 (37.0)	CIELAB (43.4)
	a^*L^*	CIELAB (21.4)	CIEDE2000 (21.8)	CMC (29.9)	CIE94 (33.0)
	b^*L^*	CIELAB (19.8)	CIEDE2000 (25.6)	CMC (36.5)	CIE94 (38.3)
	All planes	CIEDE2000 (24.8)	CIELAB (33.5)	CMC (36.2)	CIE94 (38.8)
All centers	All planes	CIEDE2000 (32.6)	CMC (32.8)	CIELAB (38.0)	CIE94 (39.4)
ΔE formula		Order of color centers predicted with highest to lowest accuracy			
CIELAB	green (27.7)	gray (29.7)	yellow (30.0)	blue (33.5)	red (39.2)
CMC	yellow (22.4)	gray (23.4)	green (26.6)	red (35.9)	blue (36.2)
CIE94	gray (30.7)	green (31.8)	red (34.2)	yellow (38.1)	blue (38.8)
CIEDE2000	yellow (19.6)	gray (21.4)	blue (24.8)	green (29.5)	red (34.2)

were well represented by using a mathematical model of chromaticity ellipses, which were found to extend in a regular way with the increase of visual scales. The results for the ellipse contours and their different areas at individual color centers imply that there is a lack of local and overall uniformity of color-difference metrics in the CIELAB space.

The visual and colorimetric scales were compared in CIELAB ΔE and threshold units. A detailed analysis showed that the colorimetric magnitudes were typically linear with the visual ones, but not in the same ratio for different directions and color centers. In the CIELAB space most of the opposite directions had symmetric color-differ-

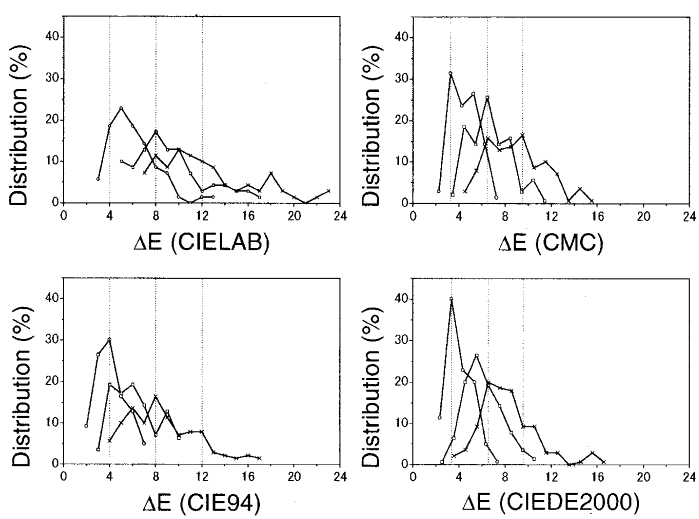


FIG. 11. Distribution of color pairs, in percentage of the total pairs, predicted to scatter in the successive intervals of 1 ΔE unit across the whole data set by individual color-difference formulas. The dotted vertical lines indicate the predicted color differences corresponding to the three reference visual scales of 4.0 (○), 8.0 (□), and 12.0 (×) CIELAB ΔE units, respectively.

ence metrics, except in the $\pm\Delta L^*$ directions because of the asymmetry of visual characterization between $+\Delta L^*$ and $-\Delta L^*$ directions in the studied color space. For the visual data obtained in this study, the Euclidean distance in the CIELAB space underestimated the color difference, whereas when using the threshold unit, the colorimetric magnitude was perceptually underestimated. Hence, the metrics of the present color system are only subadditive, at most, for the threshold step—they are not a really linear uniform space viewed from this point.

Three advanced color-difference formulas, CMC, CIE94, and CIEDE2000, together with the basic CIELAB equation, were tested using the obtained visual data in their original forms of $k_L = k_C = k_H = 1$. All formulas performed better at the gray and yellow centers than at red and blue ones, with the green center between them. For the combined data set of all color centers under the viewing condition of the present study, the CIEDE2000 performed a little better than the CMC, followed by CIELAB, with CIE94 faring the worst.

ACKNOWLEDGMENTS

The authors thank Professor M. R. Luo for providing technical guidance and all observers for their cooperation. The authors are also full of gratitude to the anonymous referees for their detailed and thoughtful comments, which helped to improve the article.

1. CIE. Colorimetry. 2nd ed. CIE Publication No. 15.2. Vienna: Central Bureau of the CIE; 1986.
2. Clarke FJJ, McDonald R, Rigg B. Modification to JPC79 colour-difference formula. J Soc Dyers Col 1984;100:128–132.
3. CIE. Technical report: Industrial colour-difference evaluation. CIE Publication No. 116. Vienna: Central Bureau of the CIE; 1995.
4. Luo MR, Cui G, Rigg B. The development of the CIE 2000 colour difference formula: CIEDE2000. Color Res Appl 2001;26:340–350.
5. Berns RS, Alman DH, Reniff L, Snyder GD, Balonon-Rosen MR. Visual determination of suprathreshold color-difference tolerances using probit analysis. Color Res Appl 1991;16:297–316.
6. Witt K. Geometric relations between scales of small colour differences. Color Res Appl 1999;24:78–92.
7. Kim DH, Nobbs JH. New weighting functions for the weighted CIELAB colour difference formula. Proc Color 97 Kyoto 1997;1:446–449.
8. Luo MR, Rigg B. BFD (ℓ:c) colour-difference formula. Part I—Development of the formula. J Soc Dyers Col 1987;103:86–94.
9. Witt K. CIE guidelines for coordinated future work on industrial colour-difference evaluation. Color Res Appl 1995;20:399–403.
10. Robertson AR. CIE guidelines for coordinated research on colour-difference evaluation. Color Res Appl 1978;3:149–151.
11. Indow T, Robertson AR, von Grunau M, Fielder GH. Discrimination ellipsoids of aperture and simulated surface colors by matching and paired comparison. Color Res Appl 1992;17:6–23.
12. Strocka D, Brockes A, Paffhausen W. Influence of experimental parameters on the evaluation of color-difference ellipsoids. Color Res Appl 1983;8:169–175.
13. Alman DH, Berns RS, Snyder GD, Larsen WA. Performance testing of color-difference metrics using a color tolerance data set. Color Res Appl 1989;14:139–151.
14. Indow T, Morrison ML. Construction of discrimination ellipsoids for surface colors by the method of constant stimuli. Color Res Appl 1991;16:42–56.
15. Stromeyer III CF, Cole GR, Kronauer RE. Second-site adaptation in the red-green chromatic pathways. Vision Res 1985;25:219–237.
16. Cowan WB, Wyszecki G, Yaguchi H. Probability summation among color channels. J Opt Soc Am A 1984;1:1307.
17. Rich DC, Billmeyer Jr FW. Small and moderate color differences. IV. Color-difference-perceptibility ellipses in surface-color space. Color Res Appl 1983;8:31–39.
18. Cheung M, Rigg B. Color difference ellipsoids for five CIE colour centers. Color Res Appl 1986;11:185–195.
19. Witt K. Parametric effects on surface color-difference evaluation at threshold. Color Res Appl 1990;15:189–199.
20. Melgosa M, Hita E, Poza AJ, Alman DH, Berns RS. Suprathreshold color-difference ellipsoids for surface colors. Color Res Appl 1997; 22:148–155.
21. MacAdam DL. Visual sensitivities to color differences in daylight. J Opt Soc Am 1942;32:247–274.
22. Nagy AL, Eskew RT, Boynton RM. Analysis of color-matching ellipses in a cone-excitation space. J Opt Soc Am A 1987;4:756–768.
23. Wyszecki G, Fielder GH. New color matching ellipses. J Opt Soc Am 1971;61:1135–1152.
24. Luo MR, Rigg B. BFD (ℓ:c) colour-difference formula. Part II—Performance of the formula. J Soc Dyers Col 1987;103:126–132.
25. Guan SS, Luo MR. Investigation of parametric effects using small colour differences. Color Res Appl 1999;24:331–343.
26. Guan SS, Luo MR. A color-difference formula for assessing large colour differences. Color Res Appl 1999;24:344–355.
27. Guan SS, Luo MR. Investigation of parametric effects using large colour differences. Color Res Appl 1999;24:356–368.
28. Yaguchi H, Mohri Y, Ishiwata M, Shioiri S. Temporal characteristics of colour discrimination. In: Dickinson C, Murray I, Carden D, editors. John Dalton's colour vision legacy. London: Taylor & Francis; 1997. p 173–177.
29. Yaguchi H, Masuda I, Shioiri S, Miyake Y. Analysis of the color discrimination data in the physiological based color space. Die Farbe 1993; 39:105–114.
30. Kuehni RG. Color space and its divisions. Color Res Appl 2001;26: 209–222.
31. Witt K. Linearity and additivity of small color differences. Color Res Appl 1995;20:36–43.
32. Kuehni RG. Industrial color difference: Progress and problems. Color Res Appl 1990;15:261–265.